



## ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

### **Χρήση στατιστικής σχεσιακής μάθησης με σκοπό την μοντελοποίηση της εμφάνισης και της εξέλιξης χρόνιων παθήσεων**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

**ΝΙΚΟΛΑΟΥ ΠΙΖΑΝΙΑ**

**Επιβλέπων :** Δημήτριος Κουτσούρης  
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιανουάριος 2014





## ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

### Χρήση στατιστικής σχεσιακής μάθησης με σκοπό την μοντελοποίηση της εμφάνισης και της εξέλιξης χρόνιων παθήσεων

#### ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

**ΝΙΚΟΛΑΟΥ ΠΙΖΑΝΙΑ**

**Επιβλέπων :** Δημήτριος Κουτσούρης  
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την

.....  
Δημήτριος Κουτσούρης  
Καθηγητής Ε.Μ.Π.

.....  
Γεώργιος Μασόπουλος  
Καθηγητής Ε.Μ.Π.

.....  
Δημήτριος Φωτιάδης  
Καθηγητής Παν.  
Ιωαννίνων

Αθήνα, Ιανουάριος 2014



(Υπογραφή)

.....

**ΠΙΖΑΝΙΑΣ ΝΙΚΟΛΑΟΣ**

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

**© 2013 – All rights reserved**

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς το συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

## **ΠΕΡΙΛΗΨΗ**

Σκοπός της παρούσης εργασίας είναι η συνοπτική παρουσίαση ορισμένων σύγχρονων μεθόδων σχεσιακής μάθησης οι οποίες βασίζονται στην χρήση Μπεϋζιανών αλλά και Μαρκοβιανών δικτύων. Στο πρώτο κεφάλαιο γίνεται μια μικρή ιστορική ανασκόπηση της σχεσιακής μάθησης. Στο δεύτερο κεφάλαιο εισάγονται οι έννοιες των γραφικών μοντέλων, του συμπερασμού και της μάθησης. Στο τρίτο κεφάλαιο γίνεται μια εισαγωγή στην Μαρκοβιανή λογική και στους τρόπους με τους οποίους μπορεί αυτή να χρησιμοποιηθεί στις διαδικασίες μάθησης και συμπερασμού. Στο τέταρτο κεφάλαιο, προκειμένου να γίνει περισσότερο αντιληπτή η χρησιμότητα ειδικά των Μπεϋζιανών δικτύων, παρουσιάζονται τέσσερα παραδείγματα εφαρμογών από τον χώρο της ιατρικής. Τέλος, στο πέμπτο κεφάλαιο γίνεται μια μικρή αναφορά στις μελλοντικές κατευθύνσεις και προοπτικές των συγκεκριμένων τεχνικών.

## **ABSTRACT**

The purpose of this thesis is the brief presentation of some contemporary methods of relational learning which are based on the use of Bayesian and Markov networks. In the first chapter we present a retrospective overview of relational learning. In the second chapter we introduce the concepts of graphical models, inference and learning. In the third chapter we introduce Markov logic and the ways in which it can be used in learning procedures and inference. In the fourth chapter, in order to exhibit the usefulness of Bayesian networks specifically, four examples are being presented from the world of modern medicine. In the fifth and final chapter we discuss the possible future directions of the pre mentioned techniques.

## ΠΕΡΙΕΧΟΜΕΝΑ

1. Εισαγωγή	3
1.1 Γενικά	3
1.2 Μία μικρή ιστορική ανασκόπηση της σχεσιακής μάθησης	4
1.3 Οι σύγχρονες τάσεις	5
1.4 Στατιστική σχεσιακή μάθηση	6
2. Γραφικά μοντέλα	8
2.1 Εισαγωγικά	9
2.2 Αναπαράσταση	9
2.2.1 Bayesian δίκτυα	10
2.2.2 Υποθέσεις δεσμευμένης (υπό συνθήκες) ανεξαρτησίας σε Bayesian δίκτυα	13
2.2.3 Δίκτυα Markov	16
2.2.4 Ανεξαρτησία σε δίκτυα Markov	19
2.3 Συμπερασμός	21
2.3.1 Ο συμπερασμός ως βελτιστοποίηση	25
2.3.2 Ο ακριβής συμπερασμός ως βελτιστοποίηση	27
2.3.2.1 Το πρόβλημα της βελτιστοποίησης	30
2.3.2.2 Χαρακτηρισμός σταθερού σημείου	32
2.3.3 Αναδραστική διάδοση πίστεως σε δίκτυα Markov με συζεύξεις	34
2.3.4 Προσεγγιστικός συμπερασμός που βασίζεται στην δειγματοληψία	40
2.3.4.1 Μέθοδοι Monte Carlo για αλυσίδες Markov	40
2.3.4.2 Αλυσίδες Markov για γραφικά μοντέλα	42
2.3.4.3 Δειγματοληψία Gibbs	43
2.3.4.4 Κατασκευή μιας αλυσίδας Markov	45
2.3.4.5 Δημιουργία δειγμάτων	45
2.3.4.6 Γενικές παρατηρήσεις	47
2.4 Μάθηση	47
2.4.1 Εκτίμηση παραμέτρων σε Bayesian δίκτυα	48
2.4.1.1 Εκτίμηση παραμέτρων μεγίστης πιθανοφάνειας	48
2.4.1.2 Εκτίμηση Bayesian παραμέτρων	50
2.4.2 Εκτίμηση παραμέτρων σε δίκτυα Markov	52
2.4.3 Μαθαίνοντας την δομή Bayesian δικτύου	54
2.4.3.1 Αποτελέσματα δομής	55
2.4.3.2 Αναζήτηση	60
2.5 Συμπεράσματα	65
3. Μαρκοβιανή λογική (Markov logic): Ένα ενοποιητικό πλαίσιο εργασίας για στατιστική σχεσιακή μάθηση	66
3.1 Η ανάγκη για ένα ενοποιητικό πλαίσιο εργασίας	66
3.2 Δίκτυα Markov	68
3.3 Λογική πρώτης τάξεως	69
3.4 Μαρκοβιανή λογική	72
3.5 Προσεγγίσεις SRL	79
3.5.1 Κατασκευή μοντέλων βασιζόμενη σε γνώση	79

3.5.2 Άλλες προσεγγίσεις λογικού προγραμματισμού	80
3.5.3 Πιθανοτικά σχεσιακά μοντέλα	81
3.5.4 Σχεσιακά δίκτυα Markov	82
3.5.5 Λογιστική αναδρομή δομής	83
3.5.6 Δίκτυα σχεσιακής εξάρτησης	83
3.5.7 Επικαλύψεις και πιθανοτικά μοντέλα σχέσεων οντοτήτων	83
3.5.8 BLOG	84
3.6 Εργασίες SRL	84
3.6.1 Συλλογική κατηγοριοποίηση	84
3.6.2 Πρόβλεψη συνδέσμου	85
3.6.3 Συμπλεγματοποίηση βασιζόμενη σε συνδέσμους	85
3.6.4 Μοντελοποίηση κοινωνικών δικτύων	86
3.6.5 Ταυτοποίηση αντικειμένων	86
3.7 Συμπερασμός	87
3.8 Μάθηση	90
3.9 Πειράματα	93
3.9.1 Συστήματα	94
3.9.1.1 Λογική	95
3.9.1.2 Πιθανότητα	96
3.9.1.3 Επαγωγικός λογικός προγραμματισμός	98
3.9.1.4 Λογική Markov	99
3.9.2 Αποτελέσματα	99
3.9.2.1 Εκπαίδευση με MC-MLE	99
3.9.2.2 Εκπαίδευση με ψευδο-πιθανοφάνεια	100
3.9.2.3 Συμπερασμός	101
3.9.2.4 Σύγκριση συστημάτων	101
3.10 Συμπεράσματα	102
4. Εφαρμογές πιθανοτικών μοντέλων στην διάγνωση και εξέλιξη ασθενειών	104
Εφαρμογή 1. Συμπερασμός διάγνωσης (Diagnostic reasoning)	104
Εφαρμογή 2. Συμπερασμός πρόγνωσης (Prognostic reasoning)	105
Εφαρμογή 3. Επιλογή θεραπείας	107
Εφαρμογή 4. Ανακάλυψη λειτουργικών αλληλεπιδράσεων	108
5. Μελλοντικές κατευθύνσεις και προοπτικές	109
Βιβλιογραφία	110



## 1.Εισαγωγή

### 1.1 Γενικά

Στην πλειοψηφία της βιβλιογραφίας που αφορά την στατιστική μάθηση (statistical learning), είναι συνήθως πρακτική τα διάφορα δεδομένα να αναπαρίστανται με την μορφή σημείων ενός πολυδιάστατου χώρου.

Για οποιαδήποτε συγκεκριμένη μεμονωμένη εργασία (isolated task), όπως για παράδειγμα την ανίχνευση ενός προσώπου σε μία εικόνα ή τον χαρακτηρισμό ενός e-mail ως επιθυμητού ή μη επιθυμητού (spam), συνήθως μπορούμε να θεωρήσουμε τα σχετικά με την εργασία αυτή χαμηλού επιπέδου χαρακτηριστικά (low-level features) (π.χ. pixels, φίλτρα, λέξεις, URLs) και να επιλύσουμε το πρόβλημα χρησιμοποιώντας τα καθιερωμένα εργαλεία των διαφόρων μεθόδων διανυσματικής αναπαράστασης.

Αυτή η θεώρηση, αν και είναι εξαιρετικά βολική για την ανάπτυξη και ανάλυση αλγορίθμων, έχει το μειονέκτημα ότι συνήθως αποκρύπτει την δομή της λογικής που διέπει τα υπό επεξεργασία δεδομένα (data), η οποία όμως είναι κρίσιμης σημασίας για την επίλυση πιο γενικών και περίπλοκων προβλημάτων.

Μπορεί ας πούμε να μην θέλουμε απλώς να ανιχνεύσουμε ένα πρόσωπο μέσα σε μία εικόνα αλλά να το αναγνωρίσουμε κιόλας. Για παράδειγμα να αναγνωρίσουμε "το πρόσωπο μίας ψηλής γυναίκας η οποία στέκεται δίπλα από ένα άγαλμα" ή "το πρόσωπο ενός μικρού παιδιού που τρέχει" κλπ. ή για την περίπτωση του e-mail να θέλουμε να ανιχνεύσουμε όχι απλώς ότι είναι επιθυμητό (not-spam) αλλά ότι πρόκειται συγκεκριμένα κιόλας για "πρόσκληση από τον προϊστάμενο μας στο αυριανό συμβούλιο μαζί με άλλους τρεις συναδέλφους" ή μία "πρόσκληση από τον γείτονα μας για το πάρτι που θα κάνει το ερχόμενο Σαββάτο" .

Ο απώτερος σκοπός μας επομένως είναι, να μην είμαστε απλώς σε θέση να μπορούμε να απαντάμε σε ερωτήσεις τύπου "ναι ή όχι" αλλά να μπορούμε να παράγουμε και να χειριζόμαστε δομές δεδομένων οι οποίες περιλαμβάνουν αντικείμενα (objects) τα οποία περιγράφονται από τις ιδιότητες τους (attributes), την συμμετοχή τους σε "σχέσεις" με άλλα αντικείμενα και την συμμετοχή τους σε "γεγονότα" .

Η πρόκληση επομένως είναι να αναπτυχθεί ο ανάλογος φορμαλισμός, τα μοντέλα και οι αλγόριθμοι που θα αποτελέσουν τον ζητούμενο "συλλογισμό" (reasoning) για την επεξεργασία αυτού του τύπου δομών δεδομένων οι οποίες περιλαμβάνουν αλληλοσχετιζόμενα αντικείμενα.

Κατά τον χειρισμό πραγματικών δεδομένων, όπως είναι οι εικόνες και τα κείμενα, απαιτείται αναπόφευκτα η ικανότητα χειρισμού της αβεβαιότητας που πηγάζει από την ύπαρξη θορύβου αλλά και τις ατελείς πληροφορίες (πχ. ορθογραφικά λάθη για την περίπτωση των κειμένων).

Στα προβλήματα συσχετισμού, η αβεβαιότητα προκύπτει σε πολλά και διαφορετικά επίπεδα.

Εκτός από την αβεβαιότητα που σχετίζεται με τις ιδιότητες ενός αντικειμένου, ενδεχομένως να υπάρχει αβεβαιότητα και ως προς τον τύπο και την ταυτότητα του αντικειμένου ή ακόμα και ως προς το πλήθος (αριθμό) των μελών μίας συλλογής πολλών αντικειμένων.

Η επίλυση προβλημάτων συσχετισμού απαιτεί να γνωρίζουμε το πώς να χειριζόμαστε την αβεβαιότητα που υφίσταται σε όλα τα παραπάνω επίπεδα.

Σκοπός αυτής της εργασίας είναι να παρουσιάσουμε ορισμένα στατιστικά μοντέλα διαδικασιών σχεσιακής μάθησης (relational learning).

Τα μοντέλα αυτά, θα τα αναπτύξουμε με ένα συμπαγή και διαισθητικό τρόπο ο οποίος θα αντικατοπτρίζει αυτές τις σχεσιακές δομές (relational structures) αλλά ταυτοχρόνως θα υποστηρίζει επαρκώς τις διαδικασίες μάθησης και συμπερασμού (inference).

Στην πλειοψηφία τους, τα μοντέλα αυτά, βασίζονται σε συνδυασμούς γραφικών μοντέλων, πιθανοτικών γραμματικών και λογικών τύπων.

## 1.2 Μία μικρή ιστορική ανασκόπηση της σχεσιακής μάθησης

Οι πρώτες προσπάθειες πάνω στον τομέα της μηχανικής μάθησης (MM) συνήθως εστίαζαν στην εκμάθηση ντετερμινιστικών λογικών ιδεών (concepts).

Ένα από τα πρώτα συστήματα σχεσιακής μάθησης ήταν το σύστημα αναζήτησης μάθησης Winston (Winston's search learning system).

Το συγκεκριμένο (τύπου "on-line") σύστημα εκπαιδευόταν χρησιμοποιώντας μία ακολουθία παραδειγμάτων που ήταν σημασμένα ως "αρνητικά" ή ως "θετικά".

Αρχικά, το σύστημα συντηρούσε μία "τρέχουσα υπόθεση" με την μορφή ενός σημασιολογικού δικτύου.

Όταν παρουσιαζόταν ένα νέο παράδειγμα, το σύστημα έκανε μία πρόβλεψη χρησιμοποιώντας την τρέχουσα υπόθεση.

Εάν η πρόβλεψη ήταν σωστή, δεν γινόταν καμία αλλαγή στην υπόθεση.

Εάν όμως η πρόβλεψη ήταν λανθασμένη, τότε εντοπιζόταν το σύνολο των διαφορών μεταξύ της τρέχουσας υπόθεσης και του υπό εξέταση παραδείγματος.

Στην περίπτωση που το υπό εξέταση παράδειγμα ήταν "θετικό", οι διαφορές αυτές χρησιμοποιούνταν για την γενίκευση/διεύρυνση της ιδέας (concept).

Στην περίπτωση που το υπό εξέταση παράδειγμα ήταν "αρνητικό", οι διαφορές χρησιμοποιούνταν για την εξειδίκευση της ιδέας και όχι για την γενίκευση της.

Αργότερα υπήρξαν και πιο προχωρημένα συστήματα σχεσιακής μάθησης που και αυτά όμως χρησιμοποιούσαν μια παρόμοια αναπαράσταση για τις ιδέες (concepts) η οποία βασιζόταν στην λογική.

Αυτή η προσέγγιση MM σταμάτησε να είναι δημοφιλής για αρκετά χρόνια εξαιτίας της δυσκολίας χειρισμού του θορύβου αλλά και των δεδομένων μεγάλης κλίμακας.

Σε αυτό το χρονικό διάστημα, η κοινότητα που ασχολείτο με την MM έστρεψε την προσοχή της σε στατιστικές μεθόδους οι οποίες αγνοούσαν τους ενδεχόμενους συσχετισμούς μεταξύ των δεδομένων (πχ. νευρωνικά δίκτυα, δέντρα αποφάσεων και γενικευμένα γραμμικά μοντέλα).

Αυτές οι μέθοδοι βοήθησαν στο να αναπτυχθεί μεγάλη ακρίβεια σε πολλά προβλήματα που αφορούσαν την σε χαμηλό επίπεδο όραση και την επεξεργασία φυσικής γλώσσας. Το χαρακτηριστικό τους όμως ήταν ότι επικεντρώνονταν στην προτασιακή αναπαράσταση ή στην αναπαράσταση βάσει ιδιοτήτων/τιμών.

Μεγάλη εξαίρεση υπήρξε η κοινότητα που ασχολείτο με τον επαγωγικό λογικό προγραμματισμό (Inductive Logic Programming – ILP).

Η ILP κοινότητα επικέντρωσε τις προσπάθειες της στην εκμάθηση πρώτης τάξεως (ντετερμινιστικών) κανόνων από συσχετισμένα δεδομένα.

Αρχικά η προσοχή της στράφηκε αποκλειστικά στην προσπάθεια σύνθεσης προγραμμάτων μέσω παραδειγμάτων και ήδη υπάρχουσας γνώσης.

Πρόσφατες έρευνες οδήγησαν στην ανακάλυψη κανόνων οι οποίοι προέκυψαν από μεγάλες βάσεις δεδομένων.

Αυτοί οι κανόνες συνήθως χρησιμοποιούνται για προβλέψεις και μπορούν να ερμηνευθούν πιθανοτικά.

Η κοινότητα ILP είχε πολλές επιτυχίες σε διάφορες περιοχές εφαρμογών, συμπεριλαμβανομένων και της ανακάλυψης δισδιάστατων δομικών "προειδοποιήσεων" (alerts) για καρκινογένεση, τρισδιάστατων φαρμακοφόρων για την σχεδίαση φαρμάκων καθώς και την ανάλυση βάσεων δεδομένων χημικών στοιχείων.

### 1.3 Οι σύγχρονες τάσεις

Προσφάτως, η ILP και η MM κοινότητα άρχισαν να συνεργάζονται πολύ στενά συμπληρώνοντας η μία την άλλη.

Πολλοί ILP ερευνητές, πλέον, αναπτύσσουν στοχαστικές και πιθανοτικές αναπαραστάσεις και αλγόριθμους ενώ αντίστοιχα στους πιο παραδοσιακούς MM κύκλους, ερευνητές οι οποίοι στο παρελθόν επικεντρώνονταν σε μαθησιακούς αλγόριθμους που βασίζονταν μόνο σε ιδιότητες, τιμές ή προτασιακή λογική, πλέον εξερευνούν μεθόδους προκειμένου οι αλγόριθμοι τους να αποκτήσουν και την ικανότητα συσχέτισης πληροφοριών.

Προφανώς το ιδανικό και ταυτοχρόνως ζητούμενο, είναι να συνεχιστεί αυτή η τάση αλληλοσυμπλήρωσης και ανταλλαγής ιδεών μεταξύ των δύο κοινοτήτων.

Σύμφωνα με αυτή την επιδίωξη θα προσπαθήσουμε στην παρούσα εργασία, να συνδυάσουμε την σχεσιακή με την στατιστική μάθηση.

Ένα από τα ισχυρότερα κίνητρα για να χρησιμοποιήσει κάποιος σχεσιακά μοντέλα, είναι η δυνατότητα τους να μπορούν να μοντελοποιούν τις εξαρτήσεις μεταξύ σχετικών περιπτώσεων (instances).

Διαισθητικά, θα θέλαμε να χρησιμοποιήσουμε τις πληροφορίες που διαθέτουμε για κάποιο ζήτημα, προκειμένου να βγάλουμε συμπεράσματα για κάποια άλλα σχετιζόμενα με αυτό αντικείμενα.

Για παράδειγμα, ας θεωρήσουμε την περίπτωση των δεδομένων που υπάρχουν στο διαδίκτυο.

Θα θέλαμε να είμαστε σε θέση να μπορούμε να μεταδώσουμε τις πληροφορίες που περιέχει κάποιο έγγραφο (document), για ένα συγκεκριμένο θέμα, προς τα συνδεδεμένα σε αυτό έγγραφα στα οποία μας οδηγεί αλλά και προς τα συνδεδεμένα έγγραφα που μας οδηγούν σε αυτό.

Με την σειρά τους δε αυτά, να μπορούν να μεταδώσουν πληροφορίες σε άλλα έγγραφα και ούτω καθ' εξής.

Πολλοί ερευνητές έχουν προτείνει μία διαδικασία που ακολουθεί αυτή την ιδέα της μετάδοσης πληροφορίας με σχετικιστικό τρόπο.

Ο Chakrabarti και οι συνεργάτες του έχουν περιγράψει έναν αλγόριθμο "relaxation labeling" ο οποίος κάνει χρήση των γειτονικών πληροφοριακών δεσμών.

Ο αλγόριθμος αυτός ξεκινά με την σήμανση να καθορίζεται από έναν text-based ταξινομητή (classifier) ο οποίος είναι κατασκευασμένος από το σύστημα εκπαίδευσης.

Στην συνέχεια χρησιμοποιεί την ήδη υπολογισμένη κλάση των γειτονικών εγγράφων προκειμένου να ξέρει σε ποια κλάση να κατατάξει το υπό εξέταση έγγραφο.

Σε όλη αυτή την διαισθητικότητα που εμφανώς διέπει τέτοιου είδους συστήματα διεργασιών, μπορεί να αποδοθεί ξεκάθαρη σημασιολογία (semantics) αν χρησιμοποιηθούν πιθανοτικά γραφικά μοντέλα.

## 1.4 Στατιστική σχεσιακή μάθηση

Η στατιστική σχεσιακή μάθηση (Statistical Relational Learning - SRL) είναι ένα ανερχόμενο πεδίο έρευνας.

Η έρευνα στον χώρο του SRL προσπαθεί να επιτύχει την αναπαράσταση, την συλλογιστική και την μάθηση σε περιοχές με σχετικότητα μεγάλης πολυπλοκότητας και έντονα πιθανοτική δομή.

Άλλοι όροι που έχουν χρησιμοποιηθεί πρόσφατα συμπεριλαμβάνουν τους όρους "πιθανοτική λογική μάθηση" και "πολυσχεσιακή εξόρυξη δεδομένων".

Επίσης, πολλές από τις εργασίες (tasks) που είναι γνωστές ως " προβλήματα δομημένης πρόβλεψης" , επικαλύπτουν σε μεγάλο βαθμό προβλήματα τα οποία σχετίζονται με την έρευνα στον χώρο του SRL.

Τα περισσότερα προτεινόμενα SRL συστήματα μπορούν να διακριθούν με διάφορα κριτήρια.

Οι πιο συνηθισμένοι φορμαλισμοί αναπαράστασης βασίζονται είτε στην λογική (πχ. rule-based formalisms) είτε σε κάποιο πλαίσιο (frame-based) (πχ. objected-oriented formalisms).

Η πιθανοτική σημασιολογία βασίζεται κυρίως σε γραφικά μοντέλα ή στοχαστικές γραμματικές.

Παλαιότερες SRL προσεγγίσεις συνήθως ορίζονταν με όρους προσανατολισμένων γραφικών μοντέλων (πχ. Bayesian δίκτυα) ενώ πρόσφατα υπάρχει ένα αυξανόμενο ενδιαφέρον για μη προσανατολισμένα μοντέλα (πχ. δίκτυα Markov).

Τα προσανατολισμένα μοντέλα μπορούν να αναπαραστήσουν περίπλοκα γενετικά μοντέλα ενώ τα μη προσανατολισμένα μοντέλα μπορούν να αναπαραστήσουν μη αιτιατές εξαρτήσεις.

Υπάρχουν βεβαίως και άλλες δυνατές εναλλακτικές λύσεις, όπως είναι για παράδειγμα τα δίκτυα εξάρτησης και τα μικτά δίκτυα (προσανατολισμένα μαζί με μη προσανατολισμένα).

Η λογική ερμηνεία για τις περισσότερες SRL γλώσσες (πχ. πιθανοτικά σχεσιακά μοντέλα, Bayesian λογικά προγράμματα, σχεσιακά μοντέλα Markov) συνήθως γίνεται με όρους "ελαχίστων μοντέλων Hebrand" και η πιθανοτική σημασιολογία συνήθως γίνεται με όρους σημασιολογίας "πιθανών κόσμων".

Κάποιες από τις πρώτες προσεγγίσεις, όπως η βασιζόμενη σε γνώση κατασκευή μοντέλων (Knowledge-Based Model Construction - KBMC), στηρίζονταν σε διαδικαστική σημασιολογία, η οποία προσδιορίζει την σημασία μίας έκφρασης με βάση την σχέση της με άλλες εκφράσεις.

Η σημασιολογία πολλών από τα SRL συστήματα δίνεται υπό όρους ενός γραφικού μοντέλου στο οποίο δεν έχει ανατεθεί κάποιος ρόλος ή όπως λέμε ενός γραφικού μοντέλου αναφοράς.

Συνεπώς, ένας τρόπος εξαγωγής συμπερασμάτων σε αυτά τα μοντέλα είναι να γίνει ο κατάλληλος πιθανοτικού τύπου συμπερασμός στο μοντέλο βασικού επιπέδου.

Μία απλή βελτιστοποίηση τύπου KBMC επιτυγχάνεται αν κάνουμε χρήση αναζήτησης (query) κατά την πορεία κατασκευής του δικτύου.

Αντί να κατασκευάσουμε απευθείας ολόκληρο το μοντέλο βασικού επιπέδου, η κατασκευή μπορεί να γίνει πιο αποδοτικά δημιουργώντας κάθε φορά μόνο το τμήμα του δικτύου που χρειάζεται για να απαντηθεί η έρευνα. Αυτό όμως δεν εκμεταλλεύεται καμία από τις εγγενείς δομές εντός του πιθανοτικού μοντέλου.

Ο Pfeffer και οι συνεργάτες του παρατήρησαν ότι σε πολλές περιπτώσεις τα μοντέλα μπορούν να αποδομηθούν σε ασθενώς συζευγμένα συστήματα και κατά αυτό τον τρόπο να αναδειχθεί το πως οι διεπιφάνειες μεταξύ των στοιχείων μπορούν να χρησιμοποιηθούν ώστε να περιλαμβάνεται διαδικασία συμπερασμού εντός των στοιχείων.

Κάτι τέτοιο επιτρέπει την επαναχρησιμοποίηση και την αποθήκευση των συμπερασμών με την μορφή κρυφής μνήμης και έτσι μπορούμε να οδηγηθούμε σε σημαντικές βελτιώσεις της απόδοσης κατά την διάρκεια συμπερασμού.

Πιο γενικού τύπου προσεγγίσεις, όπως είναι η απαλοιφή μεταβλητών πρώτης τάξεως, συνδυάζουν την απαλοιφή μεταβλητών με την ενοποίηση και επιτρέπουν να εκτελεσθεί μία πιο lifted διαδικασία συμπερασμού.

Είναι προφανές ότι η μάθηση είναι βασικό συστατικό για κάθε προσέγγιση SRL.

Η ισχύς της δομημένης αναπαράστασης έγκειται στην ιεραρχική φύση των στατιστικών μοντέλων.

Το πλεονέκτημα των ιεραρχικών μοντέλων που τα κάνει να ξεχωρίζουν από τα επίπεδα στατιστικά μοντέλα είναι η δυνατότητα από κοινού χρήσης παραμέτρων καθώς και η δέσμευση παραμέτρων.

Η από κοινού χρήση παραμέτρων ανακύπτει όταν εν δυνάμει διακριτές παράμετροι του μοντέλου δεσμεύονται ώστε να είναι ίδιες.

Κάτι τέτοιο μπορεί να συμβεί για παράδειγμα σε ένα κρυφό μοντέλο Markov. Λόγω της Μαρκοβιανής υπόθεσης, οι παράμετροι που καθορίζουν το επόμενο στάδιο είναι οι ίδιες κάθε χρονική στιγμή, συνεπώς δεν απαιτούνται διακριτές παράμετροι οι οποίες να είναι δεικτοδοτημένες με συγκεκριμένες τιμές του χρόνου  $t$ , αρκεί να έχουμε απλώς ένα μόνο σύνολο από παραμέτρους  $\theta_{t+1/t}$ .

Αυτή η δέσμευση παραμέτρων δεν μας δίνει απλώς ένα συμπαγές μοντέλο για πολυπληθείς κατηγορίες (κλάσεις) κατανομών αλλά επιπλέον επιτρέπει και τον εύρωστο (robust) υπολογισμό παραμέτρων.

Σε αντίθεση με τα παραδοσιακά σενάρια ML όπου στο εκπαιδευόμενο σύστημα δίνεται ως είσοδος μια ακολουθία από ανεξάρτητες και ισόνομες (Independent and Identically Distributed- I.I.D.) παρατηρήσεις, η είσοδος στους SRL μαθησιακούς αλγορίθμους είναι συνήθως μία μόνο ισχυρώς συνδεδεμένη περίπτωση.

Εάν δεν υπήρχε η από κοινού χρήση των παραμέτρων, μία τέτοια είσοδος δεν θα ήταν ιδιαίτερα χρήσιμη για την πραγματοποίηση στατιστικού συμπερασμού.

Επειδή όμως σε πολλά σημεία μέσα στο μοντέλο χρησιμοποιούνται οι ίδιες παράμετροι, υπάρχει ακόμα και σε αυτή την περίπτωση η δυνατότητα να εξάγουμε χρήσιμα στατιστικά συμπεράσματα από τα δεδομένα προκειμένου να τα χρησιμοποιήσουμε στην συνέχεια σε διαδικασίες στατιστικού συμπερασμού.

Η επιλογή μοντέλου αποτελεί ένα πρόβλημα-πρόκληση για το SRL.

Συγκεκριμένα συνηθισμένα ζητήματα ανακύπτουν κατά επανάληψη, με διάφορες μορφές, σε πολλά SRL συστήματα.

Ένα από τα πιο συνηθισμένα ζητήματα είναι η κατασκευή και η συγκέντρωση χαρακτηριστικών.

Η πλούσια ποικιλία στην δομή σε συνδυασμό με την ανάγκη για μία συμπαγή παραμετροποίηση, οδηγεί στην ανάγκη κατασκευής σχεσιακών χαρακτηριστικών (ή

συγκεντρώσεων) τα οποία θα είναι σε θέση να δίνουν μια εικόνα της τοπικής "γειτονιάς" κάποιας τυχαίας μεταβλητής.

Επειδή είναι αδύνατον να οριστούν με σαφή τρόπο παράγοντες επί όλων των δυνατών γειτονιών, μπορούν να χρησιμοποιηθούν οι συγκεντρώσεις οι οποίες μας δίνουν έναν διαισθητικό τρόπο για την περιγραφή της σχεσιακής γειτονιάς.

Συνηθισμένες συγκεντρώσεις περιλαμβάνουν την λήψη της μέσης ή της πιθανότερης (προσδοκώμενης) τιμής κάποιας ιδιότητας γειτονιάς.

Υπάρχει όμως δυνατότητα και για πιο σύνθετες συγκεντρώσεις οι οποίες θα αφορούν ειδικά κάποια περιοχή.

Οι συγκεντρώσεις έχουν επίσης μελετηθεί και ως μέσο για την προτασιοποίηση (propositionalizing) προβλημάτων σχεσιακής κατηγοριοποίησης .

Στην SRL κοινότητα, οι Perllich και Provost έχουν μελετήσει εκτενώς το ζήτημα της συγκέντρωσης ενώ οι Popescu και Ungar έχουν εργαστεί στον τομέα της ανακάλυψης στατιστικών κατηγορημάτων.

Ένα ακόμα συνηθισμένο ζήτημα που οι ερευνητές έχουν αρχίσει να επεξεργάζονται είναι η αβεβαιότητα στην δομή.

Πολλές από τις πρώτες SRL προσεγγίσεις λάμβαναν υπόψη τους την περίπτωση όπου υπάρχουν ήδη κάποια μοναδική λογική ερμηνεία (ή κάποιος σχετικιστικός σκελετός) η οποία ορίζει ένα σύνολο τυχαίων μεταβλητών και κάποια πιθανοτική κατανομή επί των στιγμιαίων τιμών των τυχαίων μεταβλητών.

Η αβεβαιότητα στην δομή ενισχύει την αβεβαιότητα στην σχεσιακή ερμηνεία.

Οι Koller και Pfeffer παρουσίασαν πολλές μορφές, συμπεριλαμβανομένης της αριθμητικής ανακρίβειας όπου υπάρχει μία κατανομή επί του αριθμού των σχετιζομένων αντικειμένων.

Ο Gatur και οι συνεργάτες του μελέτησαν μαθησιακά μοντέλα με δομική αβεβαιότητα και έδειξαν πως αυτές οι αναπαραστάσεις μπορούσαν να υποστηριχθούν από ένα πιθανοτικό και βασιζόμενο σε λογική σύστημα.

Οι Pasula και Russell μελέτησαν την αβεβαιότητα στην ταυτότητα, που είναι μία μορφή δομικής αβεβαιότητας η οποία επιτρέπει την μοντελοποίηση της αβεβαιότητας που έχει να κάνει με την ταυτότητα μίας αναφοράς (reference).

Τα περισσότερα από αυτά τα μοντέλα βασίζονται στην υπόθεση "κλειστού κόσμου" προκειμένου να ορίσουν την σημασιολογία για τα μοντέλα.

Πιο πρόσφατα, ο Milch και οι συνεργάτες του ερεύνησαν την χρήση μη παραμετρικών μοντέλων τα οποία επιτρέπουν την ύπαρξη άπειρου αριθμού αντικειμένων και υποστηρίζουν και το μοντέλο "ανοικτού κόσμου".

Άλλες νέες ευέλικτες προσεγγίσεις περιλαμβάνουν τα άπειρα σχεσιακά μοντέλα του Kemp και της ερευνητικής ομάδας αυτού καθώς και εκείνα του Xu και των συνεργατών του.

## 2.Γραφικά μοντέλα

Τα πιθανοτικά γραφικά μοντέλα είναι ένα κομψό πλαίσιο εργασίας που συνδυάζει την αβεβαιότητα και τις λογικές δομές προκειμένου να αναπαραστήσει με ένα ενιαίο τρόπο πολύπλοκα φαινόμενα του πραγματικού κόσμου.



Το εργασιακό αυτό πλαίσιο είναι πολύ γενικό, με την έννοια ότι πολλά από τα συνήθως προτεινόμενα στατιστικά μοντέλα (φίλτρα Kalman, κρυφά μοντέλα Markov, μοντέλα Ising) μπορούν να περιγραφούν σαν γραφικά μοντέλα.

Τα γραφικά μοντέλα τυγχάνουν ιδιαίτερου ενδιαφέροντος τις δύο τελευταίες δεκαετίες τόσο λόγω της ευελιξίας και των δυνατοτήτων αναπαράστασης που προσφέρουν, όσο και λόγω της αυξανόμενης δυνατότητας τους για αποτελεσματική μάθηση και συμπερασμό σε μεγάλα δίκτυα.

## 2.1 Εισαγωγικά

Τα γραφικά μοντέλα έχουν γίνει πλέον ένα ιδιαίτερα δημοφιλές εργαλείο για την μοντελοποίηση της αβεβαιότητας. Αυτό συμβαίνει διότι μας δίνουν την δυνατότητα να προσεγγίσουμε την αβεβαιότητα χρησιμοποιώντας τις αρχές της θεωρίας πιθανοτήτων, καθώς επίσης και την δυνατότητα αποτελεσματικής αντιμετώπισης της πολυπλοκότητας με την χρήση της θεωρίας γράφων.

Οι δύο πιο κοινοί τύποι γραφικών μοντέλων είναι τα Bayesian δίκτυα (επίσης γνωστά και ως αιτιατά δίκτυα) και τα δίκτυα Markov (επίσης γνωστά και ως τυχαία πεδία Markov, Markov Random Fields-MRF).

Σε υψηλότερο επίπεδο, σκοπός μας είναι να αναπαραστήσουμε επαρκώς μία από κοινού κατανομή (joint distribution)  $P$  πάνω σε ένα σύνολο τυχαίων μεταβλητών  $X = \{X_1, \dots, X_n\}$ .

Ακόμα και στην απλούστερη περίπτωση όμως, όπου αυτές οι μεταβλητές είναι δυαδικές, μία από κοινού κατανομή θα απαιτούσε τον προσδιορισμό  $2^n$  τιμών που είναι οι πιθανότητες των  $2^n$  διαφορετικών τρόπων αναθέσεως των τιμών  $x_1, \dots, x_n$ .

Συνήθως όμως υπάρχει κάποιου είδους δομή στην κατανομή που μας επιτρέπει να διακρίνουμε την αναπαράσταση της στα επί μέρους δομικά στοιχεία που την αποτελούν.

Η δομή που εκμεταλλεύονται τα γραφικά μοντέλα είναι οι ιδιότητες της ανεξαρτησίας η οποία διέπει πολλά φαινόμενα του πραγματικού κόσμου.

Οι ιδιότητες της ανεξαρτησίας της κατανομής, μπορούν να χρησιμοποιηθούν για να αναπαραστήσουν πολυδιάστατες κατανομές με ένα ενιαίο τρόπο.

Τα πιθανοτικά γραφικά μοντέλα μας παρέχουν μια γλώσσα μοντελοποίησης γενικού χαρακτήρα που μας επιτρέπει να εκμεταλλευτούμε τέτοιου είδους δομές στην αναπαράστασή μας.

Ο συμπερασμός στα πιθανοτικά γραφικά μοντέλα μας παρέχει τους μηχανισμούς ώστε να μπορούμε να συνδυάσουμε όλα αυτά τα στοιχεία (components) με ένα συνεκτικό πιθανοτικό τρόπο.

Η αποτελεσματική μάθηση (εκτίμηση παραμέτρων και επιλογή μοντέλων) στα πιθανοτικά γραφικά μοντέλα πραγματοποιείται από την συμπαγή παραμετροποίηση.

Στην συνέχεια θα καλύψουμε ζητήματα αναπαράστασης, συμπερασμού και μάθησης.

## 2.2 Αναπαράσταση

Οι δύο πιο συνηθισμένες κατηγορίες γραφικών μοντέλων είναι τα "Bayesian δίκτυα" και τα "δίκτυα Markov" .

Η σημασιολογία που διέπει τα Bayesian δίκτυα στηρίζεται στους κατευθυνόμενους (προσανατολισμένους) γράφους και για αυτό τον λόγο ονομάζονται και προσανατολισμένα γραφικά μοντέλα.

Η σημασιολογία που διέπει τα δίκτυα Markov στηρίζεται σε μη κατευθυνόμενους (μη προσανατολισμένους) γράφους.

Τα δίκτυα Markov ονομάζονται και μη προσανατολισμένα γραφικά μοντέλα.

Είναι δυνατόν, αν και όχι τόσο συνηθισμένο, να χρησιμοποιηθούν και μεικτές, προσανατολισμένες και μη προσανατολισμένες, αναπαραστάσεις. Κάτι τέτοιο όμως, δεν θα αποτελέσει αντικείμενο αυτής της εργασίας.

Βασική έννοια για την αναπαράσταση είναι η "υπό συνθήκες ανεξαρτησία".

Έστω  $X, Y$  και  $Z$  σύνολα τυχαίων μεταβλητών. Το σύνολο  $X$  είναι υπό συνθήκες **ανεξάρτητο** από τα σύνολα  $Y$  και  $Z$  αν

$$P(X = x, Y = y / Z = z) = P(X = x / Z = z)P(Y = y / Z = z)$$

για όλες τις τιμές  $x \in Val(X)$ ,  $y \in Val(Y)$  και  $z \in Val(Z)$ .

Χρησιμοποιούμε τον συμβολισμό  $(X \perp Y / Z)$  για να πούμε ότι το σύνολο  $X$  είναι υπό συνθήκες ανεξάρτητο του συνόλου  $Y$  δεδομένου του συνόλου  $Z$ .

Μερικές φορές, όταν είναι προφανές από τα συμφραζόμενα, λέμε απλώς "ανεξάρτητο" ενώ στην πραγματικότητα εννοούμε "υπό συνθήκες ανεξάρτητο".

### 2.2.1 Bayesian δίκτυα

Ο πυρήνας της αναπαράστασης με Bayesian δίκτυα (Bayesian Networks-BN) είναι ο προσανατολισμένος ακυκλικός γράφος (Directed Acyclic Graph - DAG)  $G$ .

Οι κόμβοι (κορυφές) του  $G$  είναι οι τυχαίες μεταβλητές ενώ οι κλάδοι δείχνουν την απευθείας επίδραση που έχει ο ένας κόμβος πάνω στον άλλο.

Ένας τρόπος που μπορούμε να δούμε αυτόν τον γράφο είναι σαν μία δομή δεδομένων που μας παρέχει τον σκελετό για να αναπαραστήσουμε την από κοινού κατανομή με ένα παραγοντοποιημένο τρόπο.

Έστω ότι  $G$  είναι ένας γράφος BN επί των μεταβλητών  $X_1, \dots, X_n$ .

Κάθε τυχαία μεταβλητή  $X_i$  του δικτύου ακολουθεί μία δεσμευμένη κατανομή (Conditional Probability Distribution - CPD) ή αλλιώς όπως λέμε ακολουθεί ένα "τοπικό πιθανοτικό μοντέλο".

Η δεσμευμένη κατανομή πιθανότητας για την τυχαία μεταβλητή  $X_i$  δεδομένων των γονέων της στον γράφο (συμβολίζουμε  $\mathbf{Pa}X_i$ ) είναι

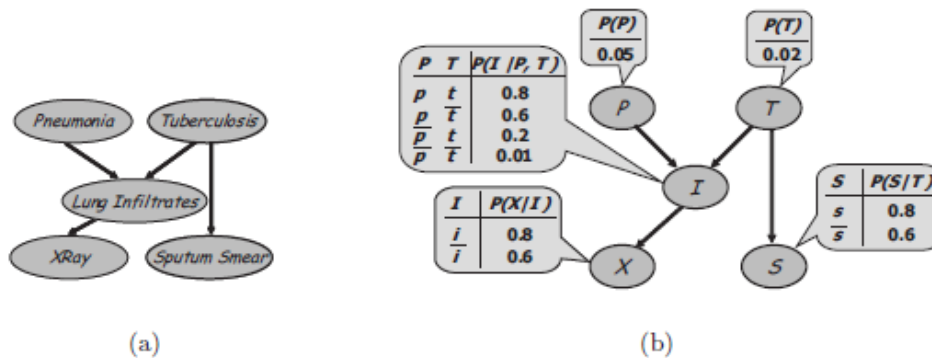
$$P(X_i / \mathbf{Pa}X_i).$$

Οι δεσμευμένες κατανομές πιθανοτήτων (CPD) μπορούν να περιγραφούν με πολλούς διαφορετικούς τρόπους.

Ένας συνηθισμένος τρόπος αναπαράστασης μίας CPD είναι με την μορφή ενός πίνακα που περιέχει μία γραμμή για κάθε πιθανό σύνολο τιμών για τους γονείς του κόμβου, το οποίο περιγράφει την πιθανότητα των διαφορετικών τιμών που μπορεί να λάβει το  $X_i$ . Αυτοί είναι συνήθως πίνακες πολυωνυμικών κατανομών.

Ένας άλλος τρόπος είναι να αναπαραστήσουμε τις κατανομές με την βοήθεια της δομής ενός δέντρου ή με δομή noisy-OR ή noisy-MAX.





Σχήμα 2.1

Σχετικά με το σχήμα 2.1 :

Στο 2.1a φαίνεται ένα απλό Bayesian δίκτυο που δείχνει δύο πιθανές ασθένειες, πνευμονία και φυματίωση, κάθε μία από τις οποίες μπορεί να προκαλέσει διηθήσεις του πνεύμονα.

Οι διηθήσεις του πνεύμονα μπορούν να φανούν σε μία ακτινογραφία ενώ υπάρχει και ένα εξειδικευμένο test που αφορά την φυματίωση.

Όλες οι τυχαίες μεταβλητές είναι Boolean (λογικές).

Στο 2.1b φαίνεται το ίδιο Bayesian δίκτυο μαζί με τον πίνακα δεσμευμένων πιθανοτήτων.

Οι πιθανότητες που φαίνονται αντιστοιχούν στην πιθανότητα να λάβει η τυχαία μεταβλητή την τιμή true (αληθές), δεδομένων των αληθοτιμών που έχουν λάβει οι γονείς.

Η δεσμευμένη πιθανότητα να λάβει η τυχαία μεταβλητή την τιμή false (ψευδής) είναι απλώς 1 μείον την πιθανότητα να είναι true.

Θεωρούμε το Bayesian δίκτυο που φαίνεται στο σχήμα 2.1.

Πρόκειται για ένα απλό παράδειγμα που δείχνει τις αλληλεπιδράσεις μεταξύ δύο δυνατών ασθενειών, της πνευμονίας και της φυματίωσης.

Οι δύο αυτές ασθένειες μπορούν εξίσου να οδηγήσουν σε διηθήσεις του πνεύμονα.

Υπάρχουν δύο tests που μπορούν να γίνουν. Το ένα test είναι μέσω ακτινογραφίας να διαπιστώσουμε εάν ο ασθενής πάσχει από διηθήσεις του πνεύμονα, το άλλο test γίνεται με την λήψη επιχρίσματος προκειμένου να διαπιστωθεί εάν ο ασθενής πάσχει από φυματίωση.

Το σχήμα 2.1a δείχνει την δομή της εξάρτησης μεταξύ των μεταβλητών.

Υποθέτουμε ότι όλες οι μεταβλητές είναι Boolean.

Το σχήμα 2.1b δείχνει τις δεσμευμένες κατανομές για κάθε μία από τις τυχαίες μεταβλητές.

Χρησιμοποιούμε τα αρχικά P, T, I, X και S για συντομία.

Στις ρίζες (roots) έχουμε τις προκαταρκτικές πιθανότητες ώστε ο ασθενής να έχει κάποια από αυτές τις δύο ασθένειες. Η πιθανότητα ο ασθενής να μην πάσχει α priori από κάποια από τις δύο ασθένειες είναι 1 μείον την πιθανότητα να πάσχει από κάποια από αυτές.

Για απλούστευση φαίνονται μονάχα οι πιθανότητες για την περίπτωση της αληθούς τιμής.

Ομοίως, οι δεσμευμένες πιθανότητες για τους υπόλοιπους κόμβους δίνουν την πιθανότητα η τυχαία μεταβλητή να είναι αληθής για διάφορες δυνατές εκφράσεις των γονέων.

Έστω  $G$  ο γράφος ενός Bayesian δικτύου επί των μεταβλητών  $X_1, \dots, X_n$ .

Λέμε ότι μία κατανομή  $P_B$  επί του ίδιου χώρου παραγοντοποιεί σύμφωνα με τον  $G$  εάν η  $P_B$  μπορεί να εκφραστεί ως ένα γινόμενο της μορφής

$$P_B(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Pa}_{X_i}) \quad (2.1)$$

Ένα Bayesian δίκτυο είναι ένα διατεταγμένο ζεύγος της μορφής  $(G, \theta_G)$  όπου η κατανομή  $P_B$  παραγοντοποιεί επί του  $G$  και ορίζεται ως ένα σύνολο δεσμευμένων κατανομών που περιγράφουν τους κόμβους του γράφου  $G$ . Με  $\theta_G$  συμβολίζουμε το σύνολο των κόμβων.

Η παραπάνω εξίσωση ονομάζεται κανόνας της αλυσίδας για Bayesian δίκτυα (chain rule for Bayesian networks) και μας δίνει μία μέθοδο προσδιορισμού της πιθανότητας για κάθε πλήρη ανάθεση επί του συνόλου των τυχαίων μεταβλητών:

Κάθε είσοδος στην ένωση μπορεί να υπολογιστεί σαν ένα γινόμενο παραγόντων.

Ένας παράγοντας για κάθε μεταβλητή.

Κάθε παράγοντας αναπαριστά μια δεσμευμένη πιθανότητα για την μεταβλητή δεδομένων των γονέων της στο δίκτυο.

Το Bayesian δίκτυο του σχήματος 2.1α περιγράφει την ακόλουθη παραγοντοποίηση:

$$P(P, T, I, X, S) = P(P)P(T)P(I | P, T)P(X | I)P(S | T).$$

Μερικές φορές είναι χρήσιμο να σκεφτόμαστε ένα Bayesian δίκτυο σαν να περιγράφουμε μία γενετική διαδικασία.

Μπορούμε να δούμε τον γράφο σαν την κωδικοποίηση μίας δειγματοληπτικής γενετικής διαδικασίας που γίνεται στην φύση, όπου η τιμή για κάθε μεταβλητή επιλέγεται από την φύση χρησιμοποιώντας κάποια κατανομή που εξαρτάται μόνο από τους γονείς.

Με άλλα λόγια, κάθε μεταβλητή (κόμβος του δικτύου) είναι στοχαστική συνάρτηση των γονέων της.

### 2.2.2 Υποθέσεις δεσμευμένης (υπό συνθήκης) ανεξαρτησίας σε Bayesian δίκτυα

Ένας άλλος τρόπος που μπορούμε να δούμε ένα Bayesian δίκτυο είναι σαν αναπαράσταση ενός συνόλου δεσμευμένων (υπό συνθήκη) υποθέσεων ανεξαρτησίας μίας κατανομής.

Αυτές οι δεσμευμένες υποθέσεις ανεξαρτησίας ονομάζονται " τοπικές υποθέσεις Markov" (local Markov assumptions).

Χωρίς να μπορούμε σε πολλές λεπτομέρειες, θα πούμε ότι ο συγκεκριμένος τρόπος αντίληψης των εν λόγω δικτύων είναι σαν να λέμε ότι ένα Bayesian δίκτυο πραγματοποιεί μία παραγοντοποίηση της κατανομής.

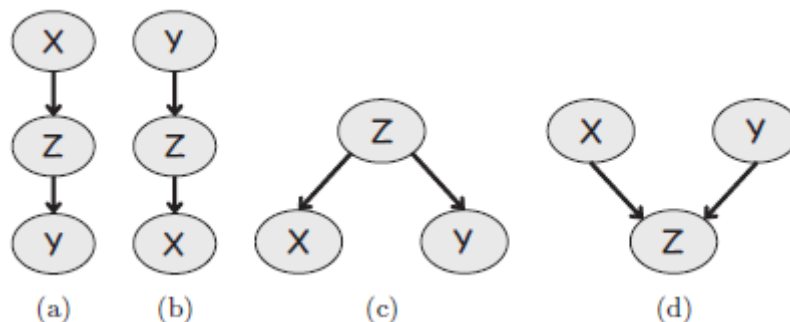
Θεωρώντας δεδομένη μία BN δομή δικτύου  $G$  επί των τυχαίων μεταβλητών  $X_1, \dots, X_n$ , θα συμβολίζουμε ως  $NonDescendantsX_i$  εκείνες τις μεταβλητές του γράφου (δηλαδή τους κόμβους) που δεν είναι απόγονοι (παιδιά) της μεταβλητής  $X_i$ .

Τότε ο γράφος  $G$  κωδικοποιεί το ακόλουθο σύνολο δεσμευμένων υποθέσεων ανεξαρτησίας, οι οποίες ονομάζονται " τοπικές υποθέσεις Markov" :

Για κάθε μεταβλητή  $X_i$  ισχύει ότι

$$X_i \perp NonDescendantsX_i / \mathbf{Pa}X_i .$$

Δηλαδή οι τοπικές υποθέσεις Markov μας λένε ότι κάθε κόμβος  $X_i$  είναι ανεξάρτητος από τους κόμβους (μεταβλητές) που δεν είναι απόγονοι του (nondescendants), δεδομένων πάντοτε των γονέων του.



Σχήμα 2.2

Σχετικά με το σχήμα 2.2:

Στο 2.2a βλέπουμε μία έμμεση αιτιατή δράση/επιρροή (causal effect), στο 2.2b βλέπουμε μία έμμεση επιρροή απόδειξης (evidential effect), στο 2.2c μία συνηθισμένη αιτιατότητα και στο 2.2d μία συνηθισμένη επιρροή.

Το BN του σχήματος 2.1a περιγράφει τις ακόλουθες τοπικές υποθέσεις Markov:

$$(P \perp T / \emptyset), (T \perp P / \emptyset), (X \perp \{P, T, S\} / I) \text{ και } (S \perp \{P, I, X\} / T).$$

Αυτοί δεν είναι οι μόνοι ισχυρισμοί περί ανεξαρτησίας που βρίσκονται υπό κωδικοποιημένη μορφή σε ένα δίκτυο.

Μία γενική διαδικασία που ονομάζεται d-separation (από το directed separation – κατευθυνόμενος χωρισμός) μπορεί να μας πεί εάν ένας ισχυρισμός ανεξαρτησίας "πρέπει" να ισχύει "για κάθε" κατανομή που είναι συνεπής με τον γράφο  $G$ .

Παρ' όλα αυτά, να σημειώσουμε ότι άλλες ανεξαρτησίες μπορεί να ισχύουν για "μερικές" κατανομές που είναι συνεπείς με τον γράφο  $G$ , αυτό μπορεί να οφείλεται στην συγκεκριμένη επιλογή παραμέτρων του δικτύου.

Επιστρέφοντας στον ορισμό της διαδικασίας d-separation, είναι χρήσιμο να δούμε την πιθανοτική επιρροή σαν ένα είδος ροής (flow) μέσα στον γράφο.

Αυτού του είδους η ανάλυση μας λέει πότε η επιρροή του  $X$  μπορεί να ρεύσει μέσω του  $Z$  και να επηρεάσει τα όσα πιστεύουμε για το  $Y$ .

Θα θεωρήσουμε ότι η ροή γίνεται μέσω μη προσανατολισμένων μονοπατιών του γράφου.

Έστω ένα απλό μονοπάτι τριών κόμβων  $X$ - $Y$ - $Z$ .

Εάν η επιρροή μπορεί να ρεύσει από το  $X$  στο  $Y$  μέσω του  $Z$ , τότε λέμε ότι το μονοπάτι  $X$ - $Y$ - $Z$  είναι ενεργό.

Διακρίνουμε τέσσερις περιπτώσεις:

1. Αιτιατό μονοπάτι  $X \rightarrow Z \rightarrow Y$ : Ενεργό αν και μόνο αν δεν

παρατηρείται το  $Z$

2. Αποδεικτικό μονοπάτι  $X \leftarrow Z \leftarrow Y$ : Ενεργό αν και μόνο αν δεν

παρατηρείται το  $Z$

3. Απλή αιτιατότητα  $X \leftarrow Z \rightarrow Y$ : Ενεργό αν και μόνο αν δεν

παρατηρείται το  $Z$

4. Απλή επιρροή  $X \rightarrow Z \leftarrow Y$ : Ενεργό αν και μόνο αν παρατηρείται

το  $Z$  ή κάποιος από τους απογόνους του

Μια δομή του τύπου  $X \rightarrow Z \leftarrow Y$  (όπως αυτή του σχήματος 2.2d) ονομάζεται *v-structure*.

Στο BN του σχήματος 2.1a, το μονοπάτι  $P \rightarrow I \rightarrow X$  είναι ενεργό αν δεν παρατηρείται το  $I$ .

Από την άλλη, το μονοπάτι  $P \rightarrow I \leftarrow T$  είναι ενεργό αν παρατηρείται το  $I$ .

Τώρα ας θεωρήσουμε ένα μεγαλύτερο μονοπάτι  $X_1 - \dots - X_n$ .

Καταλαβαίνουμε ότι για να ρεύσει επιρροή από το  $X_1$  στο  $X_n$  πρέπει πρώτα να ρεύσει μέσα από όλους τους ενδιάμεσους κόμβους του μονοπατιού.

Με άλλα λόγια το  $X_1$  μπορεί να επηρεάσει το  $X_n$  εάν κάθε μονοπάτι δύο ακμών  $X_{i-1}—X_i—X_{i+1}$  κατά μήκος του συνολικού μονοπατιού επιτρέπει την ροή επιρροής μέσα από αυτό.

Έστω ότι ο  $G$  είναι μία δομή BN και  $X_1—\dots—X_n$  είναι ένα μονοπάτι σε αυτόν.

Έστω επίσης ότι  $E$  είναι ένα υποσύνολο των κόμβων του  $G$ .

Το μονοπάτι  $X_1—\dots—X_n$  είναι ενεργό, δεδομένων των κόμβων του υποσυνόλου  $E$ , εάν

1. Όποτε έχουμε μία  $v$ -structure  $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$ , τότε το  $X_i$  ή κάποιος

από τους απογόνους του ανήκουν στο  $E$ .

2. Κανένας άλλος κόμβος του μονοπατιού δεν ανήκει στο  $E$ .

Καταλαβαίνουμε ότι αυτή η ροή επιρροής που εξετάζουμε, θα υπάρχει και στους γράφους όπου μεταξύ δύο κόμβων υπάρχουν περισσότερα από ένα μονοπάτια.

Ένας κόμβος μπορεί να επηρεάσει έναν άλλο κόμβο εάν υπάρχει οποιοδήποτε μονοπάτι κατά μήκος του οποίου μπορεί να ρεύσει επιρροή.

Σύμφωνα με τα παραπάνω μπορούμε να καταλάβουμε τι ακριβώς σημαίνει η διαδικασία  $d$ -separation, η οποία μας δίνει έναν τρόπο διαχωρισμού των κόμβων ενός προσανατολισμένου γράφου (εξ ου και ο όρος  $d$ -separation, από το directed separation).

Έστω  $X, Y, Z$  τρία σύνολα κόμβων του  $G$ . Λέμε ότι τα  $X$  και  $Y$  είναι  $d$ -separated (δεδομένου του  $Z$ ) και συμβολίζουμε  $d\text{-sep}G(X; Y / Z)$ , αν δεν υπάρχει ενεργό μονοπάτι μεταξύ οποιουδήποτε κόμβου  $X \in X$  και  $Y \in Y$  δεδομένου του  $Z$ .

Τέλος, ένα σημαντικό θεώρημα που σχετίζει τις ανεξαρτησίες μίας κατανομής με την παραγοντοποίηση μίας κατανομής είναι το ακόλουθο:

Έστω  $G$  ένας BN γράφος επί ενός συνόλου τυχαίων μεταβλητών  $X$  και  $P$  να είναι μία από κοινού κατανομή επί του ιδίου διαστήματος.

Εάν όλες οι τοπικές ιδιότητες Markov που αφορούν τον  $G$  ισχύουν και στο  $P$ , τότε το  $P$  παραγοντοποιείται σύμφωνα με το  $G$ .

Έστω  $G$  ένας BN γράφος επί ενός συνόλου τυχαίων μεταβλητών  $X$  και  $P$  να είναι μία από κοινού κατανομή επί του ιδίου διαστήματος.

Εάν το  $P$  παραγοντοποιείται σύμφωνα με το  $G$ , τότε όλες οι τοπικές ιδιότητες Markov που αφορούν τον  $G$  ισχύουν και στο  $P$ .

### 2.2.3 Δίκτυα Markov

Η δεύτερη πιο συνηθισμένη κατηγορία πιθανοτικών γραφικών μοντέλων είναι τα " δίκτυα Markov" (Markov Networks - MN) ή όπως αλλιώς ονομάζονται " τυχαία πεδία Markov" (Markov Random Fields - MRF).

Τα δίκτυα Markov στηρίζονται σε μη προσανατολισμένα γραφικά μοντέλα και είναι χρήσιμα για την μοντελοποίηση φαινομένων στα οποία δεν μπορούμε να διαπιστώσουμε κάποια συγκεκριμένη κατευθυντικότητα ως προς την αλληλεπίδραση μεταξύ των μεταβλητών.

Επιπλέον, τα μη προσανατολισμένα μοντέλα πολλές φορές μας παρέχουν μια διαφορετική και ταυτοχρόνως απλούστερη προοπτική των προσανατολισμένων μοντέλων όσον αφορά την δομή ανεξαρτησίας (independence structure) και της διαδικασίας συμπερασμού (inference task).

Ο τρόπος για να αναπαραστήσουμε αυτήν την ιδέα είναι με έναν μη προσανατολισμένο (μη κατευθυνόμενο) γράφο.

Όπως και σε ένα Bayesian δίκτυο, έτσι και στην περίπτωση του γράφου ενός δικτύου Markov (τον συμβολίζουμε με  $H$ ), οι κόμβοι αναπαριστούν τις μεταβλητές ενώ οι ακμές μας δείχνουν την απευθείας πιθανοτική αλληλεπίδραση μεταξύ των γειτονικών μεταβλητών (κόμβων).

Αυτό που πρέπει να δούμε τώρα είναι το πώς μπορούμε να παραμετροποιήσουμε αυτόν τον μη προσανατολισμένο γράφο.

Η δομή του γράφου περιγράφει τις ποιοτικές ιδιότητες της κατανομής.

Προκειμένου να αναπαραστήσουμε την κατανομή, πρέπει να συσχετίσουμε την δομή του γράφου με ένα σύνολο παραμέτρων με τον ίδιο τρόπο που οι δεσμευμένες κατανομές πιθανοτήτων (CPD) χρησιμοποιήθηκαν για την παραμετροποίηση της δομής του κατευθυνόμενου γράφου.

Παρ' όλα αυτά όμως η διαδικασία παραμετροποίησης των δικτύων Markov δεν μπορεί να γίνει με τόσο διαισθητικό τρόπο όπως στην περίπτωση των Bayesian δικτύων καθώς οι παράγοντες (factors) δεν αντιστοιχούν σε πιθανότητες ή σε δεσμευμένες πιθανότητες.

Η πιο γενική μορφή παραμετροποίησης είναι ο παράγοντας (factor):

Έστω  $D$  ένα σύνολο τυχαίων μεταβλητών.

Ορίζουμε ως ένα παράγοντα (factor) μία συνάρτηση που έχει πεδίο ορισμού το  $Val(D)$  και πεδίο τιμών το  $IR^+$ .

Έστω  $H$  ένα δίκτυο Markov. Μία κατανομή  $P_H$  παραγοντοποιεί (factorizes) επί του  $H$  αν σχετίζεται με :

1. Ένα σύνολο υποσυνόλων  $D_1, \dots, D_m$ , όπου κάθε  $D_i$  είναι ένας πλήρης υπογράφος του  $H$ .
2. Παράγοντες  $\pi_1[D_1], \dots, \pi_m[D_m]$ .

τέτοιους ώστε

$$P_{\mathcal{H}}(X_1, \dots, X_n) = \frac{1}{Z} P'(X_1, \dots, X_n),$$

όπου

$$P'_{\mathcal{H}}(X_1, \dots, X_n) = \pi_1[\mathbf{D}_1] \times \pi_2[\mathbf{D}_2] \times \dots \times \pi_m[\mathbf{D}_m]$$

είναι ένα μη κανονικοποιημένο μέτρο

και

$$Z = \sum_{X_1, \dots, X_n} P'_{\mathcal{H}}(X_1, \dots, X_n)$$

είναι μία κανονικοποιημένη σταθερά που ονομάζεται "συνάρτηση διαμέρισης" (partition function).

Μία κατανομή  $P$  που παραγοντοποιεί επί του  $H$  ονομάζεται επίσης και κατανομή Gibbs επί του  $H$ .

Η ονομασία "κατανομή Gibbs" έχει τις ρίζες της στην στατιστική φυσική.

Παρατηρούμε ότι αυτός ο ορισμός είναι παρόμοιος με τον ορισμό παραγοντοποίησης για τα Bayesian δίκτυα.

Στην περίπτωση των Bayesian δικτύων, αποδομήσαμε την κατανομή σε ένα γινόμενο δεσμευμένων κατανομών πιθανοτήτων.

Στην περίπτωση των δικτύων Markov, ο μοναδικός περιορισμός για τις παραμέτρους εντός κάποιου παράγοντα είναι να είναι "μη αρνητικές".

Καθώς κάθε πλήρης υπογράφος είναι ένα υποσύνολο κάποιας κλίκας, μπορούμε να απλοποιήσουμε την παραμετροποίηση εισάγοντας παράγοντες μόνο για τις ομάδες παρά για τις υποκλίκες.

Συγκεκριμένα, έστω  $C_1, \dots, C_k$  οι κλίκες στο  $H$ .

Μπορούμε να παραμετροποιήσουμε την  $P$  χρησιμοποιώντας ένα σύνολο παραγόντων  $\pi_1[C_1], \dots, \pi_k[C_k]$ .

Αυτοί οι παράγοντες ονομάζονται "δυναμικά κλίκας" (clique potentials).

Σε αυτό το σημείο θα μπορούσε να παρασυρθεί κάποιος και να σκεφτεί ότι τα δυναμικά κλίκας αναπαριστούν τις περιθώριες πιθανότητες (marginal probabilities) των υπό εξέταση τυχαίων μεταβλητών.

Κάτι τέτοιο όμως θα ήταν λάθος!

Είναι σημαντικό να σημειώσουμε ότι, αν και φαίνεται ευκολότερη, η παραμετροποίηση χρησιμοποιώντας δυναμικά κλίκας μπορεί να κάνει δυσνόητη (obscure) την δομή που είναι παρούσα στην αρχική παραμετροποίηση και μπορεί να οδηγήσει σε εκθετική αύξηση του μεγέθους της αναπαράστασης.

Συνήθως μας βολεύει να θεωρούμε έναν ελαφρώς διαφορετικό τρόπο καθορισμού των δυναμικών χρησιμοποιώντας έναν λογαριθμικό μετασχηματισμό.

Συγκεκριμένα, μπορούμε να ξαναγράψουμε έναν παράγοντα  $\pi[\mathbf{D}]$  σαν

$$\pi[\mathbf{D}] = \exp(-\epsilon[\mathbf{D}]),$$

όπου το

$$\epsilon[D] = -\ln \pi[D]$$

συνήθως ονομάζεται " συνάρτηση ενέργειας" (energy function).

Η χρήση της λέξης " ενέργεια" προέρχεται από την στατιστική φυσική όπου η πιθανότητα μίας φυσικής κατάστασης (πχ. ο καθορισμός ενός συνόλου ηλεκτρονίων) εξαρτάται από τον αντίστροφο της ενέργειας της.

Με την λογαριθμική αναπαράσταση έχουμε

$$P_H(X_1, \dots, X_n) \propto \exp \left[ - \sum_{i=1}^m \epsilon_i[D_i] \right].$$

Η λογαριθμική αναπαράσταση σιγουρεύει ότι η κατανομή πιθανότητας θα είναι θετική.

Επιπλέον, οι λογαριθμικές παράμετροι μπορούν να λάβουν οποιαδήποτε πραγματική τιμή.

Μια υποκατηγορία δικτύων Markov που συναντάμε πολλές φορές είναι τα "pairwise Markov networks", τα οποία αναπαριστούν κατανομές όπου όλοι οι παράγοντες είναι επί μεμονωμένων μεταβλητών ή επί ζεύγους μεταβλητών.

Πιο συγκεκριμένα, ένα δίκτυο Markov με συζεύξεις επί ενός γράφου  $H$  σχετίζεται με ένα σύνολο "δυναμικών κόμβων" (node potentials)  $\{\pi[X_i] : i = 1, \dots, n\}$  και με ένα σύνολο "δυναμικών ακμών" (edge potentials)  $\{\pi[X_i, X_j] : (X_i, X_j) \in H\}$ .

Η συνολική κατανομή είναι, όπως πάντοτε, το κανονικοποιημένο γινόμενο όλων των δυναμικών (των κόμβων και των ακμών).

Τα τυχαία πεδία Markov (Markov Random Fields - MRF) είναι ελκυστικά λόγω της απλότητας τους και επειδή οι οριακές αλληλεπιδράσεις είναι ένα σπουδαίο ειδικό ζήτημα που ανακύπτει πολλές φορές στην πράξη.

Στο σχήμα 2.3a φαίνεται ένα απλό δίκτυο Markov.

Σε αυτό το παράδειγμα έχουμε τυχαίες μεταβλητές που περιγράφουν την κατάσταση φυματώσης τεσσάρων ασθενών.

Ασθενείς που έχουν έρθει σε επαφή, συνδέονται με μη προσανατολισμένες ακμές.

Οι ακμές δείχνουν τις πιθανότητες μετάδοσης της νόσου.

Για παράδειγμα, ο ασθενής-1 έχει έρθει σε επαφή με τον ασθενή-2 και τον ασθενή-3 αλλά δεν έχει έρθει σε επαφή με τον ασθενή-4.

Στο σχήμα 2.3b φαίνεται το ίδιο δίκτυο Markov μαζί με τα δυναμικά των κόμβων και των ακμών.

Χρησιμοποιούμε τους συμβολισμούς  $P_1, P_2, P_3$  και  $P_4$  για τους τέσσερις ασθενείς.

Στην συγκεκριμένη περίπτωση, όλα τα δυναμικά των κόμβων και των ακμών είναι ίδια, αν και αυτό δεν είναι απαραίτητο.

Τα δυναμικά των κόμβων δείχνουν ότι είναι πιθανότερο οι ασθενείς να μην μολυνθούν.



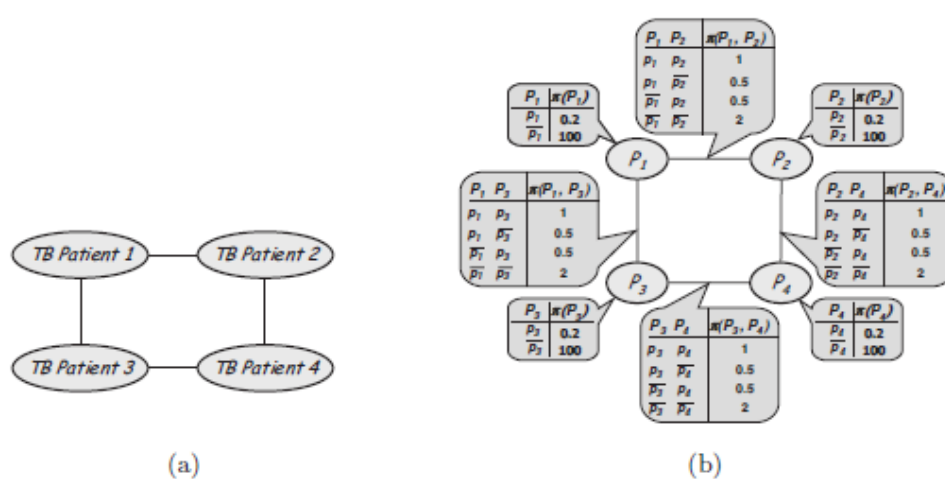
Τα δυναμικά των ακμών δείχνουν ότι είναι πολύ πιθανό δύο άνθρωποι να βρίσκονται στην ίδια κατάσταση ασθένειας, δηλαδή είτε να νοσούν και οι δύο είτε να μην νοσεί κανένας από αυτούς.

Με περισσότερη παρατήρηση βλέπουμε ότι είναι πιθανότερο να μην είναι μολυσμένος κανένας από τους δύο.

### 2.2.4 Ανεξαρτησία σε δίκτυα Markov

Όπως και στην περίπτωση των Bayesian δικτύων, έτσι και στην περίπτωση των δικτύων Markov, η δομή του γράφου περιέχει σε κωδικοποιημένη μορφή ένα σύνολο υποθέσεων ανεξαρτησίας.

Στα δίκτυα Markov η πιθανοτική επιρροή ρέει κατά μήκος των μη προσανατολισμένων μονοπατιών του γράφου αλλά μπορεί να βρει εμπόδιο και να σταματήσει εάν θέσουμε περιορισμούς/προϋποθέσεις στους παρεμβαλλόμενους κόμβους.



Σχήμα 2.3

Περί του σχήματος 2.3:

Στο 2.3a βλέπουμε ένα δίκτυο Markov που περιγράφει την κατάσταση της φυματίωσης σε τέσσερις ασθενείς. Οι ακμές μεταξύ των ασθενών δείχνουν ποιοι ασθενείς έχουν έρθει σε επαφή μεταξύ τους.

Στο 2.3b βλέπουμε το ίδιο δίκτυο Markov μαζί με τις τιμές των δυναμικών των κόμβων και των ακμών.

Μπορούμε να ορίσουμε δύο σύνολα υποθέσεων ανεξαρτησίας, τις "τοπικές ιδιότητες Markov" (local Markov properties) και τις "οικουμενικές ιδιότητες Markov" (global Markov properties).

Οι τοπικές ιδιότητες Markov σχετίζονται με κάθε κόμβο του γράφου και βασίζονται στην ιδέα ότι μπορούμε να εμποδίσουμε κάθε είδους επιρροή πάνω σε ένα κόμβο

θέτοντας τις ανάλογες συνθήκες σε όλους τους κόμβους με τους οποίους γειτονεύει άμεσα.

Έστω  $H$  ένας μη προσανατολισμένος γράφος.

Τότε για κάθε κόμβο  $X \in X$ , η κάλυψη Markov (Markov blanket) του κόμβου  $X$  (συμβολίζουμε ως  $N_H(X)$ ) είναι το σύνολο των γειτονικών κόμβων του  $X$ . Δηλαδή το σύνολο των κόμβων που ενώνονται απευθείας με τον  $X$  μέσω κοινής ακμής.

Ορίζουμε τις "τοπικές ανεξαρτησίες Markov" που σχετίζονται με τον  $H$  ότι είναι

$$\mathcal{I}_\ell(H) = \{(X \perp\!\!\!\perp X - \{X\} - N_H(X) \mid N_H(X)) : X \in X\}.$$

Με άλλα λόγια, οι υποθέσεις Markov δηλώνουν ότι ο  $X$  είναι ανεξάρτητος από τους υπόλοιπους κόμβους του γράφου, δεδομένων των άμεσα γειτονικών του κόμβων.

Το MN του σχήματος 2.3a περιγράφει τις ακόλουθες υποθέσεις Markov :

$$(P1 \perp\!\!\!\perp P4 \mid \{P2, P3\}), (P2 \perp\!\!\!\perp P3 \mid \{P1, P4\}), (P3 \perp\!\!\!\perp P2 \mid \{P1, P4\}), (P4 \perp\!\!\!\perp P1 \mid \{P2, P3\}).$$

Προκειμένου να ορίσουμε τις οικουμενικές ιδιότητες Markov, πρέπει πρώτα να ορίσουμε ποια είναι τα ενεργά μονοπάτια σε ένα μη προσανατολισμένο γράφο.

Έστω  $H$  ένα δίκτυο Markov και  $X_1 - \dots - X_k$  ένα μονοπάτι σε αυτό.

Έστω επίσης  $E \subseteq X$  ένα σύνολο " παρατηρούμενων μεταβλητών" .

Το μονοπάτι  $X_1 - \dots - X_k$  είναι ενεργό, δεδομένου του συνόλου  $E$ , αν καμία από τις τυχαίες μεταβλητές  $X_i, i = 1, \dots, k$  δεν ανήκει στο  $E$ .

Χρησιμοποιώντας αυτό τον ορισμό μπορούμε στην συνέχεια να ορίσουμε την έννοια του " διαχωρισμού" (separation) σε έναν μη κατευθυνόμενο γράφο.

Πρόκειται για το ανάλογο του d-separation.

Λέμε ότι ένα σύνολο κόμβων  $Z$  χωρίζει (separates) το  $X$  και το  $Y$  στο  $H$ , γράφουμε  $sep_H(X; Y \mid Z)$  αν δεν υπάρχει ενεργό μονοπάτι μεταξύ οποιουδήποτε κόμβου  $X \in X$  και οποιουδήποτε κόμβου  $Y \in Y$ , δεδομένου του  $Z$ .

Ορίζουμε τις " οικουμενικές υποθέσεις Markov" (global Markov assumptions) που σχετίζονται με το  $H$  ως

$$I(H) = \{(X \perp\!\!\!\perp Y \mid Z) : sep_H(X; Y \mid Z)\}$$

Όπως και στην περίπτωση των Bayesian δικτύων, έτσι και εδώ μπορούμε να συνδέσουμε τις "τοπικές ιδιότητες Markov" με τις "οικουμενικές ιδιότητες Markov" .

Για την ακρίβεια, οι υποθέσεις (assumptions) είναι ισοδύναμες αλλά μόνο για θετικές κατανομές.

Με έναν πρόχειρο τρόπο θα μπορούσαμε να πούμε ότι μία κατανομή είναι θετική εάν κάθε δυνατή από κοινού παράθεση έχει πιθανότητα γνήσια μεγαλύτερη του μηδενός. Αρχίζουμε με το ανάλογο του θεωρήματος 2.7 το οποίο μας έλεγε ότι μια κατανομή Gibbs ικανοποιεί τις οικογενειακές ανεξαρτησίες που σχετίζονται με τον γράφο.

Έστω  $P$  μια κατανομή επί του  $X$  και  $H$  ένα δίκτυο Markov επί του  $X$ .

Εάν η  $P$  είναι μια κατανομή Gibbs επί του  $H$ , τότε όλες οι τοπικές ιδιότητες Markov που σχετίζονται με το  $H$  ισχύουν και στο  $P$ .

Για την αντίθετη κατεύθυνση, όπου ξεκινάμε από τις οικογενειακές ιδιότητες ανεξαρτησίας μιας κατανομής και καταλήγουμε στην παραγοντοποίηση, έχουμε το θεώρημα Hammersley-Clifford.

Σε αντίθεση με τα Bayesian δίκτυα, αυτή η κατεύθυνση δεν ισχύει στην γενική περίπτωση παρά μόνο όταν, επιπλέον, ισχύει η υπόθεση ότι η  $P$  είναι θετική κατανομή.

Έστω  $P$  μία θετική κατανομή επί του  $X$  και  $H$  ένα δίκτυο Markov επί του  $X$ .

Εάν όλοι οι περιορισμοί ανεξαρτησίας που προκύπτουν από το  $H$  ισχύουν για την  $P$ , τότε η  $P$  είναι μία κατανομή Gibbs επί του  $H$ .

Αυτό το αποτέλεσμα μας δείχνει ότι, για την περίπτωση των θετικών κατανομών, η οικογενειακή ιδιότητα Markov συνεπάγεται ότι η κατανομή παραγοντοποιείται σύμφωνα με την δομή του δικτύου.

Δηλαδή, για αυτή την κλάση κατανομών, έχουμε ότι μια κατανομή  $P$  παραγοντοποιείται επί ενός δικτύου Markov  $H$  αν και μόνο αν όλες οι ανεξαρτησίες που συνεπάγονται από το  $H$  ισχύουν και στην  $P$ .

Η υπόθεση της "θετικότητας" είναι απαραίτητη προκειμένου να ισχύει αυτό το συμπέρασμα.

### 2.3 Συμπερασμός

Τόσο τα κατευθυνόμενα όσο και τα μη κατευθυνόμενα γραφικά μοντέλα αναπαριστούν μια πλήρη από κοινού κατανομή πιθανότητας επί του  $X$ .

Θα περιγράψουμε κάποιους από τους κυριότερους τύπους αναζητήσεων (query types) στους οποίους κάποιος θα ήθελε να μπορεί να απαντήσει έχοντας στην διάθεση του μια από κοινού κατανομή και θα συζητήσουμε περί της υπολογιστικής πολυπλοκότητας που ενέχει η απάντηση τέτοιων αναζητήσεων όταν χρησιμοποιούμε γραφικά μοντέλα.

Ο πιο συνηθισμένος τύπος αναζήτησης είναι η "αναζήτηση δεσμευμένης πιθανότητας" (conditional probability query)  $P(Y / E = e)$ .

Μία τέτοια αναζήτηση αποτελείται από δύο μέρη:

Το πρώτο μέρος είναι οι αποδείξεις και πρόκειται για ένα υποσύνολο τυχαίων μεταβλητών  $E$  του δικτύου μαζί με μία παράθεση  $e$  σε αυτές τις μεταβλητές.

Το δεύτερο μέρος είναι η αναζήτηση (query) και πρόκειται για ένα υποσύνολο τυχαίων μεταβλητών  $Y$  του δικτύου.  
Σκοπός μας είναι να υπολογίσουμε την

$$P(Y | E = e) = \frac{P(Y, e)}{P(e)}$$

Δηλαδή θέλουμε να υπολογίσουμε την κατανομή επί των τιμών  $y$  του  $Y$ , με την προϋπόθεση (δέσμευση) ότι  $E = e$ .

Ένας άλλος τύπος αναζήτησης που προκύπτει συχνά είναι να βρούμε την "πιο πιθανή ανάθεση" (most probable assignment) για κάποιο υποσύνολο μεταβλητών.

Όπως και με τις αναζητήσεις δεσμευμένης πιθανότητας, έτσι και εδώ έχουμε την απόδειξη  $E = e$ .

Σε αυτή την περίπτωση όμως προσπαθούμε να υπολογίσουμε την πιο πιθανή ανάθεση σε κάποιο υποσύνολο των εναπομεινάντων μεταβλητών.

Αυτό το πρόβλημα έχει δύο παραλλαγές, με την πρώτη παραλλαγή να είναι μία ειδική περίπτωση της δεύτερης παραλλαγής (η οποία όμως έχει ιδιαίτερη σημασία).

Η πιο απλή παραλλαγή αυτής της εργασίας (task) είναι οι αναζητήσεις "πιο πιθανής εξήγησης" (Most Probable Explanation queries – MPE queries).

Μια αναζήτηση MPE προσπαθεί να βρεί την πιο πιθανή ανάθεση για όλες τις μεταβλητές που δεν είναι "αποδεικτικές" (non-evidence variables).

Πιο συγκεκριμένα, εάν υποθέσουμε ότι  $W = X - E$ , σκοπός μας είναι να βρούμε την πιο πιθανή ανάθεση για τις μεταβλητές του  $W$  δεδομένου ότι  $E = e$ .

Γράφουμε  $\operatorname{argmax}_w P(w, e)$ , όπου γενικά το  $\operatorname{argmax}_x f(x)$  αντιπροσωπεύει την τιμή του  $x$  για την οποία η  $f(x)$  είναι μεγιστική (maximal).

Να σημειωθεί ότι μπορεί να υπάρχουν περισσότερες της μίας ανάθεσης που να έχουν την υψηλότερη ύστερη πιθανότητα (the highest posterior probability).

Σε αυτήν την περίπτωση, μπορούμε να αποφασίσουμε αν σκοπός της MPE θα είναι να μας δώσει το σύνολο των πιθανών αναθέσεων ή να μας δώσει ένα αυθαίρετο μέλος αυτού του συνόλου.

Στην δεύτερη παραλλαγή, την λεγόμενη maximum a posteriori (MAP) αναζήτηση, έχουμε ένα υποσύνολο μεταβλητών  $Y$  οι οποίες μορφοποιούν την αναζήτηση μας.

Σκοπός μας είναι να βρεθεί η πιο πιθανή ανάθεση για τις μεταβλητές του  $Y$  δεδομένου ότι  $E = e$ , δηλαδή το  $\operatorname{argmax}_y P(y | e)$ .

Αυτή η κλάση αναζητήσεων είναι προφανώς πιο γενική από τις MPE αναζητήσεις, οπότε ίσως να μην είναι προφανές για ποιο λόγο μας ενδιαφέρει η κλάση των MPE αναζητήσεων ως ειδική περίπτωση.

Η διαφορά γίνεται προφανής εάν γράψουμε επί τούτου την έκφραση για μία γενική MAP αναζήτηση.

Εάν υποθέσουμε ότι  $Z = X - Y - E$ , τότε σκοπός της MAP αναζήτησης είναι να υπολογιστεί το

$$\operatorname{argmax}_Y \sum_Z P(Y, Z | e)$$

Οι αναζητήσεις MAP εμπεριέχουν τόσο πρόσθεση όσο και μεγιστοποίηση.

Κατά κάποιο τρόπο δηλαδή περιέχουν στοιχεία αναζήτησης δεσμευμένης πιθανότητας αλλά και αναζήτησης MPE.

Αυτός ο συνδυασμός καθιστά το έργο της αναζήτησης MAP δυσκολότερο σε σχέση με άλλες εργασίες.

Συγκεκριμένα, υπάρχουν τεχνικές και αναλύσεις για την MPE εργασία που δεν γενικεύονται ως την περίπτωση της MAP εργασίας.

Αυτή η παρατήρηση, σε συνδυασμό με το γεγονός ότι η περίπτωση MPE είναι λογικά η πιο συνηθισμένη δίνει στην MPE τόση αξία ώστε να την θεωρήσουμε σαν μία ξεχωριστή εργασία.

Να σημειωθεί ότι στην βιβλιογραφία που αφορά τόσο την στατιστική όσο και τα γραφικά μοντέλα, ο όρος MAP χρησιμοποιείται συχνά εννοώντας MPE. Η διαφορά όμως εν γένει θα είναι προφανής από τα συμφοραζόμενα.

Ένα γραφικό μοντέλο μπορεί να χρησιμοποιηθεί προκειμένου να απαντηθούν όλοι οι τύποι αναζητήσεων που περιγράφηκαν παραπάνω.

Αυτό που κάνουμε απλώς είναι να δημιουργούμε την από κοινού κατανομή και να αφαιρούμε την ένωση (στην περίπτωση της αναζήτησης δεσμευμένης πιθανότητας), να ψάχνουμε για την πιο πιθανή είσοδο (στην περίπτωση της MPE αναζήτησης) ή να κάνουμε και τα δύο (στην περίπτωση της αναζήτησης MAP).

Παρ' όλα αυτά, η συγκεκριμένη προσέγγιση δεν είναι ικανοποιητική για το πρόβλημα του συμπερασμού καθώς οδηγεί στην εκθετική αύξηση της από κοινού κατανομής την οποία το γραφικό μοντέλο σχεδιάστηκε για να αποφεύγει.

Υποθέτουμε ότι έχουμε να κάνουμε με ένα σύνολο παραγόντων  $F$  επί ενός συνόλου μεταβλητών  $X$ .

Αυτό το σύνολο παραγόντων ορίζει μία εν δυνάμει μη κανονικοποιημένη συνάρτηση

$$P_{\mathcal{F}}(\mathcal{X}) = \prod_{\phi \in \mathcal{F}} \phi \quad (2.2)$$

Για ένα Bayesian δίκτυο χωρίς αποδεικτικά (evidence), οι παράγοντες είναι απλώς οι δεσμευμένες κατανομές και η κατανομή  $P_{\mathcal{F}}$  είναι μια κανονικοποιημένη κατανομή.

Για ένα Bayesian δίκτυο  $B$  με αποδεικτικά  $E=e$ , οι παράγοντες είναι δεσμευμένες κατανομές που περιορίζονται από το  $e$  και  $P_{\mathcal{F}}(X) = P_B(X, e)$ .

Για ένα δίκτυο Markov  $H$  (με ή χωρίς αποδεικτικά), οι παράγοντες είναι τα (περιορισμένα-restricted) δυναμικά συμβατότητας (compatibility potentials) και  $P_{\mathcal{F}}$  είναι η μη κανονικοποιημένη κατανομή  $P_H$  πριν την διαίρεση με την συνάρτηση διαμερισμού (partition function).

Να τονιστεί ότι οι περισσότερες πράξεις που μπορεί να κάνει κάποιος σε μια κανονικοποιημένη κατανομή μπορούν να γίνουν και σε μία μη κανονικοποιημένη κατανομή.

Γι' αυτό μπορούμε να περιθωριοποιήσουμε (marginalize) την  $P_{\mathcal{F}}$  σε ένα υποσύνολο των μεταβλητών βγάζοντας έξω (summing out) τις υπόλοιπες.

Μπορούμε επίσης να θεωρήσουμε την δεσμευμένη πιθανότητα

$$P_{\mathcal{F}}(X / Y) = P_{\mathcal{F}}(X, Y) / P_{\mathcal{F}}(Y).$$

Οπότε για αυτό το μέρος της παρούσης εργασίας θα χειριζόμαστε την  $P_{\mathcal{F}}$  σαν να είναι μια κατανομή, αγνοώντας το γεγονός ότι μπορεί να είναι μη κανονικοποιημένη.

Στην χειρότερη περίπτωση, η πολυπλοκότητα του πιθανοτικού συμπερασμού είναι αναπόφευκτη.

Στα επόμενα θεωρούμε ότι το σύνολο των παραγόντων  $\{\phi \in F\}$  του γραφικού μοντέλου που ορίζει την επιθυμητή κατανομή μπορεί να σε έναν πολυωνυμικό αριθμό bits (εννοώντας τον αριθμό των μεταβλητών).

Τα επόμενα προβλήματα αποφάσεων (decision problems) είναι NP-complete :

1. Δοθέντων, μιας κατανομής  $P_F$  επί του συνόλου  $X$ , μιας μεταβλητής  $X \in X$  και μιας τιμής  $x \in Val(X)$ , να αποφασιστεί αν  $P_F(X = x) > 0$ .
2. Δοθείσης μιας κατανομής  $P_F$  επί του συνόλου  $X$  και ενός αριθμού  $\tau$ , να αποφασιστεί αν υπάρχει ανάθεση  $x$  για την  $X$  τέτοια ώστε  $P_F(x) > \tau$ .

Το επόμενο πρόβλημα είναι #P-complete :

Δοθέντων, μιας κατανομής  $P_F$  επί του συνόλου  $X$ , μιας μεταβλητής  $X \in X$  και μιας τιμής  $x \in Val(X)$  να υπολογιστεί η  $P_F(X = x)$ .

Τα παραπάνω αποτελέσματα φαίνονται μάλλον δυσάρεστα καθότι μας λένε ότι κάθε τύπος συμπερασμού σε γραφικά μοντέλα είναι NP δυσκολίας ή ακόμα δυσκολότερα. Για την ακρίβεια, ακόμα και το απλό πρόβλημα υπολογισμού της κατανομής επί μίας δυαδικής μεταβλητής είναι NP δυσκολίας.

Υποθέτοντας ότι η καλύτερη υπολογιστική απόδοση που μπορούμε να επιτύχουμε για προβλήματα NP δυσκολίας είναι εκθετική για την χειρότερη περίπτωση, φαίνεται σαν να μην υπάρχει ελπίδα για αποδοτικούς αλγόριθμους ακόμα και για την πιο απλή περίπτωση συμπερασμού.

Παρ' όλα αυτά όμως, όπως θα δούμε παρακάτω, η αύξηση (blowup) της χειρότερης περίπτωσης μπορεί να αποφευχθεί.

Για όλα τα υπόλοιπα μοντέλα, θα καταφύγουμε σε προσεγγιστικές τεχνικές συμπερασμού.

Να σημειώσουμε ότι τα αποτελέσματα χειρότερης περίπτωσης για προσεγγιστικό συμπερασμό είναι επίσης αρνητικά :

Το επόμενο πρόβλημα είναι NP δυσκολίας για οποιοδήποτε  $\epsilon \in (0, 1/2)$ :

Δοθέντων, μιας κατανομής  $P_F$  επί του συνόλου  $X$ , μιας μεταβλητής  $X \in X$  και μιας τιμής  $x \in Val(X)$ , να βρεθεί αριθμός  $\tau$  τέτοιος ώστε

$$|P_F(X = x) - \tau| \leq \epsilon.$$

Ευτυχώς, πολλοί τύποι ακριβούς συμπερασμού μπορούν να πραγματοποιηθούν αποδοτικά για μια ιδιαίτερης σημασίας κατηγορία γραφικών μοντέλων (μικρό εύρος δέντρου - low treewidth) που θα ορίσουμε στην συνέχεια.

Για ένα μεγάλο αριθμό μοντέλων όμως, ο ακριβής συμπερασμός δεν είναι εφικτός οπότε καταφεύγουμε σε προσεγγίσεις.

Μιλώντας με την ευρύτερη έννοια υπάρχουν δύο κύρια πλαίσια εργασίας (frameworks) για πιθανοτικό συμπερασμό :

Το πλαίσιο που βασίζεται στην βελτιστοποίηση (optimization - based) και το πλαίσιο που βασίζεται στην δειγματοληψία (sampling - based).

Αλγόριθμοι ακριβούς συμπερασμού έχουν προκύψει ιστορικά από τον τομέα του δυναμικού προγραμματισμού, με προσεκτική αποφυγή των επαναλαμβανόμενων υπολογισμών.

Εδώ θα κάνουμε μία, κατά κάποιο τρόπο, μη συμβατική προσέγγιση παρουσιάζοντας τον ακριβή και τον προσεγγιστικό συμπερασμό σε ένα ενοποιημένο πλαίσιο εργασίας που στηρίζεται στην βελτιστοποίηση.

Για τον λόγο αυτό θα ξεκινήσουμε θεωρώντας προσεγγιστικό συμπερασμό και στην πορεία θα παρουσιάσουμε τις συνθήκες υπό τις οποίες μπορούν να εξαχθούν ακριβή συμπεράσματα.

### 2.3.1 Ο συμπερασμός ως βελτιστοποίηση

Οι μέθοδοι που κατατάσσονται στο πλαίσιο εργασίας βελτιστοποίησης βασίζονται σε μία απλή αρχή :

Καθόρισε μια κλάση-στόχο (target class) που να περιέχει "εύκολες" κατανομές  $Q$  και μετά ψάξε για κάποιο συγκεκριμένο μέλος  $Q$  της κλάσης αυτής το οποίο να είναι η "καλύτερη" προσέγγιση της κατανομής  $P_F$ .

Οι αναζητήσεις (queries) τότε μπορούν να απαντηθούν κάνοντας συμπερασμό με το  $Q$  παρά με την  $P_F$ .

Οι συγκεκριμένοι αλγόριθμοι που αναφέρονται στην βιβλιογραφία διαφέρουν μεταξύ τους σε πολλά λεπτά σημεία.

Οι περισσότεροι όμως από αυτούς μπορούν να θεωρηθούν ως βελτιστοποίηση μίας συνάρτησης-στόχου (target function) προκειμένου να μετρηθεί η ποιότητα της προσέγγισης.

Ας υποθέσουμε ότι θέλουμε να προσεγγίσουμε την  $P_F$  με μια άλλη κατανομή  $Q$ .

Καταλαβαίνουμε ότι θέλουμε να επιλέξουμε την  $Q$  έτσι ώστε να είναι πολύ κοντά στην  $P_F$ .

Υπάρχουν πολλοί δυνατοί τρόποι για να μετρήσουμε την απόσταση μεταξύ δύο κατανομών, όπως το Ευκλείδιο μέτρο ( $L_2$ ) ή η απόσταση  $L_1$ .

Η κύρια πρόκληση όμως είναι ότι έχουμε ως σκοπό να αποφύγουμε να κάνουμε συμπερασμό με την κατανομή  $P_F$ .

Συγκεκριμένα, δεν μπορούμε να υπολογίσουμε αποδοτικά τις περιθώριες κατανομές της  $P_F$ .

Γι' αυτό χρειαζόμαστε μεθόδους που να μας επιτρέπουν να βελτιστοποιούμε την απόσταση (τεχνικά μιλώντας, την απόκλιση) μεταξύ του  $Q$  και της  $P_F$  χωρίς να απαιτείται να απαντήσουμε σε δύσκολες αναζητήσεις στην  $P_F$ .

Αυτή η απαίτηση ίσως να φαίνεται αδύνατον να ικανοποιηθεί a priori.

Παρ' όλα αυτά όμως, υπάρχει ένα μέτρο της απόστασης, η σχετική εντροπία (relative entropy) ή αλλιώς KL-απόκλιση (KL-divergence) που μας επιτρέπει να εκμεταλλευτούμε την δομή της  $P_F$  χωρίς να κάνουμε συμπερασμό με αυτή.

Γενικώς, η σχετική εντροπία μεταξύ των κατανομών  $P_1$  και  $P_2$  ορίζεται ως

$$D(P_1||P_2) = E_{P_1} \left[ \ln \frac{P_1(X)}{P_2(X)} \right]$$

Η σχετική εντροπία είναι πάντοτε μη αρνητική και ισούται με 0 αν και μόνο αν  $P_1 = P_2$ .

Συνεπώς μπορούμε να χρησιμοποιήσουμε την σχετική εντροπία σαν ένα μέτρο της απόστασης και έτσι να ψάξουμε να βρούμε μία προσέγγιση  $Q$  για την  $P_F$ , η οποία να την ελαχιστοποιεί.

Η σχετική εντροπία όμως δεν είναι συμμετρική



$$D(P_1|P_2) \neq D(P_2|P_1)$$

Ίσως ,a priori, να φαίνεται ότι το  $ID(P_F//Q)$  είναι ένα πιο κατάλληλο μέτρο για τον προσεγγιστικό συμπερασμό, καθώς μία από τις κυριότερες (σε επίπεδο θεωρητικής πληροφορικής) επιβεβαιώσεις/επαληθεύσεις για την σχετική εντροπία είναι ο αριθμός των bits που χάνεται καθώς κωδικοποιείται μία κατανομή "πραγματικού" μηνύματος  $P_F$  χρησιμοποιώντας μια προσέγγιση  $Q$ .

Ωστόσο, ο υπολογισμός της επονομαζόμενης " M-προβολής  $Q$  της  $P_F$  " (M-projection  $Q$  of  $P_F$ ), γράφουμε  $\operatorname{argmin}_Q ID(P_F//Q)$ , είναι στην πραγματικότητα ισοδύναμο με το να κάνουμε συμπερασμό στην  $P_F$ .

Κατά κάποιον παράδοξο τρόπο όμως, όπως θα δείξουμε παρακάτω, αυτή η ιδέα δεν εφαρμόζεται στην λεγόμενη I-προβολή (I-projection) σύμφωνα με την οποία :

Μπορούμε να εκμεταλλευτούμε την δομή της  $P_F$  για την αποδοτική βελτιστοποίηση του  $\operatorname{argmin}_Q ID(P_F//Q)$  "χωρίς" να κάνουμε συμπερασμό στην  $P_F$ .

Ένας επιπλέον λόγος για να χρησιμοποιήσουμε την σχετική εντροπία σαν μέτρο της απόστασης, στηρίζεται στο ακόλουθο αποτέλεσμα που συσχετίζει την σχετική εντροπία  $ID(P_F//Q)$  με την συνάρτηση διαμερισμού  $Z$  :

Ισχύει ότι

$$\ln Z = F[P_F, Q] + D(Q|P_F) \quad (2.3)$$

όπου  $F[P_F, Q]$  είναι το συναρτησιακό (functional) της ενέργειας :

$$F[P_F, Q] = \sum_{\phi \in \mathcal{F}} E_Q[\ln \phi] + H_Q(\mathcal{X}).$$

Αυτή η πρόταση έχει πολλές σημαντικές συνέπειες.

Να σημειωθεί ότι ο όρος  $\ln Z$  δεν εξαρτάται από το  $Q$ , οπότε η ελαχιστοποίηση της σχετικής εντροπίας  $ID(P_F//Q)$  είναι ισοδύναμη με την μεγιστοποίηση του συναρτησιακού της ενέργειας  $F[P_F, Q]$ .

Ο όρος " συναρτησιακό της ενέργειας" σχετίζεται με ιδέες της στατιστικής φυσικής και είναι το αντίθετο αυτού που ονομάζεται "ελεύθερη ενέργεια Helmholtz" (Helmholtz free energy).

Στην συνέχεια αυτού του κεφαλαίου παρουσιάζουμε το πρόβλημα εύρεσης μιας καλής προσέγγισης  $Q$  ως πρόβλημα μεγιστοποίησης του συναρτησιακού της ενέργειας ή ισοδύναμα ως πρόβλημα ελαχιστοποίησης της σχετικής εντροπίας.

Το συναρτησιακό της ενέργειας έχει να κάνει με τις προσδοκίες στο  $Q$ .

Όπως θα δείξουμε, επιλέγοντας την προσέγγιση  $Q$  ,που επιτρέπει αποδοτικό συμπερασμό, μπορούμε να αποτιμήσουμε το συναρτησιακό της ενέργειας καθώς και να το βελτιστοποιήσουμε αποτελεσματικά.

Επιλέον, καθώς  $ID(Q//P_F) \geq 0$ , έχουμε ότι  $\ln Z \geq F[P_F, Q]$ .

Αυτό σημαίνει ότι το συναρτησιακό της ενέργειας είναι ένα "κάτω φράγμα" της τιμής του λογαρίθμου της συνάρτησης διαμερισμού  $Z$ , για κάθε  $Q$  που μπορεί να επιλέξουμε.

Για ποιο λόγο όμως αυτό το συμπέρασμα είναι σημαντικό?



Να υπενθυμίσουμε ότι σε προσανατολισμένα μοντέλα η συνάρτηση διαμερισμού  $Z$  είναι η πιθανότητα του αποδεικτικού (evidence).

Ο υπολογισμός της συνάρτησης διαμερισμού είναι συχνά το δυσκολότερο μέρος σε έναν συμπερασμό και γι' αυτό το συγκεκριμένο θεώρημα μας δείχνει ότι αν έχουμε μία καλή προσέγγιση (που σημαίνει το  $ID(Q//P_F)$  να είναι μικρό) τότε μπορούμε να βρούμε μία "καλή" προσέγγιση του  $Z$  με κάποιο "καλό" κάτω φράγμα.

Το γεγονός ότι αυτή η προσέγγιση είναι ένα κάτω φράγμα, παίζει σημαντικό ρόλο όσον αφορά τις παραμέτρους μάθησης (learning parameters) των γραφικών μοντέλων.

### 2.3.2 Ο ακριβής συμπερασμός ως βελτιστοποίηση

Πριν θεωρήσουμε τις μεθόδους προσεγγιστικού συμπερασμού, θα παρουσιάσουμε την χρήση μιας τροποποιημένης προσέγγισης για την εξαγωγή μιας διαδικασίας ακριβούς συμπερασμού (exact inference).

Οι ιδέες που θα παρουσιαστούν εδώ, θα μας χρειαστούν παρακάτω στην συζήτηση περί μεθόδων προσεγγιστικού συμπερασμού.

Ο σκοπός του ακριβούς συμπερασμού είναι να υπολογίσουμε τα περιθώρια (marginals) της κατανομής.

Για να το κάνουμε αυτό θα πρέπει να σιγουρευτούμε ότι το σύνολο των κατανομών  $Q$  είναι αρκετά εκφραστικό ώστε να μπορεί να αναπαραστήσει την κατανομή  $P_F$  που έχουμε ως στόχο.

Αντί να προσεγγίσουμε την  $P_F$ , η λύση του προβλήματος βελτιστοποίησης μετασχηματίζει την αναπαράσταση της κατανομής, από ένα γινόμενο παραγόντων σε μια πιο χρήσιμη μορφή  $Q$  η οποία μας δίνει απευθείας τα επιθυμητά περιθώρια.

Για να το επιτύχουμε αυτό, θα χρειαστεί να κάνουμε βελτιστοποίηση επί του συνόλου  $Q$  των κατανομών που περιλαμβάνει την  $P_F$ .

Κατόπιν, εάν κάνουμε μια αναζήτηση επί αυτού του (βελτιστοποιημένου) συνόλου, είναι σίγουρο ότι θα βρούμε μια κατανομή  $Q^*$  για την οποία θα ισχύει ότι  $ID(Q^*//P_F) = 0$ , που σημαίνει ότι πρόκειται για το μοναδικό οικουμενικό βέλτιστο του συναρτησιακού της ενέργειας που διαθέτουμε.

Θα αναπαραστήσουμε αυτό το σύνολο χρησιμοποιώντας έναν μη κατευθυνόμενο γραφικό μοντέλο που ονομάζεται "δέντρο της κλίκας" (clique tree) για λόγους που θα εξηγηθούν στην συνέχεια.

Θεωρούμε τον μη κατευθυνόμενο γράφο που αντιστοιχεί στο σύνολο των παραγόντων (factors)  $F$ .

Σε αυτόν τον γράφο, οι κόμβοι συνδέονται εάν εμφανίζονται μαζί σε έναν παράγοντα. Σημειώνουμε ότι, εάν ένας παράγοντας είναι η δεσμευμένη κατανομή ενός προσανατολισμένου γραφικού μοντέλου τότε η "οικογένεια" θα είναι μία κλίκα στον γράφο και έτσι η σύνδεση της (connectivity) είναι πιο πυκνή σε σχέση με τον αρχικό κατευθυνόμενο γράφο αφού οι γονείς είναι συνδεδεμένοι.

Η ιδιότητα "κλειδί" για ακριβή συμπερασμό στον γράφο είναι η χορδικότητα (chordality).

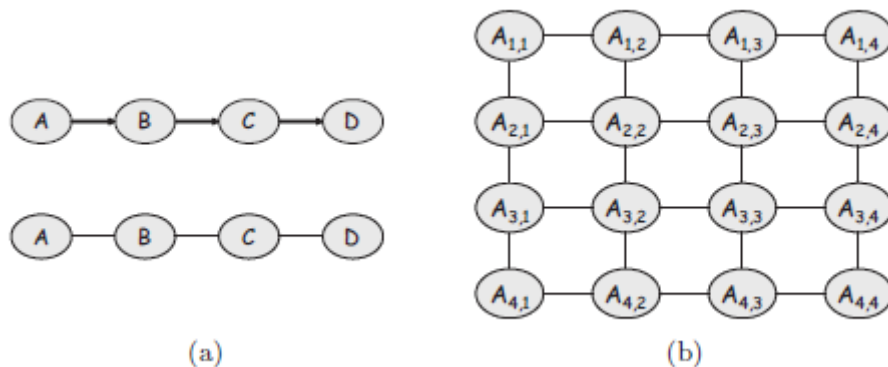
Έστω  $X_1 - X_2 - \dots - X_k - X_1$  ένας βρόχος στον γράφο.

Χορδή (chord), στον βρόχο, ονομάζουμε μια ακμή που συνδέει τους  $X_i$  και  $X_j$  όπου οι  $X_i, X_j$  είναι μη διαδοχικοί κόμβοι.

Ένας μη προσανατολισμένος γράφος  $H$  λέμε ότι είναι χορδικός (chordal) εάν κάθε βρόχος  $X_1 - X_2 - \dots - X_k - X_1$ , για  $k \geq 4$  έχει μία χορδή.

Με άλλα λόγια, ο μακρύτερος ελάχιστος βρόχος (longest minimal loop), δηλαδή αυτός που δεν περιέχει συντόμευση (shortcut), είναι ένα τρίγωνο.

Για τον λόγο αυτό, οι χορδικοί γράφοι συχνά ονομάζονται και διασταυρωμένοι (triangulated).



Σχήμα 2.4

Περί του σχήματος 2.4 :

Στο 2.4a βλέπουμε ένα Bayesian δίκτυο με αλυσιδωτή δομή και το ισοδύναμο δίκτυο Markov.

Στο 2.4b φαίνεται ένα δίκτυο Markov με δομή πλέγματος.

Οι πιο απλοί (και συχνότερα χρησιμοποιούμενοι) χορδικοί γράφοι έχουν αλυσιδωτή δομή (σχήμα 2.4a).

Τι γίνεται όμως εάν ο γράφος δεν είναι χορδικός?

Παράδειγμα μη χορδικών γράφων είναι οι γράφοι με δομή πλέγματος, οι οποίοι χρησιμοποιούνται συχνά στην όραση υπολογιστών για προβλήματα pixel-labeling (σχήμα 2.4b).

Για να κάνουμε έναν γράφο χορδικό (δηλαδή να διαθέτει διασταυρώσεις) προσθέτουμε συμπληρωματικές ακμές (fill-in edges) ώστε να δημιουργηθούν βραχυκυκλώματα μεταξύ των βρόχων.

Γενικώς υπάρχουν πολλοί τρόποι για να το επιτύχουμε αυτό και μάλιστα το να βρούμε τον ελάχιστο αριθμό ακμών που πρέπει να προστεθούν ώστε ο γράφος να γίνει χορδικός είναι ένα πρόβλημα δυσκολίας NP.

Υπάρχουν βέβαια και αρκετοί ευριστικοί αλγόριθμοι που μπορούν να δώσουν απαντήσεις στο συγκεκριμένο πρόβλημα.

Στην συνέχεια θα ορίσουμε τον "συμπλεγματοειδή γράφο" (cluster graph) ο οποίος αποτελεί την ραχοκοκαλιά για μια δομή γραφικών δεδομένων (graphical data structure) και που είναι απαραίτητος για την πραγματοποίηση συμπερασμού.

Κάθε κόμβος σε έναν γράφο είναι ένα σύμπλεγμα (cluster) που σχετίζεται με ένα υποσύνολο μεταβλητών.

Ένας τέτοιος γράφος διαθέτει μη προσανατολισμένες ακμές οι οποίες συνδέουν συμπλέγματα των οποίων οι εμβέλεις (scopes) έχουν κάποια μη κενή τομή (intersection).

Ένας γράφος  $K$  για ένα σύνολο παραγόντων  $F$ , επί ενός συνόλου  $X$ , είναι ένας μη προσανατολισμένος γράφος, κάθε κόμβος  $i$  του οποίου σχετίζεται με ένα υποσύνολο  $C_i \subseteq X$ .

Ένας συμπλεγματοειδής γράφος πρέπει να διατηρεί την οικογενειακότητα (family-preserving).

Αυτό σημαίνει ότι κάθε παράγοντας  $\varphi \in F$  πρέπει να σχετίζεται με ένα σύμπλεγμα  $C$ , γράφουμε  $\alpha(\varphi)$ , έτσι ώστε  $\text{Scope}[\varphi] \subseteq C_i$ .

Κάθε ακμή μεταξύ δυο συμπλεγμάτων  $C_i$  και  $C_j$  σχετίζεται με ένα sepset  $S_{i,j} = C_i \cap C_j$ .

Ένας κατ' ιδίαν συνδεδεμένος (singly connected) συμπλεγματοειδής γράφος (ένα δένδρο) ονομάζεται συμπλεγματοειδές δένδρο.

Έστω  $T$  ένα δένδρο επί ενός συνόλου παραγόντων  $F$ .

Λέμε ότι το  $T$  διαθέτει την ιδιότητα της "διατρέχουσας τομής" εάν όποτε υπάρχει μια μεταβλητή  $X$  τέτοια ώστε  $X \in C_i$  και  $X \in C_j$ , αυτή η μεταβλητή βρίσκεται επίσης και σε κάθε σύμπλεγμα εντός του (μοναδικού) μονοπατιού που συνδέει τα  $C_i$  και  $C_j$  εντός του  $T$ .

Ένα συμπλεγματοειδές δένδρο που ικανοποιεί την ιδιότητα της διατρέχουσας τομής ονομάζεται "δένδρο-κλίκα" (clique tree).

Κάθε χορδικός γράφος  $G$  διαθέτει ένα δένδρο-ομάδα (clique tree)  $T$ .

Η κατασκευή του δένδρου-ομάδα (clique tree) δοθέντος ενός χορδικού γράφου είναι σχετικά εύκολη υπόθεση. Απλώς ακολουθούμε τα παρακάτω βήματα :

1. Βρίσκουμε τις μεγαλύτερες δυνατές ομάδες (maximal cliques) του γράφου.
2. Τρέχουμε έναν αλγόριθμο μεγίστου εκτεινόμενου δένδρου (maximum spanning tree algorithm) στον κατάλληλο γράφο-ομάδα (clique graph).

Πιο συγκεκριμένα, κατασκευάζουμε έναν μη προσανατολισμένο γράφο του οποίου οι κόμβοι είναι οι μεγαλύτερες δυνατές κλίκες και όπου κάθε ζευγάρι κόμβων  $C_i, C_j$  συνδέεται με μια ακμή της οποίας το "βάρος" είναι  $|C_i \cap C_j|$ .

Εξαιτίας αυτής της αντιστοίχισης μπορούμε να ορίσουμε ένα πολύ σημαντικό χαρακτηριστικό για ένα γράφο, το οποίο είναι κρίσιμης σημασίας για την πολυπλοκότητα της πραγματοποίησης ακριβούς συμπερασμού.

Το εύρος δένδρου ενός χορδικού γράφου ισούται με το μέγεθος της μεγαλύτερης κλίκας μείον 1.

Το εύρος δένδρου ενός μη τριγωνοποιημένου γράφου είναι το ελάχιστο εύρος δένδρου από όλες τις τριγωνοποιήσεις του.

Να αναφέρουμε ως παράδειγμα ότι το εύρος δένδρου κάθε αλυσίδας στο σχήμα 2.4a είναι 1 και ότι το εύρος δένδρου του πλέγματος που φαίνεται στο σχήμα 2.4b είναι 4.

### 2.3.2.1 Το πρόβλημα της βελτιστοποίησης

Υποθέτουμε ότι μας δίνετε ένα δένδρο-ομάδα  $T$  για την  $P_F$ .

Αυτό σημαίνει ότι το  $T$  ικανοποιεί την ιδιότητα της διατρέχουσας τομής και την ιδιότητα διατήρησης της οικογενειακότητας.

Επιπλέον, υποθέτουμε ότι μας δίνετε ένα σύνολο δυναμικών

$Q = \{\pi_i\} \cup \{\mu_{i,j} : (C_i - C_j) \in T\}$ , όπου με  $C_i$  συμβολίζουμε τα συμπλέγματα στο  $T$ , με  $S_{i,j}$  συμβολίζουμε τους διαχωριστές κατά μήκος των συνόρων στο  $T$ , με  $\pi$  συμβολίζουμε κάποιο δυναμικό επί του  $C_i$  και με  $\mu_{i,j}$  συμβολίζουμε κάποιο δυναμικό επί των  $S_{i,j}$ .

Το σύνολο των δυναμικών ορίζει μια κατανομή  $Q$  σύμφωνα με το  $T$  που δίνεται από την σχέση

$$Q(\mathcal{X}) = \frac{\prod_{C_i \in T} \pi_i}{\prod_{(C_i - C_j) \in T} \mu_{i,j}}.$$

(2.4)

Να σημειωθεί ότι εκ κατασκευής, το  $Q$  μπορεί να αναπαραστήσει την  $P_F$  απλώς θέτοντας τα κατάλληλα δυναμικά  $\pi_i$  ίσα με τους παράγοντες  $\phi_i$  και το  $\mu_{i,j}$  ίσο με 1.

Εμείς όμως θα θεωρήσουμε μια διαφορετική και πιο χρήσιμη αναπαράσταση.

Το σύνολο των δυναμικών  $Q$  λέμε ότι είναι σταθμισμένο (calibrated) όταν για κάθε  $(C_i - C_j) \in T$  το δυναμικό  $\mu_{i,j}$  επί του  $S_{i,j}$  είναι το περιθώριο (marginal) του  $\pi_i$  (και του  $\pi_j$ ).

Έστω  $Q$  ένα σύνολο σταθμισμένων δυναμικών για το  $T$  και  $Q$  κατανομή που ορίζεται από την εξίσωση (2.4).

Τότε  $\pi_i[c_i] = Q(c_i)$  και  $\mu_{i,j}[s_{i,j}] = Q(s_{i,j})$ .

Με άλλα λόγια, τα δυναμικά αντιστοιχούν σε περιθώρια (marginals) της κατανομής  $Q$  που ορίζεται από την εξίσωση (2.4).

Τώρα, εάν το  $Q$  είναι ένα σύνολο μη σταθμισμένων δυναμικών για το  $T$  και  $Q$  είναι η κατανομή που ορίζεται από την (2.4) μπορούμε να κατασκευάσουμε το  $Q'$ , ένα σύνολο σταθμισμένων δυναμικών που αναπαριστούν την  $Q$ , χρησιμοποιώντας απλώς τα κατάλληλα περιθώρια (marginals) της  $Q$ .

Από την στιγμή που θα επικεντρώσουμε την προσοχή μας σε σταθμισμένα δέντρα-κλικών (calibrated clique trees), μπορούμε να ξαναγράψουμε το συναρτησιακό της ενέργειας σε μια παραγοντοποιημένη μορφή, ως ένα άθροισμα όρων κάθε ένας από τους οποίους εξαρτάται απευθείας μόνο από ένα από τα δυναμικά στο  $Q$ .

Αυτή η μορφή καταδεικνύει την δομή εντός της κατανομής και για αυτό αποτελεί ένα καλύτερο σημείο αφετηρίας για περαιτέρω ανάλυση.

Όπως θα δούμε, αυτή η μορφή θα είναι επίσης και η βάση για τις προσεγγίσεις μας σε επόμενες ενότητες.

Δοθέντος ενός δέντρου –ομάδας  $T$  με ένα σύνολο δυναμικών  $Q$  και μια ανάθεση  $\alpha$  που απεικονίζει τους παράγοντες της  $P_F$  σε συμπλέγματα στο  $T$ , ορίζουμε το παραγοντοποιημένο συναρτησιακό της ενέργειας

$$\tilde{F}[P_F, Q] = \sum_i E_{\pi_i}[\ln \pi_i^0] + \sum_{C_i \in T} H_{\pi_i}(C_i) - \sum_{(C_i - C_j) \in T} H_{\mu_{i,j}}(S_{i,j}) \quad (2.5)$$

όπου

$$\pi_i^0 = \prod_{\phi, \alpha(\phi)=i} \phi.$$

Πριν αποδείξουμε ότι το συναρτησιακό της ενέργειας είναι ισοδύναμο με την παραγοντοποιημένη του μορφή, ας κατανοήσουμε την μορφή του.

Ο πρώτος όρος είναι ένα άθροισμα όρων της μορφής

$$E_{\pi_i}[\ln \pi_i^0].$$

Να υπενθυμίσουμε ότι το  $\pi_i^0$  είναι ένας παράγοντας (όχι απαραίτητα μία κατανομή) επί της εμβέλειας  $C_i$  που είναι μια συνάρτηση με πεδίο ορισμού το  $\text{Val}(C_i)$  και πεδίο τιμών το  $\mathbb{R}^+$ .

Ο λογάριθμος του επομένως είναι μια συνάρτηση με πεδίο ορισμού το  $\text{Val}(C_i)$  και πεδίο τιμών το  $\mathbb{R}$ .

Το δυναμικό ομάδας (clique potential)  $\pi_i$  είναι μία κατανομή επί του  $\text{Val}(C_i)$ .

Μπορούμε επομένως να υπολογίσουμε την προσδοκία (expectation)

$$\sum_{c_i} \pi_i[c_i] \ln \pi_i^0.$$

Οι τελευταίοι δύο όροι είναι εντροπίες των κατανομών που σχετίζονται με συμπλέγματα και sepsets στο δένδρο.

Αν  $Q$  είναι ένα σύνολο σταθμισμένων δυναμικών για το  $T$  και η  $Q$  ορίζεται από την (2.4) τότε

$$\tilde{F}[P_F, Q] = F[P_F, Q].$$

Χρησιμοποιώντας αυτήν την μορφή της ενέργειας, μπορούμε πλέον να ορίσουμε το πρόβλημα της βελτιστοποίησης.

Πρώτα όμως πρέπει να ορίσουμε τον χώρο (space) επί του οποίου θα γίνει η βελτιστοποίηση.

Αν η  $Q$  είναι παραγοντοποιημένη σύμφωνα με το  $T$ , τότε μπορούμε να την αναπαραστήσουμε με ένα σύνολο σταθμισμένων δυναμικών.

Η στάθμιση είναι ένας περιορισμός για τα δυναμικά, καθώς ένα δέντρο-ομάδα είναι σταθμισμένο αν τα γειτονικά δυναμικά συμφωνούν (agree) επί της περιθώριας κατανομής στο συζευγμένο υποσύνολο (joint subset) τους.

Ως εκ τούτου, παραθέτουμε την ακόλουθη περιορισμένη διαδικασία βελτιστοποίησης (constrained optimization procedure):

CTree-Optimize

Find  $Q$   
that maximize  $\bar{F}[P_F, Q]$

subject to  $\sum_{C_i \in S_{i,j}} \pi_i = \mu_{i,j}, \quad \forall (C_i - C_j) \in T; \quad (2.6)$

$\sum_{C_i} \pi_i = 1, \quad \forall C_i \in T. \quad (2.7)$

Οι περιορισμοί (2.6) και (2.7) εξασφαλίζουν ότι τα δυναμικά στο  $Q$  είναι σταθμισμένα και αναπαριστούν επιτρεπτές κατανομές.

Μπορεί να αποδειχθεί ότι η αντικειμενική συνάρτηση είναι αυστηρώς κοίλη (concave) για τις μεταβλητές  $\pi, \mu$ .

Οι περιορισμοί ορίζουν ένα κυρτό σύνολο (convex set) (γραμμικός υποχώρος), οπότε το συγκεκριμένο πρόβλημα βελτιστοποίησης έχει μοναδικό μέγιστο.

Εφόσον το  $Q$  μπορεί να αναπαραστήσει την  $P_F$ , το εν λόγω μέγιστο παρουσιάζεται όταν  $ID(Q||P_F) = 0$ .

### 2.3.2.2 Χαρακτηρισμός σταθερού σημείου

Μπορούμε τώρα να αποδείξουμε ότι τα στάσιμα σημεία (stationary points) αυτής της περιορισμένης συνάρτησης βελτιστοποίησης, δηλαδή τα σημεία στα οποία η κλίση (gradient) είναι ορθογώνια ως προς όλους τους περιορισμούς, μπορούν να χαρακτηριστούν από ένα σύνολο αυτό-συνεπών (self-consistent) εξισώσεων.

Να υπενθυμίσουμε ότι ένα στάσιμο σημείο μιας συνάρτησης είναι είτε ένα τοπικό μέγιστο, ένα τοπικό ελάχιστο ή ένα σαγματικό σημείο.

Στο συγκεκριμένο πρόβλημα βελτιστοποίησης υπάρχει μόνο ένα ολικό μέγιστο.

Αν και δεν θα το κάνουμε εδώ, είναι δυνατόν να αποδειχθεί ότι είναι και το μοναδικό στάσιμο σημείο.

Για τον λόγο αυτό μπορούμε να ορίσουμε το ολικό βέλτιστο δηλωτικά, ως ένα σύνολο εξισώσεων, χρησιμοποιώντας τις συνήθεις μεθόδους που βασίζονται στους πολλαπλασιαστές Lagrange.

Όπως θα δείξουμε στα επόμενα, αυτός ο δηλωτικός φορμαλισμός οδηγεί σε ένα σύνολο εξισώσεων το οποίο αντιστοιχεί σε βήματα μετάδοσης μηνυμάτων (message-passing steps) μέσα στο δέντρο-ομάδα (clique tree), που είναι μία καθιερωμένη διαδικασία συμπερασμού η οποία συνήθως προκύπτει από τον δυναμικό προγραμματισμό.

Ένα σύνολο δυναμικών  $Q$  είναι ένα στάσιμο σημείο της διαδικασίας CTree-Optimize αν και μόνο αν υπάρχει ένα σύνολο παραγόντων

$$\{\delta_{i \rightarrow j}[S_{i,j}] : C_i - C_j \in T\}$$

τέτοιο ώστε

$$\delta_{i \rightarrow j} \propto \sum_{C_i - S_{i,j}} \pi_i^0 \left( \prod_{k \in N_{C_i} - \{j\}} \delta_{k \rightarrow i} \right) \quad (2.8)$$

$$\pi_i \propto \pi_i^0 \left( \prod_{j \in N_{C_i}} \delta_{j \rightarrow i} \right) \quad (2.9)$$

$$\mu_{i,j} = \delta_{j \rightarrow i} \times \delta_{i \rightarrow j}, \quad (2.10)$$

Όπου  $N_{C_i}$  είναι οι γειτονικές ομάδες (cliques) του  $C_i$  στο  $T$ .

Το παραπάνω παρουσιάζει ζητήματα που εμφανίζονται σε πολλές προσεγγίσεις οι οποίες μετατρέπουν προβλήματα μεταβλητότητας σε σχήματα μετάδοσης μηνυμάτων.

Μας παρέχει ένα χαρακτηρισμό της λύσης του προβλήματος βελτιστοποίησης με όρους εξισώσεων σταθερών σημείων (fixed-point equations) οι οποίες πρέπει να επαληθεύονται όταν βρούμε ένα μέγιστο  $Q$ .

Αυτές οι εξισώσεις σταθερών σημείων ορίζουν τις σχέσεις που πρέπει να ισχύουν μεταξύ των διαφορετικών παραμέτρων που υπεισέρχονται στο πρόβλημα βελτιστοποίησης.

Ιδιαίτερος, η (2.8) ορίζει κάθε  $\delta_{i \rightarrow j}$  υπό όρους  $\delta_{k \rightarrow j}$  και όχι υπό όρους  $\delta_{i \rightarrow j}$ .

Οι υπόλοιποι παράμετροι ορίζονται όλες με ένα μη κυκλικό τρόπο υπό όρους  $\delta_{i \rightarrow j}$ .

Η μορφή των εξισώσεων που προκύπτουν από αυτό το θεώρημα μας φανερώνουν μια επαναληπτική διαδικασία για την εύρεση ενός σταθερού σημείου, κατά την οποία βλέπουμε τις εξισώσεις ως αναθέσεις και εφαρμόζουμε επαναληπτικώς τις εξισώσεις στις τρέχουσες τιμές της αριστερής πλευράς προκειμένου να ορίσουμε μια νέα τιμή για την δεξιά πλευρά.

Αρχικοποιούμε όλα τα  $\delta_{i \rightarrow j}$  ώστε να είναι ίσα με 1 και μετά εφαρμόζουμε επαναληπτικά την (2.8), υπολογίζοντας τα  $\delta_{i \rightarrow j}$  κάθε ισότητας στην αριστερή πλευρά υπό όρους της δεξιάς πλευράς (μετατρέποντας ουσιαστικά κάθε σημείο ισότητας σε ανάθεση).

Προφανώς, μια μόνο επανάληψη αυτής της διαδικασίας συνήθως δεν επαρκεί ώστε να επαληθεύονται οι εξισώσεις, παρ' όλα αυτά όμως, υπό συγκεκριμένες συνθήκες



(οι οποίες ισχύουν για την συγκεκριμένη περίπτωση) μπορούμε να εγγυηθούμε ότι αυτή η διαδικασία συγκλίνει σε μία λύση η οποία ικανοποιεί όλες τις εξισώσεις (2.8) οπότε οι άλλες εξισώσεις μπορούν πλέον να επαληθευθούν εύκολα.

### 2.3.3 Αναδραστική διάδοση πίστεως σε δίκτυα Markov με συζεύξεις

Επικεντρωνόμαστε στην κλάση των δικτύων Markov με συζεύξεις (pairwise Markov networks).

Σε αυτά τα δίκτυα, έχουμε ένα μονοπαραγοντικό δυναμικό  $\phi_i[X_i]$  επί κάθε μεταβλητής  $X_i$  και επιπλέον ένα δυναμικό ζεύξης  $\phi_{(i,j)}[X_i, X_j]$  επί κάποιων ζευγών μεταβλητών.

Αυτά τα δυναμικά σύζευξης αντιστοιχούν στις ακμές του δικτύου Markov.

Στα παραδείγματα τέτοιων δικτύων περιλαμβάνεται και το παράδειγμα του σχήματος 2.3 για την φυματίωση καθώς και τα δίκτυα με πλέγματα που συζητήσαμε στα προηγούμενα.

Η μετατροπή ενός τέτοιου δικτύου σε συμπλεγματοκό γράφο είναι αρκετά απλή.

Για κάθε δυναμικό, εισάγουμε έναν αντίστοιχο σύμπλεγμα και θέτουμε ακμές μεταξύ συμπλεγμάτων που έχουν επικαλυπτόμενη εμβέλεια.

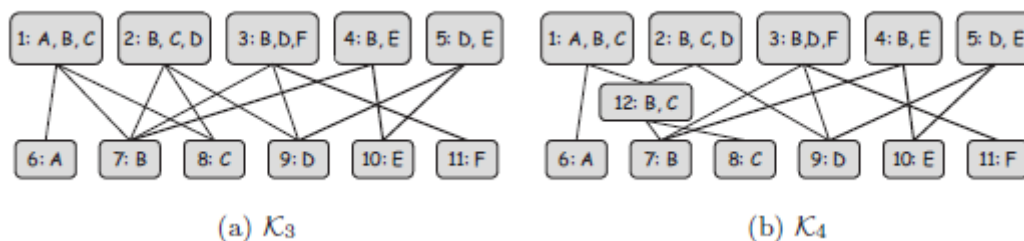
Με άλλα λόγια, υπάρχει μια ακμή μεταξύ του συμπλέγματος  $C_{(i,j)}$  που αντιστοιχεί στην ακμή  $X_i - X_j$  και στα συμπλέγματα  $C_i$  και  $C_j$  που αντιστοιχούν στους μονοπαραγοντικούς παράγοντες (univariate factors) επί των  $X_i$  και  $X_j$ .

Καθώς υπάρχει μια άμεση αντιστοιχία μεταξύ των συμπλεγμάτων στους συμπλεγματοκό γράφους και των μεταβλητών ή των ακμών στο αρχικό δίκτυο Markov, συχνά μας συμφέρει να σκεπτόμαστε τα βήματα διάδοσης (propagation steps) ως πράξεις πάνω στο αρχικό δίκτυο.

Επιπλέον, καθώς κάθε σύμπλεγμα ζευγών (pairwise cluster) έχει μόνο δύο γείτονες, θεωρούμε δύο βήματα διάδοσης κατά το μονοπάτι  $C_i - C_{(i,j)} - C_j$  ως διάδοση πληροφορίας μεταξύ των  $X_i$  και  $X_j$ .

Πράγματι, παλαιότερες εκδοχές γενικευμένης διάδοσης πίστεως δηλώνονταν υπό αυτούς τους όρους.

Αυτός ο αλγόριθμος είναι γνωστός ως "αναδραστική διάδοση πίστεως" (loopy belief propagation) καθώς χρησιμοποιεί βήματα διάδοσης που χρησιμοποιούνται από αλγορίθμους για δέντρα Markov με την διαφορά ότι εφαρμόστηκε σε δίκτυα με βρόχους (loops).



Σχήμα 2.5

Μια φυσιολογική ερώτηση είναι πώς να επεκτείνουμε αυτήν την μέθοδο σε δίκτυα που είναι πιο περίπλοκα από τα συζευγμένα δίκτυα Markov.



Όταν έχουμε μεγαλύτερα δυναμικά, μπορεί να επικαλύπτονται με τέτοιο τρόπο που να έχει ως αποτέλεσμα περίπλοκες αλληλεπιδράσεις μεταξύ τους.

Μια απλή κατασκευή δημιουργεί έναν διμερή (bipartite) γράφο

Το πρώτο στρώμα (layer) αποτελείται από "μεγάλα" συμπλέγματα, με ένα σύμπλεγμα για κάθε παράγοντα  $\phi$  στο  $F$ , του οποίου η εμβέλεια είναι  $\text{Scope}[\phi]$ .

Αυτά τα συμπλέγματα σιγουρεύουν ότι ικανοποιούμε την ιδιότητα διατήρησης της οικογενειακότητας.

Το δεύτερο στρώμα αποτελείται από "μικρά" μονοπαραγοντικά συμπλέγματα, ένα για κάθε μια τυχαία μεταβλητή.

Τέλος, τοποθετούμε μια ακμή μεταξύ κάθε μονοπαραγοντικού συμπλέγματος  $X$  του δευτέρου στρώματος και κάθε συμπλέγματος του πρώτου στρώματος που περιλαμβάνει τον  $X$ , η εμβέλεια (scope) αυτής της ακμής είναι ο ίδιος ο  $X$ .

Αυτό μπορούμε να το δούμε ως παράδειγμα στο σχήμα 2.5a .

Μπορούμε εύκολα να επιβεβαιώσουμε ότι είναι ένας κανονικός (proper) συμπλεγματολογικός γράφος.

Καταρχήν, εκ κατασκευής ικανοποιεί της ιδιότητα διατήρησης της οικογενειακότητας.

Κατά δεύτερον, οι ακμές που "δείχνουν" μια μεταβλητή  $X$  σχηματίζουν έναν υπογράφο σε σχήμα αστεριού με ακμές από το μονοπαραγοντικό σύμπλεγμα με εμβέλεια  $X$  προς όλα τα άλλα μεγάλα συμπλέγματα που περιέχουν το  $X$ .

Ονομάζουμε αυτή την κατασκευή "προσέγγιση Bethe" (Bethe approximation), για λόγους που θα φανούν παρακάτω.

Η κατασκευή αυτού του συμπλεγματολογικού γράφου είναι απλή και μπορεί να αυτοματοποιηθεί χωρίς κάποια ιδιαίτερη δυσκολία.

Μέχρι στιγμής, η συζήτηση μας περί διάδοσης πίστεως (belief propagation) υπήρξε εντελώς διαδικαστική (procedural) και είχε ως κίνητρο την ομοιότητα της με αλγορίθμους μετάδοσης μηνυμάτων (message-passing algorithms) για συμπλεγματολογικά δέντρα.

Υπάρχει όμως κάποια τυπική αιτιολόγηση για μια τέτοιου είδους προσέγγιση?

Υπάρχει κάποιος τρόπος να δούμε ότι αυτός ο αλγόριθμος μας παρέχει μια προσέγγιση για ακριβή συμπερασμό?

Σε αυτό το σημείο, θα δείξουμε ότι η διάδοση πίστεως μπορεί να αιτιολογηθεί χρησιμοποιώντας τον φορμαλισμό του συναρτησιακού της ενέργειας.

Συγκεκριμένα, τα μηνύματα που μεταδίδονται με γενικευμένη διάδοση πίστεως μπορούν να προκύψουν από τις εξισώσεις σταθερού σημείου για τα στάσιμα σημεία μίας προσεγγιστικής εκδοχής του συναρτησιακού της ενέργειας της μορφής (2.3).

Όπως θα δούμε, αυτός ο φορμαλισμός μας παρέχει σημαντική δυνατότητα κατανόησης για τον γενικευμένο αλγόριθμο διάδοσης πίστεως.

Μας επιτρέπει να κατανοήσουμε καλύτερα τις ιδιότητες σύγκλισης της γενικευμένης διάδοσης πίστεως καθώς και να χαρακτηρίσουμε τα σημεία σύγκλισης.

Επιπλέον μας δείχνει γενικεύσεις του αλγορίθμου οι οποίες έχουν καλύτερες ιδιότητες σύγκλισης ή που βελτιστοποιούν κάποια πιο ακριβή προσέγγιση του συναρτησιακού της ενέργειας.

Ο τρόπος κατασκευής θα είναι όμοιος με αυτόν της παραγράφου 2.3.2 που αφορούσε τον ακριβή συμπερασμό. Υπάρχουν όμως και κάποιες διαφορές.

Όπως είδαμε, ο σταθμισμένος συμπλεγματολογικός γράφος διατηρεί την πληροφορία που έχει μέσα της η  $P_F$ .

Όμως τα προκύπτοντα δυναμικά των συμπλεγμάτων δεν είναι ,εν γένει, τα περιθώρια (marginals) της  $P_F$ .

Για την ακρίβεια αυτά τα δυναμικά των συμπλεγμάτων ίσως να μην αναπαριστούν τα περιθώρια ουδεμίας συνεκτικής από κοινού κατανομής επί του συνόλου  $X$ .

Οπότε, μπορούμε να σκεφτούμε την γενικευμένη διάδοση πίστεως σαν την κατασκευή ενός συνόλου ψευδοπεριθωρίων κατανομών (pseudo-marginal distributions), όπου κάθε μία είναι επί των μεταβλητών ενός συμπλέγματος.

Αυτά τα ψευδοπεριθώρια είναι σταθμισμένα και γι' αυτό είναι και τοπικώς συνεπή (locally consistent) μεταξύ τους, αυτό δεν σημαίνει απαραίτητα όμως ότι πρόκειται για περιθώρια (marginals) κάποιας μεμονωμένης κύριας από κοινού κατανομής.

Το συναρτησιακό της ενέργειας  $F[P_F, Q]$  περιέχει όρους που έχουν να κάνουν με την εντροπία ολόκληρης κάποιας από κοινού κατανομής, συνεπώς δεν μπορεί να χρησιμοποιηθεί για την εκτίμηση της ποιότητας κάποιας προσέγγισης η οποία ορίζεται με όρους (ενδεχομένως μη συνεκτικών) ψευδοπεριθωρίων.

Όμως, το παραγοντοποιημένο συναρτησιακό της ελεύθερης ενέργειας (free energy functional)  $\tilde{F}[P_F, Q]$  ορίζεται υπό όρους εντροπίας συμπλεγμάτων και μηνυμάτων και γι' αυτό είναι καλώς ορισμένη (well-defined) για τα ψευδοπεριθώρια  $Q$ .

Οπότε μπορούμε να γράψουμε ένα πρόβλημα βελτιστοποίησης όπως ακριβώς κάναμε και σε προηγούμενη παράγραφο :

**CGraph-Optimize**

**Find**  $Q$   
**that maximize**  $\tilde{F}[P_F, Q]$

$$\text{subject to} \quad \sum_{C_i \in S_{i,j}} \pi_i = \mu_{i,j}, \quad \forall (C_i - C_j) \in T; \quad (2.11)$$

$$\sum_{C_i} \pi_i = 1, \quad \forall C_i \in T. \quad (2.12)$$

Το σημαντικό εδώ είναι ότι ,σε αντίθεση με αυτό που ισχύει για τα δέντρα-ομάδες, η  $\tilde{F}[P_F, Q]$  δεν είναι πλέον απλώς μια αναδιατυπωμένη μορφή της ελεύθερης ενέργειας αλλά περισσότερο μία προσέγγιση αυτής.

Οπότε, το πρόβλημα βελτιστοποίησης που έχουμε αυτή την στιγμή περιέχει δύο προσεγγίσεις :

Χρησιμοποιούμε μια προσέγγιση ,και όχι το ακριβές, του συναρτησιακού της ενέργειας και το βελτιστοποιούμε επί του χώρου των ψευδοπεριθωρίων, ο οποίος είναι μία χαλάρωση (ένας υπερχώρος) του χώρου όλων των συνεκτικών πιθανοτικών κατανομών που παραγοντοποιούν επί του συμπλεγματοκού γράφου.

Το κατά προσέγγιση συναρτησιακό της ενέργειας σε αυτή την περίπτωση είναι μία περιορισμένη μορφή (restricted form) μίας προσέγγισης που είναι γνωστή στην στατιστική φυσική ως "ελεύθερη ενέργεια Kikuchi" (Kikuchi free energy).

Είδαμε ότι το συναρτησιακό της ενέργειας είναι ένα κάτω φράγμα της λογαριθμικής συνάρτησης διαμερισμού, οπότε μεγιστοποιώντας την λαμβάνουμε καλύτερες προσεγγίσεις για την  $P_F$ .

Δυστυχώς όμως, το παραγοντοποιημένο συναρτησιακό της ενέργειας, που είναι μόνο μία προσέγγιση του πραγματικού συναρτησιακού της ενέργειας, δεν αποτελεί απαραίτητα ένα κάτω φράγμα.

Παρ' όλα αυτά όμως, πρόκειται για μια καλή στρατηγική προκειμένου να μεγιστοποιήσουμε το κατά προσέγγιση συναρτησιακό της ενέργειας.

Αυτό το πρόβλημα μεγιστοποίησης είναι το φυσικό ανάλογο του CTree-Optimize για την περίπτωση των συμπλεγματοκών γράφων.

Ένα σύνολο δυναμικών  $Q$  είναι ένα στάσιμο σημείο του CGraph-Optimize αν και μόνο αν για κάθε ακμή  $(C_i - C_j) \in K$  υπάρχουν βοηθητικά δυναμικά  $\delta_{i \rightarrow j}(S_{i,j})$  και  $\delta_{j \rightarrow i}(S_{j,i})$  τέτοια ώστε

$$\delta_{i \rightarrow j} \propto \sum_{C_i - S_{i,j}} \pi_i^0 \times \prod_{k \in N_{C_i} - \{j\}} \delta_{k \rightarrow i} \quad (2.13)$$

$$\pi_i \propto \pi_i^0 \times \prod_{j \in N_{C_i}} \delta_{j \rightarrow i} \quad (2.14)$$

$$\mu_{i,j} = \delta_{j \rightarrow i} \times \delta_{i \rightarrow j}. \quad (2.15)$$

Αυτό το θεώρημα μας δείχνει ότι μπορούμε να χαρακτηρίσουμε σημεία σύγκλισης του συναρτησιακού της ενέργειας υπό όρους αρχικών δυναμικών και μηνυμάτων μεταξύ των συμπλεγμάτων.

Μπορούμε και πάλι να ορίσουμε μια διαδικαστική παραλλαγή στην οποία να αρχικοποιήσουμε τα  $\delta_{i \rightarrow j}$  και μετά να χρησιμοποιήσουμε επαναληπτικά την (2.3) προκειμένου να επαναπροσδιορίσουμε κάθε  $\delta_{i \rightarrow j}$  υπό όρους των τρεχόντων τιμών άλλων  $\delta_{k \rightarrow i}$ .

Το θεώρημα 2.28 μας δείχνει ότι τα σημεία σύγκλισης αυτής της διαδικασίας σχετίζονται με τα στάσιμα σημεία της  $F^{\sim} [P_F, Q]$ .

Είναι σχετικά εύκολο να επαληθεύσουμε ότι η  $F^{\sim} [P_F, Q]$  είναι άνω φραγμένη και γι' αυτό αυτή η συνάρτηση πρέπει να έχει κάποιο μέγιστο.

Υπάρχουν δύο περιπτώσεις :

Το μέγιστο να είναι ένα εσωτερικό σημείο ή ένα σημείο φραγμού (μερικές από τις πιθανότητες στο  $Q$  είναι 0).

Στην πρώτη περίπτωση το μέγιστο είναι επίσης και σταθερό σημείο, που σημαίνει ότι ικανοποιεί την συνθήκη του θεωρήματος 2.28.

Στην δεύτερη περίπτωση το μέγιστο δεν είναι απαραίτητα και στάσιμο σημείο.

Αυτή η κατάσταση όμως απαντάται πολύ σπάνια στην πράξη και είναι σίγουρο ότι δεν θα εμφανιστεί εάν κάνουμε κάποιες ήπιες υποθέσεις.

Είναι σημαντικό να καταλάβουμε τι ακριβώς σημαίνουν και τι δεν σημαίνουν αυτά τα συμπεράσματα.

Αυτά τα αποτελέσματα λοιπόν, σημαίνουν μονάχα ότι τα σημεία σύγκλισης της γενικευμένης διάδοσης πίστωσης είναι στάσιμα σημεία του συναρτησιακού της ελεύθερης ενέργειας.

Δεν σημαίνουν ότι μπορούμε να φτάσουμε σε αυτά τα σημεία σύγκλισης εφαρμόζοντας τα βήματα της διάδοσης πίστωσης.

Για την ακρίβεια, κανείς δεν μας εγγυάται ότι τα βήματα μετάδοσης μηνύματος της γενικευμένης διάδοσης πίστωσης απαραίτητως βελτιώνουν την ελεύθερη ενέργεια (που είναι και ο σκοπός μας) :

Ένα βήμα μετάδοσης μηνύματος μπορεί να αυξήσει ή να ελαττώσει το συναρτησιακό της ενέργειας.

Για την ακρίβεια, εάν η γενικευμένη διάδοση πίστωσης βελτιώνει μονότονα το συναρτησιακό τότε θα συνέκλινε πάντοτε απαραίτητως.

Ποιες είναι όμως οι συνέπειες αυτού του συμπεράσματος?

Καταρχήν, μας παρέχει μια δηλωτική σημασιολογία για την γενικευμένη διάδοση πίστεως υπό όρους βελτιστοποίησης ενός στοχευμένου συναρτησιακού.

Αυτή η δηλωτική σημασιολογία ανοίγει τον δρόμο για την διερεύνηση άλλων υπολογιστικών προσεγγίσεων για την βελτιστοποίηση του ίδιου συναρτησιακού.

Θα συζητήσουμε μερικές από αυτές τις προσεγγίσεις παρακάτω.

Αυτό το αποτέλεσμα μας επιτρέπει επίσης να καταλάβουμε ποιες ιδιότητες είναι σημαντικές για αυτού του τύπου την προσέγγιση και επομένως να σχεδιάσουμε άλλες προσεγγίσεις που ενδεχομένως να είναι πιο ακριβείς ή και ακόμα και καλύτερες κατά κάποιο τρόπο.

Ως παράδειγμα, να θυμηθούμε ότι κατά την κουβέντα μας περί γενικευμένων συμπλεγματικών γράφων απαιτήσαμε να ισχύει η ιδιότητα της διατρέχουσας τομής.

Αυτή η ιδιότητα έχει δύο συνέπειες :

Η πρώτη είναι ότι το σύνολο των συμπλεγμάτων που περιέχουν κάποια μεταβλητή  $X$  συνδέονται, συνεπώς το περιθώριο επί της  $X$  θα είναι το ίδιο σε όλα αυτά τα συμπλέγματα στο σημείο στάθμισης.

Η δεύτερη είναι ότι δεν υπάρχει κύκλος με sepsets τα οποία να περιέχουν την  $X$  (sepset είναι ένα σύνολο από μεταβλητές που βρίσκονται μεταξύ δύο υπογράφων).

Μπορούμε να κάνουμε αυτή την υπόθεση βλέποντας ότι μας αποτρέπει από το να επιτρέψουμε πληροφορία σχετικά με την  $X$  να ανακυκλώνεται ατέρμονα μέσω ενός βρόχου.

Η ανάλυση ελεύθερης ενέργειας μας παρέχει μια πιο τυπική (formal) αιτιολόγηση.

Για να την καταλάβουμε, θεωρούμε πρώτα την παραγοντοποιημένη μορφή του συναρτησιακού της ελεύθερης ενέργειας όταν ο συμπλεγματικός γράφος  $K$  έχει την μορφή της προσέγγισης Bethe.

Να υπενθυμίσουμε ότι στον γράφο με προσέγγιση Bethe υπάρχουν δύο στρώματα (layers) :

Το ένα αποτελείται από συμπλέγματα που αντιστοιχούν σε παράγοντες στο  $F$  και το άλλο αποτελείται από μονοπαραγοντικά συμπλέγματα.

Όταν ο συμπλεγματικός γράφος σταθμίζεται, αυτά τα μονοπαραγοντικά συμπλέγματα έχουν την ίδια κατανομή με τους διαχωριστές (separators) που βρίσκονται ανάμεσα τους καθώς και με τους παράγοντες του πρώτου στρώματος.

Οπότε μπορούμε να συνδυάσουμε μαζί τους όρους της εντροπίας για όλους τους διαχωριστές που είναι σημασμένοι με την  $X$  και το σχετικό μονοπαραγοντικό σύμπλεγμα και να ξαναγράψουμε την ελεύθερη ενέργεια ως εξής :

Εάν  $Q = \{\pi_\phi : \phi \in F\} \cup \{\pi_i(X_i)\}$  είναι ένα σταθμισμένο σύνολο δυναμικών για το  $K$  για έναν συμπλεγματικό γράφο με προσέγγιση Bethe ο οποίος έχει συμπλέγματα  $\{C_\phi : \phi \in F\} \cup \{X_i : X_i \in X\}$ , τότε

$$\tilde{F}[P_F, Q] = \sum_{\phi \in F} E_{\pi_\phi}[\ln \phi] + \sum_{\phi \in F} H_{\pi_\phi}(C_\phi) - \sum_i (d_i - 1) H_{\pi_i}(X_i), \quad (2.16)$$

Όπου  $d_i = |\{\phi : X_i \in \text{Scope}[\phi]\}|$  είναι ο αριθμός των παραγόντων που περιέχουν την  $X_i$ .

Να σημειωθεί ότι η (2.16) είναι ισοδύναμη με την παραγοντοποιημένη ελεύθερη ενέργεια μόνο όταν το  $Q$  είναι σταθμισμένο.

Καθώς όμως ενδιαφερόμαστε μόνο για τέτοιες περιπτώσεις, μπορούμε να περνάμε άνετα από την μια μορφή στην άλλη προκειμένου να βρούμε σταθερά σημεία (fixed points) του παραγοντοποιημένου συναρτησιακού της ελεύθερης ενέργειας.

Η εξίσωση (2.16) είναι γνωστή ως "ελεύθερη ενέργεια Bethe" (Bethe free energy) και προέρχεται από την στατιστική μηχανική.

Η προσέγγιση Bethe που συζητήσαμε στα προηγούμενα είναι μια κατασκευή υπό όρους συμπλεγματού γράφου ο οποίος σχεδιάστηκε ώστε να ταιριάζει με την ελεύθερη ενέργεια Bethe.

Όπως μπορούμε να δούμε σε αυτήν την εναλλακτική μορφή, εάν η μεταβλητή  $X_i$  εμφανίζεται σε  $d_i$  συμπλέγματα στον συμπλεγματικό γράφο, τότε εμφανίζεται σε κάποιο όρο εντροπίας με θετικό πρόσημο ακριβώς  $d_i$  φορές.

Λόγω της ιδιότητας διατρέχουσας τομής, ο αριθμός των διαχωριστών που περιέχουν την  $X_i$  είναι  $d_i - 1$  (ο αριθμός των ακμών σε ένα δέντρο με  $k$  κόμβους είναι  $k - 1$ ), έτσι ώστε η  $X_i$  να εμφανίζεται σε έναν όρο εντροπίας με αρνητικό πρόσημο ακριβώς  $d_i - 1$  φορές.

Σε αυτή την περίπτωση, λέμε ότι το counting number της  $X_i$  είναι 1.

Έτσι, η προσέγγιση μας δεν υπερτιμά ή υποτιμά (overcount or undercount) την εντροπία της  $X_i$ .

Δεν είναι δύσκολο να δείξουμε ότι το αποτέλεσμα του counting number ισχύει για οποιαδήποτε προσέγγιση η οποία ικανοποιεί την ιδιότητα της διατρέχουσας τομής.

Ο λόγος επομένως που κάνει ελκυστική την ιδιότητα της διατρέχουσας τομής είναι ότι οι συμπλεγματικοί γράφοι που την ικανοποιούν μας παρέχουν μια καλύτερη προσέγγιση του συναρτησιακού της ενέργειας.

Αυτή η διαισθητική αντίληψη θέτει την βάση για βελτιωμένες προσεγγίσεις.

Συγκεκριμένα, μπορούμε να κατασκευάσουμε συναρτησιακά ενέργειας (τα οποία ονομάζονται "προσεγγίσεις ελεύθερης ενέργειας Kikuchi") που είναι της μορφής της εξίσωσης (2.5), στην οποία παρουσιάζονται επιπλέον όροι εντροπίας, τόσο με θετικά όσο και με αρνητικά πρόσημα κατά τέτοιο τρόπο ώστε να είναι σίγουρο ότι το counting number για όλες τις μεταβλητές είναι 1.

Κατά κάποιο τρόπο, η ίδια ανάλυση που κάναμε σε αυτή την παράγραφο (δηλαδή ο ορισμός ενός συνόλου εξισώσεων σταθερού σημείου για στάσιμα σημεία της προσέγγισης της ελεύθερης ενέργειας) μας οδηγεί επίσης σε αλγόριθμους διάδοσης μηνύματος που αφορούν αυτές τις πιο περιεκτικές προσεγγίσεις.

Οι κανόνες διάδοσης (propagation rules) αυτών των προσεγγίσεων, οι οποίοι επίσης εντάσσονται στην κατηγορία της γενικευμένης αναπαραγωγής πίστεως, είναι πιο πλήρεις και δεν θα συζητήσουμε περισσότερο για αυτούς σε αυτό το σημείο.

### **2.3.4 Προσεγγιστικός συμπερασμός που βασίζεται στην δειγματοληψία**

#### **2.3.4.1 Μέθοδοι Monte Carlo για αλυσίδες Markov**

Η αλυσίδα Markov Monte Carlo (Markov Chain Monte Carlo) ,για συντομία θα γράφουμε MCMC, είναι μία προσέγγιση για την δημιουργία δειγμάτων από την ύστερη (posterior) κατανομή.

Όπως έχουμε ήδη δει ,τυπικά, δεν μπορούμε να κάνουμε απευθείας δειγματοληψία από την ύστερη κατανομή.

Μπορούμε όμως να δημιουργήσουμε μία διαδικασία η οποία να δειγματοληπτεί βαθμιαία από κατανομές που είναι όλο και πιο κοντά στην ύστερη.

Ορίζουμε ένα (state graph) του οποίου οι κόμβοι είναι οι καταστάσεις του συστήματος, δηλαδή δυνατές στιγμές  $Val(X)$ .

Αυτός ο γράφος είναι πολύ διαφορετικός σε σχέση με το γραφικό μοντέλο που ορίζει την κατανομή  $P(X)$ , του οποίου οι κόμβοι αντιστοιχούν σε μεταβλητές.

Ορίζουμε μια διαδικασία η οποία διασχίζει τυχαία αυτόν τον γράφο και πηγαίνει από την μία κατάσταση στην άλλη.

Αυτή η διαδικασία ορίζεται έτσι ώστε τελικώς (μετά από αρκετά βήματα) η πιθανότητα να βρίσκεται σε οποιαδήποτε συγκεκριμένη κατάσταση να είναι η επιθυμητή ύστερη κατανομή.

Αρχίζουμε περιγράφοντας το γενικό πλαίσιο εργασίας όταν έχουμε αλυσίδες Markov και συνεχίζουμε με την εφαρμογή τους στον προσεγγιστικό συμπερασμό όταν έχουμε γραφικά μοντέλα.

Σημειώνουμε ότι σε αντίθεση με τις μεθόδους απευθείας δειγματοληψίας, οι μέθοδοι αλυσίδων Markov εφαρμόζονται εξίσου καλά τόσο σε κατευθυνόμενα όσο και σε μη κατευθυνόμενα μοντέλα.

Μια αλυσίδα Markov ορίζεται ως ένα σύνολο καταστάσεων μαζί με ένα μοντέλο μετάβασης από κατάσταση σε κατάσταση.

Η αλυσίδα ορίζει μια διαδικασία η οποία εξελίσσεται στοχαστικά από κατάσταση σε κατάσταση.

Μια αλυσίδα Markov ορίζεται μέσω ενός χώρου καταστάσεων  $Val(X)$  και ενός πιθανοτικού μοντέλου μετάβασης το οποίο ορίζει για κάθε κατάσταση  $x \in Val(X)$  μια κατανομή επόμενης κατάστασης (next-state distribution) επί του  $Val(X)$ .

Η πιθανότητα μετάβασης από την  $x$  στην  $x'$  σημειώνεται ως  $T(x \rightarrow x')$  και εφαρμόζεται όποτε η αλυσίδα βρίσκεται στην κατάσταση  $x$ .

Σημειώνουμε ότι σε αυτό τον ορισμό αλλά και στην συζήτηση που θα ακολουθήσει, περιορίζουμε το ενδιαφέρον μας στις ομογενείς αλυσίδες Markov όπου η δυναμική του συστήματος δεν αλλάζει καθώς περνάει ο χρόνος.

Μπορούμε να φανταστούμε μια τυχαία διαδικασία δειγματοληψίας η οποία να ορίζει μια ακολουθία καταστάσεων  $x^{(0)}, x^{(1)}, x^{(2)}, \dots$

Καθώς το μοντέλο μετάβασης είναι τυχαίο, η κατάσταση της διαδικασίας στο βήμα  $t$  μπορεί να θεωρηθεί ως μια τυχαία μεταβλητή  $X^{(t)}$ .

Υποθέτουμε ότι η αρχική κατάσταση  $X^{(0)}$  είναι κατανομημένη σύμφωνα με κάποια κατανομή αρχικής καταστάσεως  $P^{(0)}(X^{(0)})$ .

Μπορούμε τώρα να ορίσουμε κατανομές επί των επόμενων καταστάσεων  $P^{(1)}(X^{(1)})$ ,  $P^{(2)}(X^{(2)})$ ,  $\dots$  χρησιμοποιώντας την δυναμική αλυσίδας :

$$P^{(t+1)}(X^{(t+1)} = x') = \sum_{x \in Val(X)} P^{(t)}(X^{(t)} = x) T(x \rightarrow x'). \quad (2.17)$$

Η πιθανότητα να βρεθεί η διαδικασία στην κατάσταση  $x'$  την στιγμή  $t+1$  είναι το άθροισμα, επί όλων των πιθανών καταστάσεων  $x$  στις οποίες η αλυσίδα θα μπορούσε να βρεθεί, της πιθανότητας να βρίσκεται στην κατάσταση  $x$  την στιγμή  $t$  επί την πιθανότητα η αλυσίδα να έκανε την μετάβαση από την  $x$  στην  $x'$ .

Καθώς η διαδικασία συγκλίνει, θα περιμέναμε το  $P^{(t+1)}$  να είναι κοντά στο  $P^{(t)}$ .

Από την (2.17) καταλήγουμε στην



$$P^{(t)}(\mathbf{x}') \approx P^{(t+1)}(\mathbf{x}') = \sum_{\mathbf{x} \in \text{Val}(\mathbf{X})} P^{(t)}(\mathbf{x})T(\mathbf{x} \rightarrow \mathbf{x}').$$

Κατά την σύγκλιση θα περιμέναμε η προκύπτουσα κατανομή  $\pi(\mathbf{X})$  να είναι μια ισορροπία σχετική με το μοντέλο μετάβασης.

Δηλαδή, η πιθανότητα να βρίσκεται σε μια κατάσταση είναι ίδια με την πιθανότητα να μεταβεί σε αυτήν την κατάσταση ξεκινώντας από κάποιον τυχαία δειγματοληπτημένο πρόγονο.

Πιο επίσημα έχουμε τον παρακάτω ορισμό.

Μια κατανομή  $\pi(\mathbf{X})$  είναι μια στάσιμη κατανομή για μια αλυσίδα Markov  $T$  εάν ικανοποιεί την παρακάτω σχέση

$$\pi(\mathbf{X} = \mathbf{x}') = \sum_{\mathbf{x} \in \text{Val}(\mathbf{X})} \pi(\mathbf{X} = \mathbf{x})T(\mathbf{x} \rightarrow \mathbf{x}') \quad (2.18)$$

Θα θέλαμε να περιορίσουμε την προσοχή μας στις αλυσίδες Markov που έχουν μοναδική στάσιμη κατανομή και στην οποία υπάρχει πρόσβαση από οποιαδήποτε αρχική κατανομή  $P^{(0)}$ .

Υπάρχουν διάφορες συνθήκες που είναι επαρκείς ώστε να μας εγγυηθούν αυτή την ιδιότητα.

Η συνθήκη που χρησιμοποιείται συνήθως είναι περισσότερο "τεχνικής φύσεως" και μας λέει ότι η αλυσίδα πρέπει να είναι εργοδική.

Μιλώντας για αλυσίδες Markov όπου ο χώρος καταστάσεων  $\text{Val}(\mathbf{X})$  είναι πεπερασμένος, η ακόλουθη συνθήκη ισοδυναμεί με αυτή την απαίτηση:

Μια αλυσίδα Markov λέμε ότι είναι κανονική (regular) εάν υπάρχει κάποιος αριθμός  $k$  τέτοιος ώστε για κάθε  $\mathbf{x}, \mathbf{x}' \in \text{Val}(\mathbf{X})$ , η πιθανότητα μετάβασης από την  $\mathbf{x}$  στην  $\mathbf{x}'$  σε ακριβώς  $k$  βήματα είναι μεγαλύτερη από 0.

Ισχύει το παρακάτω συμπέρασμα.

Μια αλυσίδα Markov πεπερασμένων καταστάσεων  $T$  έχει μοναδική στάσιμη κατανομή αν και μόνο αν είναι κανονική.

Το να επιβεβαιωθεί η κανονικότητα (regularity) είναι συνήθως κάτι που φαίνεται απευθείας.

Δύο απλές συνθήκες που εγγυώνται την κανονικότητα σε αλυσίδες Markov πεπερασμένων καταστάσεων  $T$  είναι :

1) Είναι δυνατόν να φτάσουμε από οποιαδήποτε κατάσταση σε οποιαδήποτε άλλη κατάσταση χρησιμοποιώντας ένα μονοπάτι θετικής πιθανότητας στον γράφο κατάστασης.

2) Για κάθε κατάσταση  $x$ , υπάρχει μια θετική πιθανότητα μετάβασης από το  $x$  στο  $x$  σε ένα βήμα (ένας αυτοβρόχος).

Αυτές οι δύο συνθήκες μαζί, είναι ικανές αλλά όχι αναγκαίες ώστε να έχουμε εγγυημένη κανονικότητα.

Είναι γεγονός όμως ότι συνήθως ισχύουν για τις αλυσίδες που χρησιμοποιούνται στην πράξη.

### 2.3.4.2 Αλυσίδες Markov για γραφικά μοντέλα

Η θεωρία των αλυσίδων Markov μας παρέχει ένα γενικό πλαίσιο εργασίας για την δημιουργία δειγμάτων από μία στοχευμένη κατανομή  $\pi$ .

Σε αυτή την παράγραφο θα μιλήσουμε για την εφαρμογή αυτού του πλαισίου στις διεργασίες δειγματοληψίας (sampling tasks) που συναντώνται στα πιθανοτικά γραφικά μοντέλα.

Τυπικά, επιθυμούμε να δημιουργήσουμε δείγματα από την ύστερη ακολουθία  $P(X | E = e)$ .

Συνεπώς θέλουμε να ορίσουμε μια αλυσίδα για την οποία η  $P(X | e)$  είναι η στάσιμη κατανομή.

Προφανώς υπάρχουν πολλοί τρόποι για να οριστεί μια τέτοια αλυσίδα.

Θα επικεντρωθούμε στις πιο συνηθισμένες προσεγγίσεις.

Στα γραφικά μοντέλα, ορίζουμε τις καταστάσεις της αλυσίδας Markov να είναι εκφράσεις  $\xi$  του  $X$ , οι οποίες είναι συμβατές με το  $e$ .

Δηλαδή, όλες οι καταστάσεις  $\xi$  στην αλυσίδα Markov ικανοποιούν την σχέση  $\xi \in E = e$ .

Γι' αυτό, οι καταστάσεις σε μία αλυσίδα Markov είναι κάποιο υποσύνολο των πιθανών αναθέσεων στις μεταβλητές  $X$ .

Προκειμένου να ορίσουμε μια αλυσίδα Markov πρέπει να ορίσουμε μια διαδικασία που μεταβαίνει από την μία κατάσταση στην άλλη, συγκλίνοντας σε μια στάσιμη κατανομή  $\pi(\xi)$  η οποία είναι η επιθυμητή ύστερη κατανομή  $P(\xi | e)$ .

Για την περίπτωση των γραφικών μοντέλων, ο χώρος κατάστασης έχει παραγοντοποιημένη δομή (κάθε κατάσταση είναι μία ανάθεση σε πολλαπλές μεταβλητές).

Όταν ορίζουμε ένα μοντέλο μετάβασης επί αυτού του χώρου κατάστασης, μπορούμε να θεωρήσουμε μια πολύ γενική περίπτωση όπου μια μετάβαση μπορεί να γίνει από οποιαδήποτε κατάσταση σε μια άλλη οποιαδήποτε κατάσταση.

Συχνά όμως μας βολεύει να αποδομούμε το μοντέλο μετάβασης λαμβάνοντας υπόψη τις μεταβάσεις οι οποίες ενημερώνουν μια και μόνο συνιστώσα του διανύσματος κατάστασης κάθε φορά, δηλαδή μόνο μια τιμή για κάποια μεταβλητή.

Σε αυτή την περίπτωση, συχνά ορίζουμε ένα σύνολο μοντέλων μετάβασης  $T_1, \dots, T_k$ , κάθε ένα με την δική του δυναμική.

Κάποιες φορές, τα διαφορετικά μοντέλα είναι απαραίτητα καθώς κανένα μοντέλο μετάβασης από μόνο του δεν επαρκεί ώστε να σιγουρέψει την κανονικότητα.

Άλλες φορές, το να έχουμε πολλά μοντέλα μετάβασης απλώς κάνει τον χώρο κατάστασης πιο "συνδεδεμένο" και γι' αυτό επιταχύνεται η σύγκλιση προς μια στατική κατανομή.

Υπάρχουν διάφοροι τρόποι συνδυασμού αυτών των πολλαπλών μοντέλων μετάβασης σε μία αλυσίδα.



Ένας συνηθισμένος απλός τρόπος είναι να επιλέγουμε τυχαία μεταξύ αυτών σε κάθε βήμα, χρησιμοποιώντας οποιαδήποτε κατανομή.

Έτσι, για παράδειγμα, σε κάθε βήμα μπορούμε να επιλέξουμε ένα εκ των  $T_1, \dots, T_k$ , κάθε ένα με πιθανότητα  $1/k$ .

Εναλλακτικά, μπορούμε απλώς να κάνουμε ανακύκλωση μεταξύ των διαφορετικών μοντέλων μετάβασης και να παίρνουμε ένα κάθε φορά.

Προφανώς με αυτή την προσέγγιση, δεν ορίζεται μια ομογενής αλυσίδα καθώς το μοντέλο μετάβασης που χρησιμοποιείται στο  $i$  βήμα είναι διαφορετικό από το μοντέλο που χρησιμοποιείται στο  $i+1$  βήμα.

Μπορούμε όμως να δούμε την όλη διαδικασία σαν να ορίζουμε ένα και μόνο μοντέλο  $T$ , κάθε βήμα του οποίου είναι και ένα βήμα συγκέντρωσης που προκύπτει λαμβάνοντας πρώτα το  $T_1$ , μετά το  $T_2$ , ... , εώς και το  $T_k$ .

Στην περίπτωση των γραφικών μοντέλων, ορίζουμε  $X = X - E = \{X_1, \dots, X_k\}$ .

Ορίζουμε μια αλυσίδα πολλαπλών μεταβάσεων, στην οποία έχουμε ένα τοπικό μοντέλο μετάβασης  $T_i$  για κάθε μεταβλητή  $X_i \in X$ .

Έστω  $U_i = X - \{X_i\}$  και ότι με  $u_i$  συμβολίζουμε μία έκφραση στο  $U_i$ .

Το μοντέλο  $T_i$  λαμβάνει μια κατάσταση  $(u_i, x_i)$  και μεταβαίνει σε μια κατάσταση της μορφής  $\text{form}(u_i, x_i)$ .

Όπως είδαμε και νωρίτερα, μπορούμε να συνδυάσουμε τα διαφορετικά τοπικά μοντέλα μετάβασης σε ένα και μόνο οικουμενικό μοντέλο με πολλούς διαφορετικούς τρόπους.

### 2.3.4.3 Δειγματοληψία Gibbs

Η δειγματοληψία Gibbs είναι μια απλή αλλά αποτελεσματική αλυσίδα Markov για παραγοντοποιημένους χώρους κατάστασης, η οποία είναι ιδιαίτερος χρήσιμη για γραφικά μοντέλα.

Ορίζουμε το τοπικό μοντέλο μετάβασης  $T_i$  παρακάτω.

Για να το κάνουμε αυτό, απλώς "ξεχνάμε" την τιμή της  $X_i$  για την τρέχουσα κατάσταση και δειγματοληπτούμε μια νέα τιμή για αυτήν από τις ύστερες της (posterior) δεδομένου του υπολοίπου της παρούσας κατάστασης.

Έστω  $(u_i, x_i)$  μία κατάσταση στην αλυσίδα.

Ορίζουμε

$$T((u_i, x_i) \rightarrow (u_i, x'_i)) = P(x'_i | u_i). \quad (2.19)$$

Να σημειωθεί ότι η πιθανότητα μετάβασης δεν εξαρτάται από την τρέχουσα τιμή  $x_i$  της  $X_i$  αλλά μόνο από την απομένουσα κατάσταση  $u_i$ .

Η αλυσίδα Gibbs ορίζεται μέσω ενός συνόλου τοπικών μοντέλων μετάβασης.

Χρησιμοποιούμε το πολυβηματικό (multistep) μοντέλο μετάβασης προκειμένου να τις συνδυάσουμε.

Οι διαφορετικές τοπικές μεταβάσεις λαμβάνονται διαδοχικά, δηλαδή εφόσον έχει αλλάξει η τιμή για μια μεταβλητή  $X_1$ , δειγματοληπτείται η τιμή για την  $X_2$  βασιζόμενη στην νέα τιμή.

Να σημειωθεί ότι συλλέγουμε ένα μόνο δείγμα για κάθε ακολουθία της οποίας η τοπική μετάβαση έχει επιλεγεί μία φορά.

Αυτή η αλυσίδα είναι εγγυημένα κανονική (regular) όποτε η κατανομή είναι θετική, έτσι ώστε κάθε τιμή της  $X_i$  να έχει θετική πιθανότητα δεδομένης κάποιας ανάθεσης  $u_i$  στις υπόλοιπες μεταβλητές.

Σε αυτή την περίπτωση μπορούμε να μεταβούμε από κάποια οποιαδήποτε κατάσταση σε οποιαδήποτε άλλη κατάσταση με το πολύ  $k$  τοπικά βήματα μετάβασης, όπου  $k = |X - E|$ .

Η θετικότητα παρ' όλα αυτά όμως δεν είναι απαραίτητη καθώς υπάρχουν πολλά παραδείγματα μη θετικών κατανομών όπου η αλυσίδα Gibbs είναι κανονική.

Επίσης είναι εύκολο να δείξουμε ότι η ύστερη κατανομή  $P(X | e)$  είναι μια στάσιμη κατανομή αυτής της διαδικασίας (process).

Η δειγματοληψία Gibbs ταιριάζει πολύ καλά σε πολλά γραφικά μοντέλα όποτε και μπορούμε να υπολογίσουμε την πιθανότητα μετάβασης  $P(X_i | u_i)$  με πολύ αποδοτικό τρόπο.

Όπως θα δείξουμε στην συνέχεια, μια τέτοια κατανομή μπορεί να γίνει βασισόμενη μόνο στην περιοχή κάλυψης Markov (Markov blanket) της  $X_i$ .

Θα δείξουμε αυτή την ανάλυση για ένα δίκτυο Markov και η επέκταση της για Bayesian δίκτυα θα είναι προφανής.

Γενικά, μπορούμε να αποδομήσουμε την πιθανότητα μιας έκφρασης ως εξής

$$P(x_1 | x_2, \dots, x_n) = \frac{1}{Z} \prod_j \pi_j[C_j] = \frac{1}{Z} \prod_{j: X_i \in C_j} \pi_j[C_j] \prod_{j: X_i \notin C_j} \pi_j[C_j].$$

Για συντομία, θα συμβολίζουμε το  $\pi_j[x_i, u_{\langle C_j \rangle}]$  ως  $\pi_j[x_i, u]$ .

Μπορούμε τώρα να υπολογίσουμε την

$$P(x'_i | u_i) = \frac{P(x'_i, u_i)}{\sum_{x''_i} P(x''_i, u_i)} = \frac{\prod_{C_j \ni X_i} \pi_j[x'_i, u_i]}{\sum_{x''_i} \prod_{C_j \ni X_i} \pi_j[x''_i, u_i]}. \quad (2.20)$$

Αυτή η τελευταία έκφραση χρησιμοποιεί μόνο τα δυναμικά των ακμών που έχουν να κάνουν με την  $X_i$  και εξαρτάται μόνο από τις εκφράσεις της  $u_i$  της περιοχής κάλυψης της  $X_i$  ( $X_i$ 's Markov blanket).

Για την περίπτωση των Bayesian δικτύων, αυτή η έκφραση ελαττώνεται σε έναν τύπο που περιέχει μόνο τις δεσμευμένες κατανομές της  $X_i$  και των απογόνων της και η τιμή της εξαρτάται και πάλι μόνο από την ανάθεση της  $u_i$  στην περιοχή κάλυψης της  $X_i$ .

Σύμφωνα με τα παραπάνω μπορεί επομένως να υπολογιστεί πολύ αποδοτικά.

Να σημειωθεί ότι η αλυσίδα Markov που ορίζεται από γραφικό μοντέλο δεν είναι απαραίτητα κανονική και μπορεί να μην συγκλίνει σε κάποια μοναδική στάσιμη κατανομή.

Φαίνεται επομένως ότι αυτός ο τύπος περίπτωσης μπορεί να προκύψει εάν η κατανομή που ορίζεται από το γραφικό μοντέλο είναι μη θετική, δηλαδή αν οι δεσμευμένες κατανομές ή τα δυναμικά των ακμών έχουν τιμή 0.

Έστω  $H$  ένα δίκτυο Markov τέτοιο ώστε όλα τα δυναμικά των ακμών να είναι αυστηρώς θετικά. Τότε η αλυσίδα Markov που προκύπτει με δειγματοληψία Gibbs είναι κανονική (regular).

#### 2.3.4.4 Κατασκευή μιας αλυσίδας Markov

Όπως είδαμε, η χρήση μεθόδων MCMC στηρίζεται στην κατασκευή μιας αλυσίδας Markov η οποία να έχει δύο επιθυμητές ιδιότητες : κανονικότητα και στάσιμη στοχευμένη κατανομή (target stationary distribution)

Ως τώρα έχουμε περιγράψει την αλυσίδα Gibbs που είναι μια απλή αλυσίδα Markov που έχει εγγυημένα τις εν λόγω ιδιότητες υπό συγκεκριμένες προϋποθέσεις.

Η δειγματοληψία Gibbs όμως είναι εφαρμόσιμη μόνο σε ορισμένες περιπτώσεις.

Συγκεκριμένα, θα πρέπει να είμαστε σε θέση να δειγματοληπτούμε από την κατανομή  $P(X_i | u_i)$  και παρότι αυτό το βήμα δειγματοληψίας είναι εύκολο να γίνει για διακριτά γραφικά μοντέλα, υπάρχουν και κάποιοι άλλοι τύποι μοντέλων όπου αυτό το βήμα δεν είναι πρακτικό και έτσι η αλυσίδα Gibbs δεν είναι εφαρμόσιμη

Η πιο γενική μέθοδος κατασκευής αλυσίδας Markov η οποία εγγυημένα θα συγκλίνει στην επιθυμητή στάσιμη κατανομή είναι ο αλγόριθμος Metropolis-Hastings, ο οποίος όμως δεν θα μας απασχολήσει στην παρούσα εργασία.

#### 2.3.4.5 Δημιουργία δειγμάτων

Ο χρόνος δημιουργίας για μια μεγάλη αλυσίδα Markov είναι συχνά επίσης πολύ μεγάλος, οπότε ο αλγόριθμος που περιγράψαμε στα προηγούμενα πρέπει να εκτελέσει ένα μεγάλο αριθμό βημάτων δειγματοληψίας για κάθε χρήσιμο δείγμα.

Μια μεγάλης σημασίας παρατήρηση είναι ότι, αν το  $x^{(t)}$  έχει δειγματοληπτηθεί από την  $\pi$ , τότε και το  $x^{(t+1)}$  έχει επίσης δειγματοληπτηθεί από την  $\pi$ .

Οπότε, από την στιγμή που θα έχουμε "τρεξει" την αλυσίδα αρκετά ώστε να δειγματοληπτούμε από την στάσιμη κατανομή (ή από μια κατανομή που είναι κοντά σε αυτή), μπορούμε να συνεχίσουμε να δημιουργούμε δείγματα από την ίδια τροχιά και να λάβουμε ένα μεγάλο πλήθος δειγμάτων από την στάσιμη κατανομή.

Πιο επίσημα, ας υποθέσουμε ότι χρησιμοποιούμε τα  $x^{(0)}, \dots, x^{(T)}$  ως η φάση "burn-in" και κατόπιν συλλέγουμε  $M$  δείγματα  $x^{(T+1)}, \dots, x^{(T+M)}$ .

Κατά αυτόν τον τρόπο συλλέξαμε ένα σύνολο δεδομένων (data set)  $D$ , όπου  $x^m = x^{(T+m)}$  για  $m = 1, \dots, M$ .

Ας υποθέσουμε για απλότητα ότι το  $x^{(T+1)}$  έχει δειγματοληπτηθεί από την  $\pi$  και έτσι το ίδιο έχει γίνει και για όλα τα δείγματα στο σύνολο  $D$ .

Τότε, για κάθε συνάρτηση

$$f: \sum_{m=1}^M f(x^m)$$

είναι μια αμερόληπτη εκτιμήτρια για το

$$E_{\pi(\mathbf{X})}[f(\mathbf{X})].$$

Το βασικό πρόβλημα φυσικά είναι ότι διαδοχικά δείγματα από την ίδια τροχιά είναι συσχετισμένα, οπότε δεν μπορούμε να περιμένουμε την ίδια απόδοση όπως θα περιμέναμε από  $M$  ανεξάρτητα δείγματα από την  $\pi$ .

Με άλλα λόγια, η διακύμανση της εκτιμήτριας είναι σημαντικά υψηλότερη από ότι αυτή μιας εκτιμήτριας που δημιουργείται από  $M$  ανεξάρτητα δείγματα της  $\pi$ .

Μια λύση σε αυτό το πρόβλημα είναι να μην συλλέξουμε διαδοχικά δείγματα από την αλυσίδα.

Αντί αυτού, εφόσον έχουμε συλλέξει ένα δείγμα  $x^{(T)}$ , αφήνουμε την αλυσίδα να "τρέξει" για λίγο και κατόπιν συλλέγουμε ένα δεύτερο δείγμα  $x^{(T+d)}$  για κάποια κατάλληλη επιλογή του  $d$ .

Για αρκετά μεγάλο  $d$ , τα  $x^{(T)}$  και  $x^{(T+d)}$  είναι μόνο ελαφρώς συσχετισμένα και μπορούμε να τα δούμε πρακτικά σαν ανεξάρτητα δείγματα από την  $\pi$ .

Ο χρόνος  $d$  που απαιτείται για να "εξασθενήσει" η συσχέτιση εξαρτάται από τον χρόνο μίξης της αλυσίδας, οπότε αλυσίδες που σμίγουν αργά θα χρειαστούν αρχικά μεγαλύτερο  $d$  προκειμένου να παραχθούν δείγματα που είναι πιο κοντά στην ανεξαρτησία.

Παρ' όλα αυτά, τα δείγματα προέρχονται από την σωστή κατανομή για οποιαδήποτε τιμή του  $d$  και γι' αυτό είναι συνήθως καλύτερα να συμβιβαστούμε και να χρησιμοποιούμε κάποιο μικρότερο  $d$  από ότι να χρησιμοποιούμε μικρότερο "burn-in" χρόνο  $T$ .

Αυτή η μέθοδος μας επιτρέπει να συλλέξουμε μεγαλύτερο αριθμό χρήσιμων δειγμάτων με λιγότερες μεταβάσεις της αλυσίδας Markov.

Για την ακρίβεια, μπορούμε συχνά να κάνουμε ακόμα καλύτερη χρήση των δημιουργούμενων δειγμάτων κάνοντας χρήση αυτής της προσέγγισης "μονής αλυσίδας" (single-chain approach).

Αν και τα δείγματα μεταξύ των  $x^{(T)}$  και  $x^{(T+d)}$  δεν είναι ανεξάρτητα, δεν υπάρχει λόγος να τα αγνοήσουμε.

Αυτό σημαίνει ότι χρησιμοποιώντας όλα τα δείγματα  $x^{(T)}, x^{(T+1)}, \dots, x^{(T+d)}$  παράγεται μια αποδεδειγμένα καλύτερη εκτιμήτρια από ότι αν χρησιμοποιηθούν μόνο τα δύο δείγματα  $x^{(T)}$  και  $x^{(T+d)}$ .

Η διακύμανση δεν είναι υψηλότερη εάν χρησιμοποιήσουμε όλα τα δείγματα που δημιουργήσαμε αντί για κάποιο υποσύνολο αυτών, οπότε η στρατηγική του να επιλεγεί μόνο ένα υποσύνολο των δειγμάτων είναι χρήσιμο κυρίως σε περιπτώσεις όπου υπάρχει κάποιο σημαντικό κόστος σχετιζόμενο με την χρήση κάθε δείγματος οπότε και θα θέλαμε να ελαττώσουμε τον συνολικό αριθμό των χρησιμοποιούμενων δειγμάτων.

#### 2.3.4.6 Γενικές παρατηρήσεις

Η περιγραφή της χρήσης των αλυσίδων Markov που έχει γίνει ως αυτό το σημείο, εξακολουθεί να είναι αρκετά περιληπτική καθώς δεν περιέχονται οι προδιαγραφές για τον αριθμό των αλυσίδων που μπορούν να τρέξουν, τα μέτρα για την αξιολόγηση της μίξης, τεχνικές που να καθορίζουν την καθυστέρηση μεταξύ των δειγμάτων που θα τους επέτρεπε να θεωρηθούν ανεξάρτητα καθώς και άλλα πολλά.

Δυστυχώς, μέχρι αυτό το σημείο έχει γίνει μικρή θεωρητική ανάλυση που θα μπορούσε να μας βοηθήσει να δώσουμε απαντήσεις στα προηγούμενα ερωτήματα που αφορούν τις αλυσίδες που μας ενδιαφέρουν και για αυτό η εφαρμογή των αλυσίδων Markov είναι περισσότερο τέχνη παρά επιστήμη και συχνά απαιτείται έντονος πειραματισμός και χειροκίνητη ρύθμιση των παραμέτρων.

Παρ' όλα αυτά όμως, οι τεχνικές MCMC είναι οι μόνες τεχνικές που μπορεί να έχουν καλή απόδοση για αρκετά πιθανοτικά μοντέλα.

Συγκεκριμένα, σε αντίθεση με τις μεθόδους απευθείας δειγματοληψίας, δεν υποβαθμίζεται όταν η πιθανότητα των αποδεικτικών (evidence) είναι μικρή ή όταν η ύστερη κατανομή είναι κατά πολύ διαφορετική σε σχέση με την αρχική. Επιπλέον, σε αντίθεση και πάλι με την απευθείας δειγματοληψία, εφαρμόζονται τόσο σε προσανατολισμένα όσο και σε μη προσανατολισμένα μοντέλα και για αυτό αποτελούν σημαντικό μέρος των τεχνικών προσεγγιστικού συμπερασμού.

## 2.4 Μάθηση

Στην συνέχεια θα ασχοληθούμε με γραφικά μοντέλα που μπορούν να μαθαίνουν. Υπάρχουν δύο παραλλαγές για την διαδικασία της μάθησης: η εκτίμηση παραμέτρων (parameter estimation) και η μάθηση δομής (structure learning). Στην εκτίμηση παραμέτρων, υποθέτουμε ότι η ποιοτική εξάρτηση δομής του γραφικού μοντέλου είναι γνωστή. Δηλαδή στην περίπτωση του προσανατολισμένου μοντέλου το  $G$  θα είναι δεδομένο ενώ αντίστοιχα στην περίπτωση του μη προσανατολισμένου μοντέλου δεδομένο θα είναι το  $H$ . Σε αυτή την περίπτωση, η διαδικασία της μάθησης είναι απλώς να συμπληρωθούν οι παράμετροι που ορίζουν τις CPDs των ιδιοτήτων ή των παραμέτρων που ορίζουν τις συναρτήσεις δυναμικού του δικτύου Markov. Στην διαδικασία μάθησης δομής δεν απαιτείται κάποια επιπλέον είσοδος (input) παρ' όλο που ο χρήστης μπορεί να δώσει προηγούμενη γνώση σχετικά με την δομή, εφόσον αυτή υπάρχει, υπό μορφή περιορισμών. Σκοπός είναι να εξαχθεί ένα Bayesian δίκτυο ή ένα δίκτυο Markov, τόσο η δομή του όσο και οι παράμετροι του, από τα δεδομένα εκπαίδευσης (training data) και μόνο. Θα μιλήσουμε για κάθε ένα από αυτά τα προβλήματα ξεχωριστά.

### 2.4.1 Εκτίμηση παραμέτρων σε Bayesian δίκτυα

Ξεκινάμε με την εκμάθηση παραμέτρων για ένα Bayesian δίκτυο όπου η εξάρτηση της δομής είναι γνωστή. Με άλλα λόγια, μας δίνετε η δομή  $G$  που καθορίζει το σύνολο των γονέων για κάθε τυχαία μεταβλητή και σκοπός μας είναι να μάθουμε τις παραμέτρους  $\theta_G$  που ορίζουν τις CPDs για αυτή την δομή. Η μάθηση βασίζεται σε ένα συγκεκριμένο εκπαιδευτικό σύνολο  $D = \{x^1, \dots, x^m\}$  το οποίο, για τώρα, θα θεωρήσουμε ότι είναι πλήρες (δηλαδή ότι κάθε στιγμή είναι πλήρως παρατηρημένη και ότι δεν υπάρχουν τιμές που να παραλείφθηκαν). Υπάρχουν δύο προσεγγίσεις για την εκτίμηση παραμέτρων: η εκτίμηση μεγίστης πιθανοφάνειας (Maximum Likelihood Estimation - MLE) και οι Bayesian προσεγγίσεις (Bayesian approaches). Το βασικό συστατικό και για τις δύο είναι η συνάρτηση πιθανοφάνειας: η πιθανότητα των δεδομένων (data) με το μοντέλο να είναι δοσμένο. Αυτή η συνάρτηση περιγράφει την απόκριση της πιθανοτικής κατανομής σε αλλαγές της επιλογής των παραμέτρων.

Η πιθανοφάνεια ενός συνόλου παραμέτρων ορίζεται ως η πιθανότητα των δεδομένων (data) όταν είναι δεδομένο το μοντέλο.

Για ένα δίκτυο Bayesian δομής  $G$ , η πιθανοφάνεια ενός συνόλου παραμέτρων  $\theta_G$  είναι

$$L(\theta_G : \mathcal{D}) = P(\mathcal{D} | \theta_G).$$

#### 2.4.1.1 Εκτίμηση παραμέτρων μεγίστης πιθανοφάνειας

Σύμφωνα με όσα είπαμε στα προηγούμενα, μια προσέγγιση για την εκτίμηση παραμέτρων είναι η εκτίμηση παραμέτρων μεγίστης πιθανοφάνειας (maximum likelihood parameter estimation)

Εδώ, σκοπός μας είναι να βρούμε το σύνολο παραμέτρων  $\theta_G$  που μεγιστοποιεί την πιθανοφάνεια  $L(\theta_G : \mathcal{D})$ .

Για Bayesian δίκτυα, η πιθανοφάνεια μπορεί να αναλυθεί ως εξής :

$$\begin{aligned} L(\theta_G, \mathcal{D}) &= \prod_{j=1}^m P(x^j : \theta_G) \\ &= \prod_{j=1}^m \prod_{i=1}^n P(x_i^j | \mathbf{Pa}_{x_i^j} : \theta_G) \\ &= \prod_{i=1}^n \prod_{j=1}^m P(x_i^j | \mathbf{Pa}_{x_i^j} : \theta_G) \end{aligned}$$

Θα χρησιμοποιήσουμε το  $\theta_{X_i | \mathbf{Pa}_i}$  για να συμβολίσουμε το υποσύνολο των παραμέτρων που καθορίζουν την  $P(X_i | \mathbf{Pa}_i)$ .

Για την περίπτωση που οι παράμετροι είναι αποσυνδεδεμένες, δηλαδή όταν κάθε CPD είναι παραμετροποιημένη με ένα ξεχωριστό σύνολο παραμέτρων το οποίο δεν επικαλύπτει κάποιο άλλο, μας δίνεται η δυνατότητα να μεγιστοποιήσουμε κάθε σύνολο παραμέτρων ανεξάρτητα.

Μπορούμε να γράψουμε την πιθανοφάνεια ως εξής :

$$L(\theta_G : \mathcal{D}) = \prod_{i=1}^n L_i(\theta_{X_i | \mathbf{Pa}_i} : \mathcal{D}),$$

όπου η συνάρτηση τοπικής πιθανοφάνειας για την  $X_i$  είναι

$$L_i(\theta_{X_i | \mathbf{Pa}_i} : \mathcal{D}) = \prod_{j=1}^m P(x_i^j | \mathbf{pa}_i^j : \theta_{X_i | \mathbf{Pa}_i}).$$

Η απλούστερη παραμετροποίηση των CPDs είναι με την μορφή ενός πίνακα.

Ας υποθέσουμε ότι έχουμε μια μεταβλητή  $X$  με γονείς  $U$ .



Εάν αναπαραστήσουμε αυτήν την CPD  $P(X|U)$  με ένα πίνακα, τότε θα έχουμε μια παράμετρο  $\theta_{x|u}$  για κάθε συνδυασμό των  $x \in \text{Val}(X)$  και  $u \in \text{Val}(U)$ . Θα μπορούμε επομένως να γράψουμε την συνάρτηση τοπικής πιθανοφάνειας με τον ακόλουθο τρόπο :

$$\begin{aligned} L_X(\theta_{X|U} : \mathcal{D}) &= \prod_{j=1}^m \theta_{x^j|u^j} \\ &= \prod_{\mathbf{u} \in \text{Val}(U)} \left[ \prod_{x \in \text{Val}(X)} \theta_{x|\mathbf{u}}^{N_{\mathbf{u},x}} \right], \end{aligned} \quad (2.21)$$

όπου  $N_{\mathbf{u},x}$  είναι ο αριθμός των στιγμών όπου  $X = x$  και  $\text{Pa}_i = \mathbf{u}$  στο  $\mathcal{D}$ .

Με αυτό τον τρόπο καταφέρνουμε να συγκεντρώσουμε μαζί όλες τις εμφανίσεις του  $\theta_{x|u}$  μέσα στο γινόμενο για όλες τις περιπτώσεις.

Πρέπει να μεγιστοποιήσουμε αυτόν τον όρο υπό τον περιορισμό ότι για κάθε επιλογή τιμής για τους γονείς  $U$ , η δεσμευμένη πιθανότητα είναι επιτρεπτή :

$$\sum \theta_{x|\mathbf{u}} = 1 \quad \text{for all } \mathbf{u}.$$

Αυτός ο περιορισμός μας λέει ότι η επιλογή της τιμής για το  $\theta_{x|u}$  μπορεί να έχει επίδραση στην επιλογή της τιμής για το  $\theta_{x'|u}$ .

Παρ' όλα αυτά όμως, η επιλογή παραμέτρων για διάφορες τιμές  $u$  του  $U$  είναι ανεξάρτητες μεταξύ τους, οπότε και μπορούμε να μεγιστοποιήσουμε κάθε όρο εντός των αγκυλών στην σχέση (2.21) ανεξάρτητα.

Μπορούμε πλέον να αναλύσουμε την συνάρτηση τοπικής πιθανοφάνειας για μια CPD που είναι σε μορφή πίνακα, σε ένα γινόμενο απλών συναρτήσεων πιθανοφάνειας.

Είναι εύκολο να δούμε ότι κάθε μία από αυτές τις συναρτήσεις πιθανοφάνειας είναι μία πολυωνυμική πιθανοφάνεια.

Οι καταμετρήσεις στα δεδομένα για τα διαφορετικά αποτελέσματα  $x$  είναι  $\{N_{\mathbf{u},x} : x \in \text{Val}(X)\}$ .

Χρησιμοποιούμε τις MLE παραμέτρους για ένα πολυώνυμο, οι οποίες είναι

$$\hat{\theta}_{x|\mathbf{u}} = \frac{N_{\mathbf{u},x}}{N_{\mathbf{u}}},$$

όπου χρησιμοποιούμε το γεγονός ότι

$$N_{\mathbf{u}} = \sum_x N_{\mathbf{u},x}.$$

#### 2.4.1.2 Εκτίμηση Bayesian παραμέτρων

Σε πολλές περιπτώσεις, η εκτίμηση των παραμέτρων μέγιστης πιθανοφάνειας δεν είναι εύρωστη (robust) καθώς κάνει "overfit" στα εκπαιδευτικά δεδομένα (training data).

Η Bayesian προσέγγιση χρησιμοποιεί μια προηγούμενη κατανομή επί των παραμέτρων ώστε να εξομαλύνει τις μη κανονικότητες στα εκπαιδευτικά δεδομένα και αυτό την καθιστά πιο εύρωστη.

Όπως θα δούμε στην παράγραφο 2.4.3, το Bayesian πλαίσιο εργασίας μας δίνει επίσης και ένα καλό μέτρο για την αξιολόγηση της ποιότητας των διαφορετικών υποψήφιας δομών.

Χονδρικά, μπορούμε να πούμε ότι η Bayesian προσέγγιση εισάγει μία προηγούμενη κατανομή επί των αγνώστων παραμέτρων η οποία μας δίνει την δυνατότητα να ορίσουμε μία συγκεκριμένη από κοινού κατανομή επί των εν λόγω αγνώστων παραμέτρων και των συγκεκριμένων δεδομένων και εκτελεί Bayesian ρύθμιση (Bayesian conditioning) χρησιμοποιώντας τα δεδομένα ως αποδεικτικό (evidence) προκειμένου να υπολογιστεί μία ύστερη κατανομή αυτών των παραμέτρων.

Ας θεωρήσουμε το εξής παράδειγμα :

Θέλουμε να υπολογίσουμε τις παραμέτρους για ένα απλό δίκτυο δύο μεταβλητών  $X$  και  $Y$  όπου  $X$  είναι ο γονέας του  $Y$ .

Τα εκπαιδευτικά δεδομένα που διαθέτουμε, αποτελούνται από παρατηρήσεις  $x^j, y^j$  για  $j = 1, \dots, m$ .

Υποθέτουμε επίσης ότι οι CPDs αναπαριστώνται ως πολυώνυμα και έχουμε διανύσματα αγνώστων παραμέτρων  $\theta_X, \theta_{Y|X}^0$  και  $\theta_{Y|X}^1$ .

Οι εξαρτήσεις μεταξύ αυτών των μεταβλητών περιγράφονται στο δίκτυο του σχήματος 2.6. Αυτό είναι το meta-Bayesian δίκτυο που περιγράφει τους τρόπους μάθησης.

Η συγκεκριμένη δομή Bayesian δικτύου μας αποκαλύπτει συγκεκριμένα σημεία.

Για παράδειγμα, οι στιγμές είναι ανεξάρτητες, δεδομένων των αγνώστων παραμέτρων ενώ μια συνήθης υπόθεση που γίνεται είναι ότι οι μεμονωμένες μεταβλητές των παραμέτρων είναι a priori ανεξάρτητες.

Δηλαδή, πιστεύουμε ότι γνωρίζοντας την τιμή μίας παραμέτρου δεν μας λέει κάτι για κάποια άλλη, αυτό ονομάζεται ανεξαρτησία παραμέτρων.

Το κατά πόσο ταιριάζει αυτή η υπόθεση εξαρτάται από το πεδίο ορισμού και είναι κάτι το οποίο πρέπει να λαμβάνουμε σοβαρά υπόψη μας.

Αν δεχθούμε την ανεξαρτησία των παραμέτρων, μπορούμε να βγάλουμε ένα σημαντικό συμπέρασμα.

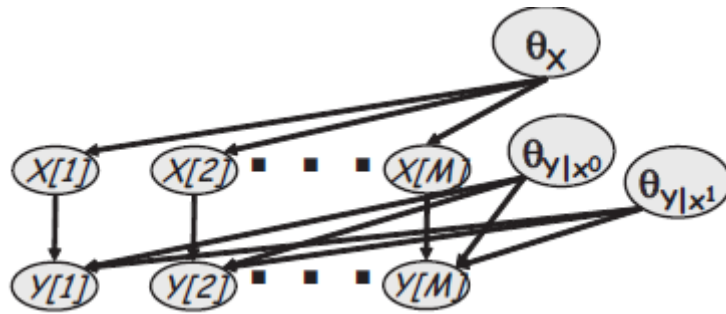
Τα πλήρη δεδομένα κάνουν d-separation των παραμέτρων για διαφορετικές CPDs.

Δεδομένου του συνόλου δεδομένων  $D$ , μπορούμε να προσδιορίσουμε την ύστερη επί της  $\theta_X$  ανεξάρτητα από την ύστερη επί της  $\theta_{Y|X}$ .

Αφού λύσουμε δηλαδή το κάθε πρόβλημα ξεχωριστά, μπορούμε να συνδυάσουμε τα αποτελέσματα.

Αυτό είναι το ανάλογο αποτέλεσμα της αποδόμησης/ανάλυσης πιθανοφάνειας για MLE προσέγγιση όπως είδαμε στην παράγραφο 2.4.1.1.





Σχήμα 2.6 (Το Bayesian δίκτυο για την εκτίμηση παραμέτρων ενός απλού Bayesian δικτύου δύο κόμβων.)

Ας θεωρήσουμε για παράδειγμα την εκπαιδευτική ρύθμιση (learning setting) που περιγράφεται στο σχήμα 2.6 όπου θεωρούμε ότι τα  $X$  και  $Y$  είναι δυαδικά.

Θέλουμε να αναπαραστήσουμε τις ύστερες  $\theta_X$  και  $\theta_{Y|x}$  σύμφωνα με τα δεδομένα που διαθέτουμε.

Εάν χρησιμοποιήσουμε μια προηγούμενη κατανομή Dirichlet επί των  $\theta_X$ ,  $\theta_{Y|x^0}$  και  $\theta_{Y|x^1}$  τότε και η ύστερη  $P(\theta_X | x^1, \dots, x^M)$  μπορεί να αναπαρασταθεί επίσης ως μια κατανομή Dirichlet.

Ας υποθέσουμε ότι  $P(\theta_X)$  είναι μία προηγούμενη κατανομή Dirichlet με υπερπαραμέτρους  $\alpha_x^0$  και  $\alpha_x^1$ ,  $P(\theta_{Y|x^0})$  είναι μια προηγούμενη Dirichlet με υπερπαραμέτρους  $\alpha_{y|x^0}^0$  και  $\alpha_{y|x^0}^1$  και  $P(\theta_{Y|x^1})$  είναι μια προηγούμενη Dirichlet με υπερπαραμέτρους  $\alpha_{y|x^1}^0$  και  $\alpha_{y|x^1}^1$ .

Όπως και στην αποδόμηση/ανάλυση για την συνάρτηση πιθανοφάνειας στην παράγραφο 2.4.1.1, οι όροι πιθανοφάνειας που έχουν το  $\theta_{Y|x}$  εξαρτώνται από όλα τα δεδομένα στοιχεία  $X^j$  έτσι ώστε  $x^j = x^0$  ενώ οι όροι που έχουν  $\theta_{Y|x}$  εξαρτώνται από όλα τα δεδομένα στοιχεία  $X^j$  έτσι ώστε  $x^j = x^1$ .

Μπορούμε να αναλύσουμε την από κοινού κατανομή επί των παραμέτρων και των δεδομένων ως εξής :

$$\begin{aligned}
 P(\theta_G, \mathcal{D}) &= P(\theta_X) L_X(\theta_X : \mathcal{D}) \\
 & P(\theta_{Y|x^1}) \prod_{j: x^j = x^1} P(y^j | x^j : \theta_{Y|x^1}) \\
 & P(\theta_{Y|x^0}) \prod_{j: x^j = x^0} P(y^j | x^j : \theta_{Y|x^0})
 \end{aligned}$$

Έτσι αυτή η από κοινού κατανομή είναι το γινόμενο τριών ξεχωριστών από κοινού κατανομών με προηγούμενη κατανομή Dirichlet για κάποια πολυωνυμική παράμετρο και δεδομένα που αντλήθηκαν από αυτό το πολυώνυμο.

Καταλήγουμε στο συμπέρασμα ότι η ύστερη κατανομή για την  $P(\theta_X | \mathcal{D})$  είναι Dirichlet με υπερπαραμέτρους  $\alpha_x^0 + N_x^0$  και  $\alpha_x^1 + N_x^1$ , η ύστερη για την  $P(\theta_{Y|x^0} | \mathcal{D})$  είναι Dirichlet με υπερπαραμέτρους  $\alpha_{y|x^0}^0 + N_{x^0,y^0}$  και  $\alpha_{y|x^0}^1 + N_{x^0,y^1}$  και η ύστερη για την  $P(\theta_{Y|x^1} | \mathcal{D})$  είναι Dirichlet με υπερπαραμέτρους  $\alpha_{y|x^1}^0 + N_{x^1,y^0}$  και  $\alpha_{y|x^1}^1 + N_{x^1,y^1}$ .

Το ίδιο μοτίβο συλλογισμού εφαρμόζεται και στην γενική περίπτωση.

Έστω  $\mathcal{D}$  ένα πλήρες σύνολο δεδομένων για το  $X$  και έστω  $G$  μια δομή δικτύου επί αυτών των μεταβλητών με CPDs υπό μορφή πίνακα.

Εάν η προηγούμενη κατανομή  $P(\theta_G)$  ικανοποιεί την ανεξαρτησία των παραμέτρων, τότε

$$P(\theta_G | \mathcal{D}) = \prod_i \prod_{\mathbf{pa}_i} P(\theta_{X_i | \mathbf{pa}_i} | \mathcal{D}).$$

Εάν η  $P(\theta_{X|u})$  είναι μια προηγούμενη κατανομή Dirichlet με υπερπαραμέτρους  $\alpha_x^1|u, \dots, \alpha_x^K|u$  τότε η ύστερη  $P(\theta_{X|u} | \mathcal{D})$  είναι μια κατανομή Dirichlet με υπερπαραμέτρους  $\alpha_x^1|u + N_{u,x}^1, \dots, \alpha_x^K|u + N_{u,x}^K$ .

Αυτό συνεπάγεται ένα μοντέλο προβλέψεων στο οποίο για την επόμενη στιγμή έχουμε ότι

$$P(X_i[m+1] = x_i | U[m+1] = \mathbf{u}, \mathcal{D}) = \frac{\alpha_{x_i|u} + N_{x_i, \mathbf{u}}}{\sum_i \alpha_{x_i|u} + N_{x_i, \mathbf{u}}}. \quad (2.22)$$

Συγκεντρώνοντας όλα αυτά μαζί, βλέπουμε ότι προκειμένου να υπολογίσουμε την πιθανότητα μιας νέας στιγμής, μπορούμε να χρησιμοποιήσουμε ένα μόνο δίκτυο παραμετροποιημένο κατά την συνήθη πρακτική, μέσω ενός συνόλου πολυωνύμων που υπολογίζονται σύμφωνα με την σχέση (2.22).

#### 2.4.2 Εκτίμηση παραμέτρων σε δίκτυα Markov

Στην γενική περίπτωση δικτύων Markov, δυστυχώς, η συνάρτηση πιθανοφάνειας δεν δύναται να αναλυθεί.

Εξαίρεση αποτελούν, ως προς αυτήν την δυσκολία, τα χορδικά δίκτυα Markov. Εμείς όμως θα ασχοληθούμε με την γενικότερη περίπτωση.

Για ένα δίκτυο με σύνολο κλικών  $D_1, \dots, D_n$ , η συνάρτηση πιθανοφάνειας δίνεται από την σχέση

$$L(\epsilon, \mathcal{D}) = \prod_{j=1}^m \frac{1}{Z} \exp \left[ - \sum_{i=1}^n \epsilon_i [\mathbf{x}_i^j] \right],$$

όπου  $x_i^j$  είναι η τιμή των παραμέτρων  $D_i$  την στιγμή  $x^j$  και

$$Z = \sum_{\mathbf{x}} \exp \left[ - \sum_{i=1}^n \epsilon_i [\mathbf{x}_i] \right]$$

είναι η σταθερά κανονικοποίησης.

Αυτή η σταθερά κανονικοποίησης ευθύνεται για την σύζευξη (coupling) των εκτιμήσεων των παραμέτρων καθώς και για την αποτελεσματική εξαγωγή λύσεως σε κλειστή μορφή.

Ευτυχώς, η αντικειμενική συνάρτηση (objective function) είναι κοίλη (concave) στο  $\epsilon$ , οπότε έχουμε ένα άνευ περιορισμών κοίλο πρόβλημα μεγιστοποίησης το οποίο μπορεί να λυθεί με απλές μεθόδους ανοδικής κλίσεως (gradient ascent methods) ή μεθόδους δευτέρας τάξεως.

Πιο συγκεκριμένα, για κάθε  $u_i \in \text{Val}(D_i)$  έχουμε μια παράμετρο  $\epsilon_{i,u_i} \in \mathbb{R}$ .

Αυτή είναι και η απλούστερη περίπτωση πλήρους παραμετροποίησης.

Συχνότερα βέβαια, οι παράμετροι μπορεί να είναι μέχρι και μηδενικές.

Κάτι τέτοιο δεν αλλάζει φυσικά την θεμελιώδη πολυπλοκότητα ή την μέθοδο εκτίμησης.

Η παράγωγος της λογαριθμικής πιθανοφάνειας ως προς το  $\epsilon_{i,u_i}$  δίνεται από την σχέση

$$\frac{\partial \log L(\epsilon, \mathcal{D})}{\partial \epsilon_{i,u_i}} = \sum_{j=1}^m \left[ P(\mathbf{u}_i | \epsilon) - \mathbf{1}\{\mathbf{x}_i^j = \mathbf{u}_i\} \right] = mP(\mathbf{u}_i | \epsilon) - N_{\mathbf{u}_i}.$$

Να σημειώσουμε ότι η κλίση (gradient) είναι μηδέν όταν οι καταμετρήσεις (counts) των δεδομένων (data) ανταποκρίνονται πλήρως στον αριθμό των αναμενόμενων καταμετρήσεων που προβλέπονται από το μοντέλο.

Στην πράξη, χρησιμοποιείται μια προηγούμενη (prior) κατανομή επί των παραμέτρων έτσι ώστε να αποφευχθεί το overfitting.

Η πιο συνηθισμένη προηγούμενη κατανομή που χρησιμοποιείται είναι η διαγώνια Γκαουσιανή κατανομή (diagonal Gaussian),  $\epsilon \sim N(0, \sigma^2 \mathbf{I})$ , η οποία προσθέτει έναν επιπλέον παράγοντα  $\frac{\epsilon_{i,u_i}}{\sigma^2}$  στην κλίση.

Για να υπολογίσουμε την πιθανότητα  $P(u_i | \epsilon)$  που χρειάζεται για την αποτίμηση της κλίσεως (gradient) χρειάζεται να κάνουμε συμπερασμό στο δίκτυο Markov.

Σε αντίθεση με τα Bayesian δίκτυα όπου οι παράμετροι των δύσκολα διαχειριζόμενων γράφων (δηλαδή γράφων με μεγάλο treewidth) μπορούν να εκτιμηθούν με απλή καταμέτρηση λόγω της τοπικής κανονικοποίησης, στην περίπτωση των μη κατευθυνόμενων γράφων απαιτείται συμπερασμός ακόμα και κατά το στάδιο της μάθησης.

Αυτό είναι ένα από τα κόστη που μας στοιχίζει η ευελιξία της οικουμενικής κανονικοποίησης σε δίκτυα Markov.

Λόγω αυτής της πρόσθετης πολυπλοκότητας, η μάθηση/εκπαίδευση μέγιστης πιθανοφάνειας ενός δικτύου Markov είναι πολύ πιο ακριβή και λιγότερο εξερευνημένη.

Στην συνέχεια θα εστιάσουμε στα Bayesian δίκτυα.

### 2.4.3 Μαθαίνοντας την δομή Bayesian δικτύου

Θεωρούμε το πρόβλημα της εκμάθησης της δομής ενός Bayesian δικτύου.

Υπάρχουν τρεις μεγάλες κατηγορίες αλγορίθμων για την εκμάθηση δομών BN :

### Προσεγγίσεις που βασίζονται σε περιορισμούς

Αυτές οι προσεγγίσεις βλέπουν ένα Bayesian δίκτυο σαν μία αναπαράσταση ανεξαρτησιών (independencies).

Προσπαθούν να κάνουν δοκιμές για υπό συνθήκη (conditional) εξαρτήσεις και ανεξαρτησίες στα δεδομένα και κατόπιν να βρουν ένα δίκτυο το οποίο να εξηγεί με τον καλύτερο δυνατό τρόπο αυτές τις εξαρτήσεις και τις ανεξαρτησίες.

Οι μέθοδοι που βασίζονται σε περιορισμούς είναι αρκετά διαισθητικοί καθώς ακολουθούν στενά τον ορισμό του Bayesian δικτύου.

Ένα ίσως μειονέκτημα αυτών των μεθόδων είναι ότι μπορεί να γίνουν πολύ ευαίσθητες σε αποτυχίες μεμονομένων δοκιμών ανεξαρτησίας.

### Προσεγγίσεις που βασίζονται σε αποτελέσματα (Score-based)

Αυτές οι προσεγγίσεις βλέπουν ένα Bayesian δίκτυο σαν τον προσδιορισμό ενός στατιστικού μοντέλου και κατόπιν πραγματοποιούν μάθηση σαν να είναι ένα πρόβλημα επιλογής μοντέλου.

Όλα αυτά γίνονται βασιζόμενα στην ίδια αρχή :

Ορίζουμε ως υπόθεση έναν χώρο που περιλαμβάνει όλα τα δυνατά μοντέλα (δηλαδή το σύνολο των δυνατών δικτυακών δομών που επιθυμούμε να λάβουμε υπόψη μας) καθώς και μια συνάρτηση αποτελεσμάτων η οποία δείχνει το πόσο καλά ταιριάζει το μοντέλο στα παρατηρούμενα δεδομένα.

Ο σκοπός μας σε υπολογιστικό επίπεδο είναι να βρούμε την δικτυακή δομή που έχει το υψηλότερο score.

Ο χώρος των Bayesian δικτύων είναι ένας συνδυαστικός χώρος που αποτελείται από έναν υπερεκθετικό αριθμό δομών ( $2^{O(n^2)}$ ).

Γι' αυτό, ακόμα και όταν έχουμε μια συνάρτηση αποτελεσμάτων δεν είναι ξεκάθαρο το πώς μπορούμε να βρούμε το δίκτυο με τα περισσότερα αποτελέσματα.

Υπάρχουν πολύ ειδικές περιπτώσεις όπου μπορούμε να βρούμε το βέλτιστο δίκτυο.

Γενικώς όμως, αυτό το πρόβλημα είναι NP δυσκολίας και καταφεύγουμε σε ευρετικές τεχνικές αναζήτησης.

Οι μέθοδοι που στηρίζονται στα αποτελέσματα λαμβάνουν απευθείας υπόψη τους ολόκληρη την δομή και γι' αυτό είναι λιγότερο ευαίσθητες σε μεμονωμένες αστοχίες οπότε και καλύτερες στο να κάνουν συμβιβασμούς μεταξύ του ποιές μεταβλητές στα δεδομένα είναι εξαρτώμενες και του κόστους που επιφέρει η πρόσθεση της ακμής.

Το μειονέκτημα αυτών των προσεγγίσεων είναι ότι γενικώς δεν είναι εγγυημένο ότι θα μας δώσουν την βέλτιστη λύση.

### Προσεγγίσεις μέσου όρου για Bayesian μοντέλα

Στην τρίτη κατηγορία προσεγγίσεων δεν επιχειρείται η εκμάθηση μίας και μόνο δομής.

Βασίζεται σε ένα Bayesian πλαίσιο εργασίας όπου περιγράφεται μία κατανομή επί των δυνατών δομών και γίνεται προσπάθεια να βρεθεί ο μέσος όρος όλων των πιθανών δομών που προβλέπονται.

Εφόσον ο αριθμός των δομών είναι τεράστιος, η εκτέλεση μιας τέτοιας εργασίας μοιάζει αδύνατη.

Για ορισμένες κατηγορίες μοντέλων αυτό μπορεί να γίνει αποδοτικά ενώ για άλλες πρέπει να καταφύγουμε σε άλλες προσεγγίσεις.

Σε αυτό το κεφάλαιο, θα επικεντρωθούμε στην δεύτερη κατηγορία προσεγγίσεων (τις βασιζόμενες σε αποτελέσματα)

### 2.4.3.1 Αποτελέσματα δομής

Όπως είδαμε στα προηγούμενα, οι μέθοδοι που στηρίζονται στα αποτελέσματα βλέπουν το πρόβλημα μάθησης της δομής σαν ένα πρόβλημα βελτιστοποίησης.

Ορίζουμε μια συνάρτηση αποτελεσμάτων που μπορεί να δίνει αποτέλεσμα (score) σε κάθε υποψήφια δομή σε σχέση με τα εκπαιδευτικά δεδομένα και κατόπιν να γίνεται αναζήτηση για την δομή με τα περισσότερα αποτελέσματα.

Όπως είναι αναμενόμενο, μια από τις σημαντικότερες αποφάσεις που πρέπει να λάβουμε είναι η επιλογή της συνάρτησης αποτελεσμάτων.

Στα αμέσως επόμενα, θα συζητήσουμε για δύο από τις πιο προφανείς επιλογές που μπορούμε να κάνουμε.

#### Αποτέλεσμα πιθανοφάνειας

Μια φυσική επιλογή ως συνάρτηση αποτελεσμάτων είναι η συνάρτηση πιθανοφάνειας που χρησιμοποιήσαμε για την εκτίμηση παραμέτρων.

Με αυτό τον τρόπο μετράται η πιθανότητα των δεδομένων όταν είναι γνωστό το μοντέλο, οπότε καταλαβαίνουμε διαισθητικά ότι θα πρέπει να βρεθεί ένα μοντέλο που να καθιστά τα δεδομένα όσο πιο πιθανά γίνεται.

Ας υποθέσουμε ότι θέλουμε να μεγιστοποιήσουμε την πιθανοφάνεια του μοντέλου.

Σκοπός μας είναι να βρούμε τον γράφο  $G$  καθώς και τις παραμέτρους  $\theta_G$  που μεγιστοποιούν την πιθανοφάνεια.

Είναι εύκολο να δείξουμε ότι προκειμένου να βρούμε το διατεταγμένο ζεύγος μέγιστης πιθανοφάνειας  $(G, \theta_G)$  ,πρέπει να βρούμε την δικτυακή δομή  $G$  που επιτυγχάνει την υψηλότερη πιθανοφάνεια όταν χρησιμοποιούμε τις MLE παραμέτρους για τον  $G$ .

Για τον λόγο αυτό ορίζουμε

$$\text{score}_L(G : \mathcal{D}) = \ell((G, \hat{\theta}_G) : \mathcal{D}),$$

όπου

$$\ell((G, \hat{\theta}_G) : \mathcal{D})$$

είναι ο λογάριθμος της συνάρτησης πιθανοφάνειας και  $\hat{\theta}_G$  είναι οι παράμετροι μέγιστης πιθανοφάνειας για τον  $G$ .

(Σε πρακτικό επίπεδο είναι ευκολότερος ο χειρισμός του λογαρίθμου της πιθανοφάνειας.)

Το πρόβλημα με τα αποτελέσματα πιθανοφάνειας είναι ότι κάνουν overfit στα εκπαιδευτικά δεδομένα.

Το μοντέλο εκπαιδεύεται ώστε να ταιριάζει ακριβώς στις προδιαγραφές της εμπειρικής κατανομής του εκπαιδευτικού συνόλου που διαθέτουμε.

Το μοντέλο αυτό δείχνει τόσο τις εξαρτήσεις που είναι αληθείς για την διέπουσα κατανομή όσο και τις εξαρτήσεις που είναι δημιουργήματα του συγκεκριμένου συνόλου στιγμών (instances) που δόθηκε ως εκπαιδευτικά δεδομένα.

Για τον λόγο αυτό, αποτυγχάνει να κάνει καλή γενίκευση με νέες περιπτώσεις δεδομένων τα οποία δειγματοληπτούνται από την διέπουσα κατανομή η οποία όμως δεν είναι πανομοιότυπη με την εμπειρική κατανομή εντός του εκπαιδευτικού συνόλου μας.

Παρ' όλα αυτά όμως, είναι λογικό να χρησιμοποιηθούν τα αποτελέσματα μεγίστης πιθανοφάνειας όταν υπάρχουν και επιπλέον μηχανισμοί που δεν επιτρέπουν υπερβολικά περίπλοκες δομές.

### Bayesian αποτελέσματα

Μια εναλλακτική συνάρτηση αποτελεσμάτων βασίζεται στην Bayesian λογική.

Η κύρια λογική στην Bayesian προσέγγιση είναι ότι όποτε έχουμε αβεβαιότητα για οτιδήποτε θα πρέπει να θέτουμε μια κατανομή επί αυτού.

Στην συγκεκριμένη περίπτωση, έχουμε αβεβαιότητα τόσο για την δομή όσο και για τις παραμέτρους.

Για τον λόγο αυτό θα ορίσουμε μια πρότερη δομή  $P(G)$  η οποία δίνει μια πρότερη πιθανότητα για μια διαφορετική επιλογή παραμέτρων από την στιγμή που δίνεται ο γράφος.

Από τον κανόνα του Bayes έχουμε

$$P(G | D) = \frac{P(D | G)P(G)}{P(D)},$$

όπου ως συνήθως, ο παρανομαστής είναι απλώς ένας παράγοντας κανονικοποίησης ο οποίος δεν βοηθά στην διάκριση μεταξύ διαφορετικών δομών.

Ορίζουμε το Bayesian αποτέλεσμα (Bayesian score) ως

$$\text{score}_B(G : D) = \log P(D | G) + \log P(G), \quad (2.23)$$

Η ικανότητα να αποδίδουμε μία πρότερη πιθανοτική κατανομή επί των δομών, μας δίνει έναν τρόπο σύμφωνα με τον οποίο μπορούμε να προτιμήσουμε κάποιες δομές έναντι άλλων.

Για παράδειγμα, μπορούμε να αποφεύγουμε τις πυκνές δομές καταλογίζοντας τους κάποιο είδος "αρνητικού αποτελέσματος" και να προτιμούμε τις αραιές δομές.

Προκύπτει όμως, ότι αυτός ο όρος στο αποτέλεσμα (score) είναι σχεδόν άσχετος συγκρινόμενος με τον δεύτερο όρο.

Ο πρώτος όρος,  $P(D|G)$ , λαμβάνει υπόψη του την αβεβαιότητα επί των παραμέτρων

$$P(\mathcal{D} | \mathcal{G}) = \int_{\Theta_{\mathcal{G}}} P(\mathcal{D} | \theta_{\mathcal{G}}, \mathcal{G}) P(\theta_{\mathcal{G}} | \mathcal{G}) d\theta_{\mathcal{G}}, \quad (2.24)$$

Όπου  $P(\mathcal{D} | \theta_{\mathcal{G}}, \mathcal{G})$  είναι η πιθανοφάνεια των δεδομένων έχοντας γνωστό το δίκτυο  $\langle \mathcal{G}, \theta_{\mathcal{G}} \rangle$  και  $P(\theta_{\mathcal{G}} | \mathcal{G})$  είναι η προηγούμενη κατανομή επί των διαφορετικών παραμέτρων των τιμών για το δίκτυο  $\mathcal{G}$ .

Αυτός ο όρος είναι η περιθώρια πιθανοφάνεια (marginal likelihood) των δεδομένων όταν είναι γνωστή η δομή από την στιγμή περιθωριοποιούμε τις άγνωστες παραμέτρους.

Να σημειωθεί ότι η περιθώρια πιθανοφάνεια είναι διαφορετική από τα αποτελέσματα μεγίστης πιθανοφάνειας.

Αν και οι δύο όροι εξετάζουν την πιθανοφάνεια των δεδομένων έχοντας γνωστή την δομή, το αποτέλεσμα μεγίστης πιθανοφάνειας μας δίνει το μέγιστο αυτής της συνάρτησης ενώ σε αντίθεση, η περιθώρια πιθανοφάνεια είναι η μέση τιμή της ίδιας συνάρτησης, όπου για αυτή την μέση κατάσταση βασιστήκαμε στο προηγούμενο μέτρο  $P(\theta_{\mathcal{G}} | \mathcal{G})$ .

Προκειμένου να πάμε ένα βήμα παραπέρα θεωρούμε ένα δίκτυο το οποίο να έχει Dirichlet προηγούμενες κατανομές, τέτοιο ώστε

$$P(\theta_{X_i | \text{pa}_i} | \mathcal{G})$$

με υπερπαραμέτρους

$$\{\alpha_{x_i^j | \mathbf{u}_i}^{\mathcal{G}} : j = 1, \dots, |X_i|\}$$

Οπότε έχουμε

$$P(\mathcal{D} | \mathcal{G}) = \prod_i \prod_{\mathbf{u}_i \in \text{Val}(\text{Pa}_{X_i}^{\mathcal{G}})} \frac{\Gamma(\alpha_{X_i | \mathbf{u}_i}^{\mathcal{G}})}{\Gamma(\alpha_{X_i | \mathbf{u}_i}^{\mathcal{G}} + N_{\mathbf{u}_i})} \prod_{x_i^j \in \text{Val}(X_i)} \left[ \frac{\Gamma(\alpha_{x_i^j | \mathbf{u}_i}^{\mathcal{G}} + N_{x_i^j, \mathbf{u}_i})}{\Gamma(\alpha_{x_i^j | \mathbf{u}_i}^{\mathcal{G}})} \right],$$

όπου

$$\alpha_{X_i | \mathbf{u}_i}^{\mathcal{G}} = \sum_j \alpha_{x_i^j | \mathbf{u}_i}^{\mathcal{G}}.$$

Στην πράξη χρησιμοποιούμε τον λογάριθμο αυτής της σχέσης επειδή είναι ευκολότερος ο χειρισμός κατά τους αριθμητικούς υπολογισμούς.

Τα Bayesian scores έχουν την τάση να μας οδηγούν σε απλούστερες δομές αλλά καθώς εισέρχονται νέα επιπλέον δεδομένα είναι σε θέση να αναγνωρίσουν ότι θα χρειαστεί κάποια πιο περίπλοκη δομή.

Δηλαδή με άλλα λόγια, "ανταλλάζει" το ταίριασμα των δεδομένων με την πολυπλοκότητα του μοντέλου.



Προκειμένου να γίνει περισσότερο κατανοητή αυτή η συμπεριφορά θα θεωρήσουμε μία προσέγγιση για τα Bayesian αποτελέσματα που θα μας αναδείξει κάπως καλύτερα τις βασικές ιδιότητες τους.

Εάν χρησιμοποιήσουμε μια προηγούμενη κατανομή Dirichlet για όλες τις παραμέτρους του δικτύου μας, τότε καθώς  $M \rightarrow \infty$ , έχουμε ότι

$$\log P(\mathcal{D} | \mathcal{G}) = \ell(\hat{\theta}_{\mathcal{G}} : \mathcal{D}) - \frac{\log M}{2} \text{Dim}[\mathcal{G}] + O(1),$$

όπου  $\text{Dim}[\mathcal{G}]$  είναι ο αριθμός των ανεξάρτητων παραμέτρων στον  $\mathcal{G}$ .

Από αυτό βλέπουμε ότι τα Bayesian αποτελέσματα τείνουν πράγματι να "ανταλλάσσουν" την πιθανοφάνεια (ταίριασμα στα δεδομένα) από την μία και την πολυπλοκότητα του μοντέλου από την άλλη.

Αυτή η προσέγγιση ονομάζεται "αποτελέσματα Bayesian κριτηρίου πληροφορίας" (Bayesian information criterion-BIC-score) :

$$\text{score}_{BIC}(\mathcal{G} : \mathcal{D}) = \ell(\hat{\theta}_{\mathcal{G}} : \mathcal{D}) - \frac{\log M}{2} \text{Dim}[\mathcal{G}]$$

Αυτό που πρέπει να ορίσουμε στην συνέχεια είναι οι προηγούμενες κατανομές που χρησιμοποιούνται για τα Bayesian αποτελέσματα.

Για την περίπτωση των προηγούμενων των δικτυακών δομών,  $P(\mathcal{G})$ , να σημειωθεί ότι παρόλο που αυτός ο όρος μοιάζει να περιγράφει την μεροληψία μας για μια συγκεκριμένη δομή, στην πραγματικότητα παίζει σχετικά μικρό ρόλο.

Όπως μπορούμε να δούμε στο θεώρημα 2.35, ο λογάριθμος της περιθώριας πιθανοφάνειας μεγαλώνει γραμμικά με τον αριθμό των παραδειγμάτων ενώ η προηγούμενη επί των δομών παραμένει σταθερή.

Συνεπώς, η προηγούμενη κατανομή επί των δομών δεν παίζει σημαντικό ρόλο κατά την ασυμπτωτική ανάλυση εφόσον δεν αποκλείει κάποια δομή (πχ. με την ανάθεση μηδενικής πιθανότητας σε αυτή).

Εν μέρει εξαιτίας αυτού, είναι σύνηθες να χρησιμοποιείται μια ομοιόμορφη (uniform) προηγούμενη κατανομή επί των δομών.

Η προηγούμενη κατανομή των δομών όμως μπορεί να κάνει κάποια διαφορά στην περίπτωση που θεωρήσουμε μικρά δείγματα οπότε και μπορεί να θέλουμε να κωδικοποιήσουμε κάποιες προτιμήσεις μας μέσα σε αυτήν την προηγούμενη κατανομή.

Για παράδειγμα, μπορεί να βαθμολογούμε αρνητικά κάποιες ακμές στον γράφο και να χρησιμοποιήσουμε προηγούμενη κατανομή

$$P(\mathcal{G}) \propto c^{|\mathcal{G}|},$$

όπου  $c$  κάποια σταθερά μικρότερη από 1 και  $|\mathcal{G}|$  είναι ο αριθμός των ακμών του γράφου.



Βλέπουμε ότι και για τις δύο επιλογές (την ομοιόμορφη κατανομή και την κατανομή με αρνητική βαθμολογία) αρκεί να χρησιμοποιήσουμε μία τιμή που να είναι ανάλογη (proportional) της προηγούμενης κατανομής αφού η σταθερά κανονικοποίησης είναι η ίδια για οποιαδήποτε επιλογή του  $G$  και κατά συνέπεια μπορεί να αγνοηθεί.

Είναι βολικό σε μαθηματικό επίπεδο, να υποθέσουμε ότι η προηγούμενη κατανομή των δομών ικανοποιεί την συνθήκη περί σπονδυλωτής δομής.

Αυτή η συνθήκη απαιτεί η προηγούμενη κατανομή  $P(G)$  να είναι ανάλογη προς ένα γινόμενο όρων, όπου ο κάθε όρος σχετίζεται και με μία οικογένεια.

Τυπικά γράφουμε,

$$P(\mathcal{G}) \propto \prod_i P(\mathbf{Pa}_{X_i} = \mathbf{Pa}_{X_i}^{\mathcal{G}}),$$

όπου με

$$P(\mathbf{Pa}_{X_i} = \mathbf{Pa}_{X_i}^{\mathcal{G}})$$

συμβολίζουμε την προηγούμενη πιθανότητα που έχει ανατεθεί για την επιλογή του συγκεκριμένου συνόλου γονέων για το  $X_i$ .

Οι προηγούμενες κατανομές δομών που ικανοποιούν αυτήν την ιδιότητα δεν αποδίδουν αρνητική βαθμολογία στις οικογενειακές ιδιότητες του γράφου (όπως είναι για παράδειγμα το βάθος του) παρά μόνο στις τοπικές ιδιότητες (όπως είναι για παράδειγμα το πλήθος των indegrees).

Στην συνέχεια θέλουμε να αναπαραστήσουμε τις προηγούμενες κατανομές των παραμέτρων.

Ο αριθμός των δυνατών δομών είναι υπερεκθετικός και αυτό μας δυσκολεύει στο να μπορέσουμε να εκμαιεύσουμε ξεχωριστές παραμέτρους για κάθε μία από αυτές.

Μία απλή προσέγγιση είναι να πάρουμε μία σταθερή κατανομή Dirichlet, πχ. Dirichlet  $(\alpha, \alpha, \alpha, \dots, \alpha)$  για κάθε παράμετρο, όπου  $\alpha$  είναι μία προκαθορισμένη σταθερά. Μία τυπική τιμή είναι η  $\alpha = 1$ .

Αυτή η προηγούμενη συνήθως αναφέρεται σαν η  $K2$  προηγούμενη, κάνοντας αναφορά στο όνομα του συστήματος όπου χρησιμοποιήθηκε για πρώτη φορά.

Μία πιο εκλεπτυσμένη προσέγγιση είναι αυτή που ονομάζεται BDe προηγούμενη κατανομή.

Εκμαιεύουμε μια προηγούμενη κατανομή  $P'$  επί ολόκληρου του χώρου πιθανοτήτων καθώς και ένα ισοδύναμο μέγεθος δείγματος  $M'$  για το σύνολο των ιδεατών δειγμάτων.

Στην συνέχεια καθορίζουμε τις παραμέτρους ως εξής :

$$\alpha_{x_i | \mathbf{pa}_i} = M' \cdot P'(x_i, \mathbf{pa}_i).$$

Με αυτή την επιλογή καταφέρνουμε να αποφύγουμε συγκεκριμένες ασυνέπειες που παρουσιάζονται από την  $K2$  προηγούμενη κατανομή.

Μπορούμε να αναπαραστήσουμε την  $P'$  σαν ένα Bayesian δίκτυο, του οποίου η δομή μπορεί να αναπαραστήσει την προηγούμενη μας κατανομή η οποία αφορά την δομή του τομέα.

Απλούστατα, εάν δεν έχουμε στην διάθεση μας πρότερη γνώση θέτουμε την  $P'$  να είναι η ομοιόμορφη κατανομή. Δηλαδή το κενό Bayesian δίκτυο με μία ομοιόμορφη περιθώρια κατανομή για κάθε μεταβλητή.

Τα BDe αποτελέσματα φαίνεται ότι ικανοποιούν μια σημαντική ιδιότητα.

Δύο δίκτυα λέγονται I-ισοδύναμα (I-equivalent) εάν κωδικοποιούν το ίδιο σύνολο ανεξάρτητων δηλώσεων (statements).

Συνεπώς βασιζόμενοι σε παρατηρούμενες ανεξαρτησίες δεν μπορούμε να διακρίνουμε/ξεχωρίσουμε I-ισοδύναμα δίκτυα μεταξύ τους.

Αυτό σημαίνει ότι βασιζόμενοι σε παρατηρήσεις διαφόρων περιπτώσεων δεδομένων, δεν περιμένουμε να ξεχωρίσουμε (δηλαδή να βρούμε διαφορές) μεταξύ ισοδύναμων δικτύων.

Τα BDe αποτελέσματα διαθέτουν την επιθυμητή ιδιότητα σύμφωνα με την οποία, τα I-ισοδύναμα δίκτυα έχουν τα ίδια αποτελέσματα ή αλλιώς ότι είναι ισοδύναμα αποτελεσματικά (score-equivalent).

### 2.4.3.2 Αναζήτηση

Πλέον έχουμε ένα καλά ορισμένο πρόβλημα βελτιστοποίησης.

Οι είσοδοι μας είναι :

1. Το εκπαιδευτικό σύνολο  $D$ .
2. Η συνάρτηση αποτελεσμάτων (scoring function), η οποία θα περιλαμβάνει προηγούμενες κατανομές εάν αυτό είναι απαραίτητο.
3. Ένα σύνολο  $G$  δυνατών δικτυακών δομών στο οποίο ενσωματώνεται κάθε προηγούμενη γνώση.

Η επιθυμητή έξοδος είναι μία δικτυακή δομή (από το σύνολο των δυνατών δομών) η οποία να μεγιστοποιεί τα αποτελέσματα.

Διαφαίνεται ότι για αυτή την συζήτηση, μπορούμε να αγνοήσουμε την συγκεκριμένη επιλογή αποτελεσμάτων.

Οι αλγόριθμοι αναζήτησης που διαθέτουμε θα εφαρμοστούν αμετάλλακτοι και στις τρεις περιπτώσεις αποτελεσμάτων.

Μία σημαντική ιδιότητα των αποτελεσμάτων η οποία επηρεάζει την επάρκεια της αναζήτησης είναι η ιδιότητα της αποδόμησης/ανάλυσης τους (decomposability).

Ένα αποτέλεσμα (score) έχει την δυνατότητα να αποδομηθεί εάν μπορούμε να γράψουμε το αποτέλεσμα μιας δικτυακής δομής  $G$  ως εξής:

$$\text{score}(G : D) = \sum_i \text{FamScore}(X_i | \mathbf{Pa}_i^G : D)$$

Όλα τα αποτελέσματα (scores) που έχουμε θεωρήσει ως τώρα είναι αποδομήσιμα.

Μια άλλη ιδιότητα που έχουν κοινή όλα αυτά τα αποτελέσματα είναι η λεγόμενη ισοδυναμία αποτελεσμάτων (score equivalence), σύμφωνα με την οποία εάν ο  $G$  είναι

ισοδύναμα ανεξάρτητος (independence-equivalent) ως προς τον  $G'$  τότε ισχύει ότι  $\text{score}(G : D) = \text{score}(G' : D)$ .

Υπάρχουν πολλές ειδικές περιπτώσεις όπου η εκμάθηση της δομής είναι εύκολα διαχειριζόμενη.

Θα αναφέρουμε δύο σημαντικές περιπτώσεις, χωρίς να εισέλθουμε σε ιδιαίτερες λεπτομέρειες :

(1) Την εκμάθηση δικτυακών δομών που έχουν μορφή δέντρου.

(2) Την εκμάθηση δικτύων που έχουν γνωστό τρόπο διάταξης επί των μεταβλητών.

Ένα δίκτυο έχει δομή δέντρου (tree-structured) εάν κάθε μεταβλητή έχει το πολύ ένα γονέα.

Σε αυτή την περίπτωση, δηλαδή για αποδομήσιμα (decomposable) και αποτελεσματικά ισοδύναμα (score-equivalent) αποτελέσματα (scores), μπορούμε να φτιάξουμε έναν μη προσανατολισμένο γράφο όπου το βάρος πάνω από μια ακμή  $X_i \rightarrow X_j$  είναι η αλλαγή που υπηρεύεται στα αποτελέσματα του δικτύου εάν προσθέσουμε την  $X_i$  σαν τον γονέα της  $X_j$ .

Να σημειωθεί ότι λόγω της ιδιότητας της ισοδυναμίας αποτελεσμάτων, στην περίπτωση που προσθέταμε την  $X_j$  ως γονέα της  $X_i$  θα είχαμε ακριβώς την ίδια αλλαγή στα αποτελέσματα του δικτύου.

Μπορούμε για αυτόν τον γράφο, να βρούμε ένα εκτεινόμενο δέντρο, με βάρη επί των ακμών του, σε πολυωνυμικό χρόνο.

Έχουμε επίσης την δυνατότητα να μετασχηματίσουμε το μη κατευθυνόμενο εκτεινόμενο δέντρο σε ένα κατευθυνόμενο εκτεινόμενο δέντρο επιλέγοντας τυχαία μια ρίζα και βάζοντας τις ακμές να απομακρύνονται από αυτήν.

Μία άλλη ενδιαφέρουσα και εύκολη στον χειρισμό περίπτωση είναι το πρόβλημα της εκμάθησης μιας BN δομής η οποία να είναι συνεπής ως προς κάποια γνωστή συνολική διάταξη επί του  $X$  και η οποία έχει φραγμένο  $d$ .

Με άλλα λόγια, περιορίζουμε την προσοχή μας σε δομές  $G$  όπου εάν  $X_i \in \text{Pa}_{X_j}^G$  τότε  $X_i < X_j$  και  $|\text{Pa}_{X_j}^G| < d$ .

Για μερικές περιοχές το να βρεθεί μια τέτοιου τύπου διάταξη είναι σχετικά εύκολο.

Σαν παράδειγμα μπορούμε να θεωρήσουμε την ροή κυριαρχίας πάνω στην σειρά με την οποία οι μεταβλητές θα λάβουν τις τιμές τους.

Σε αυτήν την περίπτωση, για κάθε  $X_i$  μπορούμε να αξιολογήσουμε κάθε δυνατό σύνολο γονέων μεγέθους  $d$  από το  $\{X_1, \dots, X_{i-1}\}$ .

Κάτι τέτοιο είναι πολυωνυμικό στο  $n$  αλλά εκθετικό στο  $d$ .

Δυστυχώς, η γενική περίπτωση εύρεσης ενός  $G^*$  που να έχει τα βέλτιστα αποτελέσματα για φραγμένο βαθμό (degree)  $d \geq 2$ , είναι πρόβλημα δυσκολίας NP (NP-hard).

Αντί να στοχεύουμε σε ένα αλγόριθμο ο οποίος θα βρίσκει πάντοτε το δίκτυο με τα μεγαλύτερα αποτελέσματα, καταφεύγουμε σε ευρετικούς αλγόριθμους οι οποίοι επιχειρούν να εντοπίσουν το καλύτερο δίκτυο, χωρίς όμως αυτό να είναι εγγυημένο ότι θα επιτευχθεί κίολας.

Για να ορίσουμε τον ευρετικό αλγόριθμο αναζήτησης, θα πρέπει να ορίσουμε τον χώρο αναζήτησης και την διαδικασία αναζήτησης.

Μπορούμε να φανταστούμε τον χώρο αναζήτησης σαν ένα γράφο, στον οποίο κάθε κόμβος είναι και μία υπονήφια προς εξέταση δικτυακή δομή ενώ οι ακμές του δείχνουν δυνατές "κινήσεις" που μπορεί να κάνει η διαδικασία αναζήτησης.

Η διαδικασία αναζήτησης ορίζει έναν αλγόριθμο ο οποίος εξερευνά τον χώρο αναζήτησης χωρίς απαραίτητως να τον δει όλο.

Η απλούστερη διαδικασία αναζήτησης είναι η άπληστη, κατά την οποία όποτε είναι ένας κόμβος, επιλέγει να μετακινήσει τον γειτονικό αυτού που έχει το μεγαλύτερο

score μέχρι να φτάσει σε κάποιον κόμβο ο οποίος έχει καλύτερο αποτέλεσμα από όλους τους γειτονικούς του.

Προκειμένου να κάνουμε πιο λεπτομερή και προσεγμένη ανάλυση, θεωρούμε ότι στην περίπτωση μας ένας κόμβος του χώρου αναζήτησης είναι μία πλήρης δικτυακή δομή  $G$  επί του  $X$ .

Υπάρχει συμβιβασμός στο πόσο πυκνά συνδέεται κάθε κόμβος με το πόσο αποτελεσματική θα είναι η αναζήτηση.

Εάν κάθε κόμβος έχει λίγους γειτονικούς, τότε η διαδικασία έχει να λάβει υπόψη της λίγες επιλογές σε κάθε σημείο της αναζήτησης. Συνεπώς, έχει την δυνατότητα να αξιολογήσει κάθε επιλογή.

Παρ' όλα αυτά όμως, μονοπάτια από τον αρχικό κόμβο προς έναν καλό κόμβο μπορεί να είναι πολύ μεγάλα σε μήκος και ταυτοχρόνως σύνθετα.

Από την άλλη πάλι, εάν κάθε κόμβος έχει πολλούς γείτονες τότε θα υπάρχουν μικρά σε μήκος μονοπάτια από κάθε σημείο προς ένα άλλο, αλλά ίσως να μην μπορούμε να τα επιλέξουμε επειδή δεν έχουμε τον χρόνο προκειμένου να αξιολογούμε όλες τις επιλογές σε κάθε βήμα.

Ένας καλός συμβιβασμός για αυτό το πρόβλημα είναι να επιλεγθεί ένας λογικά μικρός αριθμός γειτόνων για κάθε κόμβο προσέχοντας όμως ώστε η "διάμετρος" του χώρου αναζήτησης να παραμένει μικρή.

Μια φυσική επιλογή γειτόνων για μια δικτυακή δομή είναι ένα σύνολο δομών που είναι όμοιες με αυτήν, με εξαίρεση κάποιες μικρές "τοπικές" τροποποιήσεις.

Οι συνηθέστερα χρησιμοποιούμενες πράξεις που ορίζουν αυτές τις τοπικές τροποποιήσεις είναι :

1. Η προσθήκη/πρόσθεση μίας ακμής.
2. Η απαλοιφή/διαγραφή μίας ακμής.
3. Η αντιστροφή μίας ακμής.

Με άλλα λόγια, αν θεωρήσουμε τον κόμβο  $G$ , τότε οι κόμβοι που γειτονεύουν με αυτόν στον χώρο αναζήτησης είναι αυτοί στους οποίους αλλάζουμε μία ακμή είτε προσθέτοντας, είτε αφαιρώντας, είτε αντιστρέφοντας τον προσανατολισμό κάποιας.

Θεωρούμε μόνο πράξεις που έχουν ως αποτέλεσμα επιτρεπτά δίκτυα (πχ. ακυκλικά δίκτυα που ικανοποιούν οποιουσδήποτε περιορισμούς όπως είναι για παράδειγμα ο δεσμευμένος βαθμός πρόσπτωσης - indegree).

Αυτός ο ορισμός του χώρου αναζήτησης είναι αρκετά φυσικός και διαθέτει πολλές επιθυμητές ιδιότητες.

Καταρχήν πρέπει να σημειωθεί ότι η διάμετρος του χώρου αναζήτησης είναι το πολύ  $n^2$ .

Αυτό σημαίνει ότι υπάρχει ένα σχετικά μικρό μονοπάτι μεταξύ δύο οποιωνδήποτε δικτύων που επιλέγουμε.

Για να φανεί καλύτερα αυτό ας παρατηρήσουμε ότι αν θεωρήσουμε την διάσχιση ενός μονοπατιού από τον  $G_1$  στον  $G_2$ , μπορούμε να ξεκινήσουμε αφαιρώντας όλες τις ακμές στον  $G_1$  οι οποίες δεν εμφανίζονται στον  $G_2$  και μετά μπορούμε να προσθέσουμε τις ακμές που είναι στον  $G_2$  αλλά δεν υπάρχουν στον  $G_1$ .

Προφανώς, ο αριθμός των βημάτων που κάνουμε φράσσεται από τον συνολικό αριθμό των ακμών που διαθέτουμε, δηλαδή  $n^2$ .

Δεύτερον, υπενθυμίζουμε ότι τα αποτελέσματα ενός δικτύου  $G$  είναι το άθροισμα των τοπικών αποτελεσμάτων.

Οι πράξεις που θεωρούμε, έχουν ως αποτέλεσμα την αλλαγή ενός μόνο όρου τοπικών αποτελεσμάτων (για την περίπτωση της πρόσθεσης ή της αφαίρεσης μίας ακμής) ή δύο όρων (για την περίπτωση της αντιστροφής ακμής).

Συνεπώς, έχουν ως αποτέλεσμα μία τοπική μεταβολή στα αποτελέσματα ενώ η "κύρια" μάζα των αποτελεσμάτων παραμένει σταθερή.

Αυτό σημαίνει ότι υπάρχει κάποιου είδους "συνέχεια" στα αποτελέσματα γειτονικών κόμβων.

Οι μέθοδοι αναζήτησης που χρησιμοποιούνται συνήθως είναι διαδικασίες τοπικής αναζήτησης (local search procedures).

Αυτού του είδους οι διαδικασίες χαρακτηρίζονται από τον εξής σχεδιασμό: διατηρούν έναν "τρέχον" υποψήφιο (current candidate) κόμβο.

Σε κάθε επανάληψη εξερευνούν κάποιους γειτονικούς κόμβους και κατόπιν αποφασίζουν να κάνουν ένα "βήμα" προς έναν από αυτούς τους γείτονες και να τον κάνουν τον τρέχον υποψήφιο.

Αυτές οι επαναλήψεις συνεχίζονται μέχρι να υπάρξει κάποια συνθήκη αποκοπής.

Με άλλα λόγια, οι διαδικασίες τοπικής αναζήτησης μπορεί να θεωρηθεί ότι διατηρούν έναν δείκτη μέσα στον χώρο αναζήτησης, τον οποίο τον μετακινούν σε διάφορες θέσεις.

Μια από τις απλούστερες και συχνότερα χρησιμοποιούμενες διαδικασίες αναζήτησης είναι η διαδικασία άπληστης ανάβασης λόφου (greedy hill-climbing procedure).

Όπως αναφέρει το όνομα της, σε κάθε βήμα κάνουμε την κίνηση που θα μας οδηγήσει προς την μεγαλύτερη βελτίωση των αποτελεσμάτων.

Οι λεπτομέρειες αυτής της διαδικασίας φαίνονται στο σχήμα 2.7.

```

Procedure Greedy-Structure-Search (
     $\mathcal{G}_0$ , // initial network structure
     $\mathcal{D}$  // Fully observed dataset
    score, // Score
     $\mathcal{O}$ , // A set of search operators
)
1   $\mathcal{G}_{best} \leftarrow \mathcal{G}_0$ 
2  do
3     $\mathcal{G} \leftarrow \mathcal{G}_{best}$ 
4    Progress  $\leftarrow$  false
5    for each operator  $o \in \mathcal{O}$ 
6       $\mathcal{G}_o \leftarrow o(\mathcal{G})$  // Result of applying  $o$  on  $\mathcal{G}$ 
7      if  $\mathcal{G}_o$  is legal structure then
8        if score( $\mathcal{G}_o : \mathcal{D}$ ) > score( $\mathcal{G}_{best} : \mathcal{D}$ ) then
9           $\mathcal{G}_{best} \leftarrow \mathcal{G}_o$ 
10         Progress  $\leftarrow$  true
11  while Progress
12
13  return  $\mathcal{G}_{best}$ 

```

**Figure 2.7** Greedy structure search algorithm, with an arbitrary scoring function  $\text{score}(\mathcal{G} : \mathcal{D})$ .

Επιλέγουμε μια αρχική δικτυακή δομή  $G$  σαν σημείο εκκίνησης.

Αυτό το δίκτυο μπορεί να είναι το κενό δίκτυο, κάποια τυχαία επιλογή, το καλύτερο δέντρο ή ένα δίκτυο που προέκυψε από κάποια προηγούμενη γνώση.

Υπολογίζουμε τα αποτελέσματα του (score).

Κατόπιν θεωρούμε όλους τους γείτονες του  $G$  μέσα στον χώρο (δηλαδή όλα τα επιτρεπτά δίκτυα που προκύπτουν από την εφαρμογή ενός και μόνο τελεστή πάνω στον  $G$ ) και υπολογίζουμε τα αποτελέσματα για κάθε έναν από αυτούς.

Στην συνέχεια εφαρμόζουμε την αλλαγή/μεταβολή που οδηγεί στην καλύτερη βελτίωση των αποτελεσμάτων.

Συνεχίζουμε την διαδικασία μέχρι να μην υπάρχει κάποια τροποποίηση που να βελτιώνει τα αποτελέσματα.

Μπορούμε να βελτιώσουμε την απόδοση της άπληστης ανάβασης λόφου χρησιμοποιώντας πιο έξυπνους αλγόριθμους αναζήτησης.

Μερικοί από αυτούς είναι:

#### **Αναζήτηση TABU (TABU search):**

Κρατάμε μία λίστα των  $K$  πιο προσφάτως επισκεφθέντων δομών και τις αποφεύγουμε.

Πχ. εφαρμογή της καλύτερης κίνησης που θα μας οδηγήσει σε κάποια δομή η οποία δεν είναι στην λίστα.

Αυτή η προσέγγιση έχει να κάνει με τα τοπικά μέγιστα (local maxima) των οποίων ο "λόφος" (hill) έχει λιγότερες από  $K$  δομές.

#### **Τυχαίες επανεκκινήσεις (Random restarts):**

Μόλις "κολλήσουμε", εφαρμόζουμε κάποιον συγκεκριμένο αριθμό τυχαίων αλλαγών κλάδων και επανεκκινούμε την άπληστη αναζήτηση.

Στο τέλος της αναζήτησης, επιλέγουμε την καλύτερη δομή που συναντάμε οπουδήποτε πάνω στην τροχιά.

Αυτή η προσέγγιση μπορεί να ξεφύγει από το φρέαρ ενός τοπικού μεγίστου και να πάει σε κάποιο άλλο.

#### **Προσομοίωση απόψησης (Simulated annealing):**

Αποτιμούμε τελεστές σε τυχαία σειρά. Εάν ο τυχαία επιλεγμένος τελεστής συνεπάγεται ένα βήμα προς τα πάνω στον λόφο (uphill step) τότε προχωρούμε προς την προκύπτουσα δομή.

Να σημειωθεί ότι δεν χρειάζεται να είναι ο καλύτερος από τους τρέχοντες γείτονες.

Εάν ο τελεστής συνεπάγεται ένα βήμα προς τα κάτω στον λόφο (downhill step), τον εφαρμόζουμε με πιθανότητα αντιστρόφως ανάλογη της μείωσης των αποτελεσμάτων.

Μια παράμετρος θερμοκρασίας είναι αυτή που καθορίζει την πιθανότητα του να γίνουν βήματα προς τα κάτω στον λόφο.

Καθώς η αναζήτηση εξελίσσεται η θερμοκρασία μειώνεται και ο αλγόριθμος είναι λιγότερο πιθανό να κάνει ένα βήμα προς τα κάτω στον λόφο.

## **2.5 Συμπεράσματα**

Σε αυτό το κεφάλαιο έγινε μία συνεπτυγμένη περιγραφή των γραφικών μοντέλων, συμπεριλαμβανομένων της αναπαράστασης τους, αλγορίθμων συμπερασμού και αλγορίθμων μάθησης.

### 3.Μαρκοβιανή λογική (Markov logic): Ένα ενοποιητικό πλαίσιο εργασίας για στατιστική σχεσιακή μάθηση

Τα τελευταία χρόνια, εκδηλώνεται ιδιαίτερο ενδιαφέρον για την στατιστική σχεσιακή μάθηση (Statistical Relational Learning - SRL).

Συγκεκριμένα, έχουν εντοπιστεί πολλές χαρακτηριστικές εργασίες SRL και έχουν προταθεί, αντιστοίχως, πολλές προσεγγίσεις για την πραγματοποίησή τους.

Συνεπώς, είναι επιβεβλημένη η ανάγκη ύπαρξης ενός πλαισίου εργασίας το οποίο θα διευκολύνει την αμφίδρομη μεταφορά γνώσεως μεταξύ εργασιών και προσεγγίσεων, την σύγκριση μεταξύ των διαφόρων προσεγγίσεων και που θα βοηθήσει στην δημιουργία κάποιας συγκεκριμένης δομής για το όλο πεδίο.

Ένα τέτοιο προτεινόμενο πλαίσιο εργασίας είναι και η Μαρκοβιανή λογική (λογική Markov).

Σε συντακτικό επίπεδο, η Μαρκοβιανή λογική δεν ξεχωρίζει από την λογική πρώτης τάξεως, με εξαίρεση το γεγονός ότι σε κάθε τύπο ανατίθεται κάποιο βάρος.

Σημασιολογικά, ένα σύνολο τύπων Μαρκοβιανής λογικής αναπαριστά μια πιθανοτική κατανομή επί πιθανών κόσμων, με την μορφή ενός log-linear μοντέλου και συγκεκριμένα, με ένα και μόνο χαρακτηριστικό για κάθε γείωση (per grounding) ενός τύπου του συνόλου, με το αντίστοιχο βάρος.

Σκοπός μας είναι να δείξουμε τον τρόπο με τον οποίο είναι δυνατόν διάφορες προσεγγίσεις, όπως τα πιθανοτικά σχεσιακά μοντέλα, η κατασκευή μοντέλων βασισμένη σε γνώση (knowledge-based) και τα στοχαστικά λογικά προγράμματα να απεικονισθούν στην Μαρκοβιανή λογική.

Επίσης θα δείξουμε πως εργασίες όπως η συλλογική κατηγοριοποίηση, η πρόβλεψη συνδέσμων, ο σχηματισμός συμπλεγμάτων βάσει συνδέσμων, η μοντελοποίηση κοινωνικών δικτύων και η ταυτοποίηση αντικειμένων μπορούν να τυποποιηθούν επακριβώς με την Μαρκοβιανή λογική.

Στο τέλος θα αναπτυχθούν αλγόριθμοι μάθησης και συμπερασμού για Μαρκοβιανή λογική, καθώς επίσης θα παρουσιαστούν και ορισμένα πειραματικά αποτελέσματα για την περίπτωση μιας εργασίας πρόβλεψης συνδέσμου (link prediction task).

#### 3.1 Η ανάγκη για ένα ενοποιητικό πλαίσιο εργασίας

Πολλά πεδία εφαρμογών του πραγματικού κόσμου (αν όχι όλα) χαρακτηρίζονται τόσο από την ύπαρξη αβεβαιότητας όσο και από την παρουσία πολύπλοκων σχεσιακών δομών.

Η στατιστική μάθηση εστιάζει στο πρώτο ενώ η σχεσιακή μάθηση εστιάζει στο δεύτερο.

Η στατιστική σχεσιακή μάθηση (SRL) συνδυάζει τις δυνατότητες και των δύο.

Η έρευνα στην SRL επεκτάθηκε ταχύτατα τα τελευταία χρόνια, τόσο επειδή χρειάστηκε στις διάφορες εφαρμογές όσο και επειδή η στατιστική και η σχεσιακή μάθηση έχουν η κάθε μία πλέον ωριμάσει στον βαθμό που ο συνδυασμός τους μπορεί πλέον να αποτελέσει πεδίο έρευνας.

Είναι γεγονός δε, ότι έχει αναγνωριστεί ένα σημαντικός αριθμός SRL εργασιών ο οποίος συμπεριλαμβάνει συλλεκτικές κατηγοριοποιήσεις, πρόβλεψη συνδέσμων, συμπλεγματοποίηση βασισμένη σε συνδέσμους, μοντελοποίηση κοινωνικών δικτύων, αναγνώριση αντικειμένων και πολλά άλλα.



Επίσης έχει προταθεί ένα μεγάλο πλήθος SRL προσεγγίσεων (το οποίο μάλιστα συνεχώς μεγαλώνει) για διάφορα ζητήματα όπως η κατασκευή μοντέλων βασισμένη σε προϋπάρχουσα γνώση, τα στοχαστικά λογικά προγράμματα PRISM και MACCENT, τα πιθανοτικά σχεσιακά μοντέλα, τα σχεσιακά μοντέλα Markov, τα δίκτυα σχεσιακής εξάρτησης, η δομική λογιστική αναδρομή, οι συναρτήσεις σχεσιακών γενεών, ο δεσμευμένος λογικός προγραμματισμός για πιθανοτικά μοντέλα (CLP(BN)) και άλλα.

Αν και η ποικιλία των προβλημάτων και των προσεγγίσεων στον τομέα που συζητάμε είναι πολύτιμη, είναι δύσκολο για τους ερευνητές, τους σπουδαστές αλλά και τους επαγγελματίες να αναγνωρίσουν, να μάθουν και να εφαρμόσουν τα βασικά.

Συγκεκριμένα, οι σχέσεις μεταξύ διαφορετικών προσεγγίσεων και οι σχετικές τους δυνατότητες και αδυναμίες εξακολουθούν να είναι ελάχιστα κατανοητές ενώ οι καινοτομίες σε ένα είδος εργασιών ή εφαρμογών δεν μεταφέρονται εύκολα σε κάποιο άλλο είδος, με αποτέλεσμα να επιβραδύνεται ο ρυθμός εξέλιξης.

Για τούς λόγους αυτούς, υπάρχει μια αυξανόμενη πίεση για την εύρεση ενός ενοποιητικού πλαισίου εργασίας, μία κοινή γλώσσα για την περιγραφή και την συσχέτιση των διαφόρων εργασιών και προσεγγίσεων.

Προκειμένου να είναι πραγματικά χρήσιμο ένα τέτοιο πλαίσιο εργασίας, θα πρέπει να έχει τις ακόλουθες επιθυμητές ιδιότητες :

1. Το πλαίσιο εργασίας θα πρέπει να συμπεριλαμβάνει τόσο λογική πρώτης τάξεως, όσο και πιθανοτικά γραφικά μοντέλα. Ειδικά, κάποιες παρούσες ή μελλοντικές SRL προσεγγίσεις θα πέσουν εκτός "εμβέλειας".

2. Τα SRL προβλήματα θα πρέπει να αναπαριστώνται εύκολα και με απλό τρόπο στο πλαίσιο εργασίας.

3. Το πλαίσιο εργασίας θα πρέπει να διευκολύνει την χρήση του πεδίου γνώσης στην SRL.

Επειδή ο χώρος αναζήτησης για τους SRL αλγορίθμους είναι πολύ μεγάλος ακόμα και για δεδομένα τεχνητής νοημοσύνης (AI), η γνώση πεδίου είναι ιδιαίτερως κρίσιμης σημασίας για να έχουμε επιτυχή έκβαση. Αντιθέτως, η ικανότητα να συμπεριλαμβάνεται πλούσια γνώση πεδίου είναι ένα από τα πιο ελκυστικά χαρακτηριστικά της SRL.

4. Το πλαίσιο εργασίας θα πρέπει να διευκολύνει την επέκταση σε SRL τεχνικών από την στατιστική μάθηση, τον επαγωγικό λογικό προγραμματισμό, τον πιθανοτικό συμπερασμό και τον λογικό συμπερασμό. Κάτι τέτοιο θα επιταχύνει την πρόοδο στην SRL δεδομένου ότι μπορούμε να εκμεταλλευτούμε την εκτενή βιβλιογραφία που ήδη υπάρχει για αυτές τις περιοχές.

Σε αυτό το κεφάλαιο θα προτείνουμε την Μαρκοβιανή λογική (Markov logic) σαν ένα πλαίσιο εργασίας που θεωρούμε ότι πληρεί όλες τις παραπάνω προϋποθέσεις.

Θα ξεκινήσουμε κάνοντας μια σύντομη περίληψη του απαιτητού γνωστικού υποβάθρου που απαιτείται για τα δίκτυα Markov και την λογική πρώτης τάξεως. Στην συνέχεια, θα παρουσιαστεί η Μαρκοβιανή λογική και θα περιγραφεί το πώς πολλές SRL προσεγγίσεις και εργασίες μπορούν να αποδοθούν φορμαλιστικά σε αυτό το πλαίσιο εργασίας. Κατόπιν θα δείξουμε πως τεχνικές από την λογική, τον πιθανοτικό

συμπερασμό, την στατιστική και τον επαγωγικό λογικό προγραμματισμό μπορούν να χρησιμοποιηθούν για την επίτευξη πρακτικού συμπερασμού και αλγορίθμων μάθησης στην Μαρκοβιανή λογική.

Τέλος, θα δείξουμε την εφαρμογή αυτών των αλγορίθμων σε μία εργασία πρόβλεψης συνδέσμου για τον πραγματικό κόσμο και θα καταλήξουμε σε ορισμένα συμπεράσματα.

### 3.2 Δίκτυα Markov

Ένα δίκτυο Markov (επίσης γνωστό και σαν τυχαίο πεδίο Markov) είναι ένα μοντέλο για την από κοινού κατανομή των μεταβλητών  $X = (X_1, X_2, \dots, X_n) \in X$ .

Αποτελείται από ένα μη προσανατολισμένο γράφο  $G$  και ένα σύνολο συναρτήσεων δυναμικού  $\phi_k$ .

Ο γράφος έχει από έναν κόμβο για κάθε μεταβλητή και το μοντέλο έχει από μια συνάρτηση δυναμικού για κάθε ακμή του γράφου.

Μια συνάρτηση δυναμικού, είναι μία μη αρνητική συνάρτηση με πραγματικό πεδίο τιμών της κατάστασης της αντίστοιχης ακμής.

Η από κοινού κατανομή που αναπαρίσταται από ένα δίκτυο Markov δίνεται από την

$$P(X=x) = \frac{1}{Z} \prod_k \phi_k(x_{\{k\}}) \quad (3.1)$$

όπου  $x_{\{k\}}$  είναι η κατάσταση της  $k$  ακμής (δηλαδή η κατάσταση των μεταβλητών που εμφανίζονται σε αυτήν την ακμή).

Το  $Z$  είναι γνωστό ως συνάρτηση διαμερισμού (partition function) και δίνεται από την

$$Z = \sum_{x \in \mathcal{X}} \prod_k \phi_k(x_{\{k\}}).$$

Τα δίκτυα Markov συχνά είναι βολικό να αναπαρίστανται σαν log-linear μοντέλα, με κάθε δυναμικό ακμής να αντικαθίσταται από ένα εκθετικός βεβαρημένο άθροισμα των χαρακτηριστικών της κατάστασης, αυτό μας οδηγεί στην σχέση

$$P(X=x) = \frac{1}{Z} \exp \left( \sum_j w_j f_j(x) \right) \quad (3.2)$$

Ένα χαρακτηριστικό μπορεί να είναι μια οποιαδήποτε πραγματική συνάρτηση της κατάστασης.

Σε αυτό το κεφάλαιο θα επικεντρωθούμε σε δυαδικά χαρακτηριστικά  $f_j(x) \in \{0, 1\}$ .

Με την πιο άμεση μετάφραση που μπορεί να γίνει από την μορφή της συνάρτησης δυναμικού (3.1), βλέπουμε ότι υπάρχει ένα χαρακτηριστικό που αντιστοιχεί σε κάθε δυνατή κατάσταση  $x_{\{k\}}$  κάθε ακμής, με τα βάρη της να είναι  $\log \phi_k(x_{\{k\}})$ .

Αυτού του είδους η αναπαράσταση είναι εκθετική σε σχέση με τον αριθμό των ακμών. Παρ' όλα αυτά, έχουμε την ελευθερία να προσδιορίσουμε έναν πολύ

μικρότερο αριθμό χαρακτηριστικών (π.χ. λογικές συναρτήσεις της κατάστασης της ακμής), επιτρέποντας έτσι μια πιο συμπαγή αναπαράσταση συγκριτικά με την μορφή των συναρτήσεων δυναμικού, ειδικά όταν έχουμε να κάνουμε με μεγάλο αριθμό ακμών.

Αυτό το εκμεταλλεύονται τα λογικά δίκτυα Markov (Markov Logic Networks – MLNs).

Ο συμπερασμός σε δίκτυα Markov είναι #P-complete.

Η ευρύτερα χρησιμοποιούμενη μέθοδος για προσεγγιστικό συμπερασμό στα δίκτυα Markov είναι η αλυσίδα Markov Monte Carlo (Markov Chain Monte Carlo - MCMC) και συγκεκριμένα η δειγματοληψία Gibbs, η οποία εξελίσσεται δειγματοληπώντας κάθε μεταβλητή και επιστρέφοντας την αντίστοιχη περιοχή κάλυψης Markov (Markov blanket) αυτής.

Η περιοχή κάλυψης - Markov blanket - ενός κόμβου είναι το ελάχιστο σύνολο κόμβων που τον καθιστούν ανεξάρτητο από το υπόλοιπο δίκτυο. Σε ένα δίκτυο Markov πρόκειται απλώς για τους γειτονικούς (ως προς τον υπό εξέταση κόμβο) κόμβους μέσα στον γράφο.

Οι περιθώριες πιθανότητες υπολογίζονται με την καταμέτρηση επί των δειγμάτων ενώ οι δεσμευμένες πιθανότητες υπολογίζονται με την εκτέλεση του δειγματολήπτη Gibbs και με τις δεσμεύουσες μεταβλητές να παραμένουν αγκυστρωμένες στις δοθείσες τιμές τους.

Μια άλλη δημοφιλής μέθοδος για συμπερασμό σε δίκτυα Markov είναι η διάδοση πίστεως (belief propagation).

Η μέγιστη πιθανοφάνεια ή οι maximum a posteriori (MAP) εκτιμήσεις βαρών σε δίκτυα Markov δεν μπορούν να υπολογιστούν σε κλειστή μορφή. Επειδή όμως η λογαριθμική πιθανοφάνεια είναι κοίλη (concave) συνάρτηση των βαρών μπορούν να προσδιοριστούν επαρκώς με την χρήση καθιερωμένων μεθόδων που βασίζονται στην κλήση (gradient-based) ή με quasi-Newton μεθόδους βελτιστοποίησης.

Μια άλλη εναλλακτική είναι η επαναληπτική κλιμάκωση (iterative scaling).

Τα χαρακτηριστικά μπορούν να γίνουν επίσης γνωστά από τα δεδομένα, π.χ. με την άπληστη κατασκευή συζεύξεων ατομικών χαρακτηριστικών.

### 3.3 Λογική πρώτης τάξεως

Μια βάση γνώσεως (knowledge base - KB) πρώτης τάξεως είναι ένα σύνολο προτάσεων (sentences) ή τύπων (formulae) σε λογική πρώτης τάξεως.

Οι τύποι κατασκευάζονται με την χρήση τεσσάρων ειδών συμβόλων: σταθερών, μεταβλητών, συναρτήσεων και κατηγορημάτων.

Οι σταθερές αναπαριστούν αντικείμενα στο πεδίο του ενδιαφέροντος μας (π.χ. άνθρωποι: Anna, Bob, Chris κλπ.).

Οι μεταβλητές έχουν ως πεδίο ορισμού το σύνολο των αντικειμένων του πεδίου.

Οι συναρτήσεις (π.χ. Μητέρα Του) αναπαριστούν απεικονίσεις n-άδων αντικειμένων σε αντικείμενα.

Τα κατηγορήματα αναπαριστούν σχέσεις μεταξύ των αντικειμένων του πεδίου (π.χ. Φίλοι) ή ιδιότητες των αντικειμένων (π.χ. Καπνίζει).

Μια ερμηνεία (interpretation) δείχνει ποια σύμβολα αναπαριστούν τα αντικείμενα, τις συναρτήσεις και τις σχέσεις τους στο πεδίο.

Οι μεταβλητές και οι σταθερές μπορεί να είναι συγκεκριμένου τύπου και κατά συνέπεια οι μεταβλητές να έχουν ως πεδίο ορισμού μόνο το σύνολο των αντικειμένων

του αντίστοιχου τύπου ενώ οι σταθερές να μπορούν να αναπαραστήσουν μονάχα αντικείμενα του αντίστοιχου συγκεκριμένου τύπου.

Για παράδειγμα, η μεταβλητή  $X$  μπορεί να έχει σαν πεδίο ορισμού ανθρώπους (π.χ. Anna, Bob κλπ.) και η σταθερά  $C$  μπορεί να αναπαριστά μια πόλη (π.χ. Seattle, Tokyo κλπ.).

Ένας όρος (term) είναι οποιαδήποτε έκφραση η οποία αναπαριστά ένα αντικείμενο στο πεδίο.

Μπορεί να είναι μια σταθερά, μια μεταβλητή ή μια συνάρτηση που εφαρμόζεται σε μια  $n$ -άδα όρων.

Για παράδειγμα, Anna,  $X$  και Μέγιστος Κοινός Διαιρέτης ( $x, y$ ) είναι όροι.

Ένας ατομικός τύπος (atomic formula) ή άτομο (atom) είναι ένα κατηγορημα που εφαρμόζεται σε μια  $n$ -άδα όρων (π.χ. Φίλοι( $x$ , Μητέρα Του(Anna))).

Οι τύποι (formulae) κατασκευάζονται αναδρομικά από έναν ατομικό τύπο χρησιμοποιώντας λογικά συνδετικά και ποσοτικούς δείκτες (quantifiers).

Εάν  $F_1$  και  $F_2$  είναι τύποι τότε και τα επόμενα είναι τύποι:

$\neg F_1$  (άρνηση), το οποίο είναι αληθές αν και μόνο αν το  $F_1$  είναι ψευδές.

$F_1 \wedge F_2$  (σύνδεση), το οποίο είναι αληθές αν και μόνο αν και το  $F_1$  και το  $F_2$  είναι αληθή.

$F_1 \vee F_2$  (διάδεση), το οποίο είναι αληθές αν και μόνο αν το  $F_1$  ή το  $F_2$  είναι αληθές.

$F_1 \Rightarrow F_2$  (συνεπαγωγή), το οποίο είναι αληθές αν και μόνο αν το  $F_1$  είναι ψευδές ή το  $F_2$  είναι αληθές.

$F_1 \Leftrightarrow F_2$  (ισοδυναμία), το οποίο είναι αληθές αν και μόνο αν τα  $F_1$  και  $F_2$  έχουν τις ίδιες αληθοτιμές.

$\forall x F_1$  (οικουμενική ποσοτικοποίηση), το οποίο είναι αληθές αν και μόνο αν το  $F_1$  είναι αληθές για κάθε αντικείμενο  $x$  του πεδίου.

$\exists x F_1$  (ποσοτικοποίηση ύπαρξης), το οποίο είναι αληθές αν και μόνο αν το  $F_1$  είναι αληθές για τουλάχιστον ένα αντικείμενο  $x$  του πεδίου.

Οι παρενθέσεις μπορούν να χρησιμοποιηθούν για την επιβολή προτεραιότητας.

Ένα θετικό λέκρημα (literal) είναι ένας ατομικός τύπος ενώ ένα αρνητικό λέκρημα είναι ένας ατομικός τύπος με άρνηση.

Οι τύποι (formulae) σε μια KB είναι εμμέσως συζευγμένες και γι' αυτό μία KB μπορεί να θεωρηθεί ως ένας μεγάλος τύπος.

Ένας όρος γείωσης (ground term) είναι ένας όρος που δεν περιέχει μεταβλητές.

Ένα άτομο γείωσης (ground atom) ή ένα κατηγορημα γείωσης (ground predicate) είναι ένας ατομικός τύπος όλα τα ορίσματα (arguments) του οποίου είναι όροι γείωσης.

Ένας πιθανός κόσμος ή αλλιώς μία ερμηνεία Herbrand αποδίδει μια αληθοτιμή σε κάθε πιθανό κατηγορημα γείωσης.

Ένας τύπος είναι ικανοποιήσιμος αν και μόνο αν υπάρχει τουλάχιστον ένας κόσμος στον οποίο να αληθεύει.

Το βασικό πρόβλημα συμπερασμού στην λογική πρώτης τάξεως είναι να καθοριστεί αν μια KB συνεπάγεται έναν τύπο  $F$ , δηλαδή αν ο  $F$  είναι αληθής σε όλους τους κόσμους όπου η KB είναι αληθής (αυτό το συμβολίζουμε  $KB \models F$ ).

Αυτό γίνεται συνήθως με εϊς άτοπο απαγωγή: Η KB συνεπάγεται τον  $F$  αν και μόνο αν  $KB \cup \neg F$  είναι μη ικανοποιήσιμο.

Για τον αυτοματοποιημένο συμπερασμό, συνήθως μας βολεύει να μετατρέπουμε τους τύπους σε μία πιο κανονική μορφή, συνήθως την προτασιακή μορφή (γνωστή και ως συζευκτική κανονική μορφή- Conjunctive Normal Form (CNF)).

Μια KB σε προτασιακή (clausal) μορφή είναι μία σύζευξη προτάσεων, με την πρόταση να είναι μια διάζευξη λεκτημάτων.

Κάθε KB σε λογική πρώτης τάξεως μπορεί να μετατραπεί σε προτασιακή μορφή χρησιμοποιώντας μια μηχανική διαδοχή βημάτων.

Ο συμπερασμός στην λογική πρώτης τάξεως είναι ημιαποφασίσιμος (semidecidable). Εξαιτίας αυτού, οι βάσεις γνώσεως συχνά κατασκευάζονται χρησιμοποιώντας ένα δεσμευμένο υποσύνολο λογικής πρώτης τάξεως με πιο επιθυμητές ιδιότητες.

Οι ευρύτερα χρησιμοποιούμενοι περιορισμοί αφορούν στις προτάσεις Horn (Horn clauses) οι οποίες είναι προτάσεις οι οποίες περιέχουν το πολύ ένα θετικό λέκτημα.

Η γλώσσα προγραμματισμού Prolog βασίζεται στην λογική των προτάσεων Horn.

Τα προγράμματα Prolog μπορούν να αποκτηθούν ως γνώση από βάσεις δεδομένων στις οποίες κάνουμε αναζήτηση για προτάσεις Horn οι οποίες να ισχύουν για τα δεδομένα (data) που διαθέτουμε. Αυτό είναι κάτι που μελετάται στον τομέα του επαγωγικού λογικού προγραμματισμού (Inductive Logic Programming - ILP).

Στον πίνακα 3.1 φαίνεται μια απλή KB καθώς και η μετατροπή της σε προτασιακή μορφή.

Να σημειωθεί ότι αν και αυτοί οι τύποι μπορεί τυπικά να είναι αληθείς στον πραγματικό κόσμο δεν είναι πάντοτε αληθείς.

Στους περισσότερους τομείς είναι πολύ δύσκολο να καταλήξουμε σε μη τετριμμένους τύπους οι οποίοι να είναι πάντοτε αληθείς και τέτοιοι τύποι εμπεριέχουν ένα μικρό μόλις μέρος της σχετικής γνώσης.

Έτσι, παρ' όλη την εκφραστικότητα της, η καθαρή λογική πρώτης τάξεως έχει περιορισμένη εφαρμοσιμότητα σε πρακτικά προβλήματα τεχνητής νοημοσύνης.

Έχουν προταθεί πολλές ad hoc επεκτάσεις για την αντιμετώπιση αυτής της δυσκολίας.

Για την πιο περιορισμένη περίπτωση της προτασιακής λογικής, το πρόβλημα λύνεται αρκετά καλά με την χρήση πιθανοτικών γραφικών μοντέλων.

**Table 3.1** Example of a first-order knowledge base and MLN.  $Fr()$  is short for  $Friends()$ ,  $Sm()$  for  $Smokes()$ , and  $Ca()$  for  $Cancer()$

English	First-order logic	Clausal form	Wt
Friends of friends are friends	$\forall x \forall y \forall z Fr(x, y) \wedge Fr(y, z) \Rightarrow Fr(x, z)$	$\neg Fr(x, y) \vee \neg Fr(y, z) \vee Fr(x, z)$	0.7
Friendless people smoke.	$\forall x (\neg(\exists y Fr(x, y)) \Rightarrow Sm(x))$	$Fr(x, g(x)) \vee Sm(x)$	2.3
Smoking causes cancer.	$\forall x Sm(x) \Rightarrow Ca(x)$	$\neg Sm(x) \vee Ca(x)$	1.5
If two people are friends, either both smoke or neither does.	$\forall x \forall y Fr(x, y) \Rightarrow (Sm(x) \Leftrightarrow Sm(y))$	$\neg Fr(x, y) \vee Sm(x) \vee \neg Sm(y),$ $\neg Fr(x, y) \vee \neg Sm(x) \vee Sm(y)$	1.1 1.1

Στην επόμενη παράγραφο περιγράφεται ένας τρόπος για την γενίκευση αυτών των μοντέλων για την περίπτωση της πρώτης τάξεως.

### 3.4 Μαρκοβιανή λογική

Μια βάση γνώσεως πρώτης τάξεως (first-order KB) μπορεί να θεωρηθεί σαν ένα σύνολο ισχυρών περιορισμών επί του συνόλου των δυνατών κόσμων: Εάν ένας κόσμος παραβιάζει έστω και έναν τύπο (formula), τότε αυτός θα έχει μηδενική πιθανότητα.

Η βασική ιδέα στην Μαρκοβιανή λογική (Markov logic) είναι να "χαλαρώσουν" αυτοί οι περιορισμοί: Εάν ένας κόσμος παραβιάζει έναν τύπο της βάσης γνώσεως τότε αυτός θα είναι λιγότερο πιθανός αλλά όχι απίθανος (μηδενική πιθανότητα).

Όσο λιγότερους τύπους παραβιάζει ένας κόσμος τόσο πιο πιθανός είναι (δηλαδή τόσο μεγαλύτερη πιθανότητα αντιστοιχεί σε αυτόν).

Κάθε τύπος έχει ένα αντίστοιχο βάρος το οποίο αντανακλά το κατά πόσο ισχυρός είναι ο περιορισμός: Όσο μεγαλύτερο είναι το βάρος, τόσο μεγαλύτερη είναι η διαφορά, σε λογαριθμική πιθανότητα, μεταξύ ενός κόσμου ενός κόσμου ο οποίος ικανοποιεί τον τύπο και ενός κόσμου που δεν τον ικανοποιεί.

Ένα Μαρκοβιανό λογικό δίκτυο (Markov Logic Network - MLN) ορίζεται ως ένα σύνολο τύπων Μαρκοβιανής λογικής.

Τα MLNs ορίζουν πιθανοτικές κατανομές επί πιθανών κόσμων όπως θα δούμε στα αμέσως επόμενα.

Ένα λογικό δίκτυο Markov  $L$  είναι ένα σύνολο διατεταγμένων ζευγών  $(F_i, w_i)$ , όπου  $F_i$  είναι ένας τύπος σε λογική πρώτης τάξεως και  $w_i$  είναι ένας πραγματικός αριθμός. Μαζί με ένα πεπερασμένο σύνολο σταθερών  $C = \{c_1, c_2, \dots, c_{|C|}\}$ , ορίζουν ένα δίκτυο Markov  $M_{L,C}$  (βλέπε εξισώσεις (3.1) και (3.2)) ως ακολούθως:

1. Το  $M_{L,C}$  περιέχει έναν δυαδικό κόμβο για κάθε πιθανή γείωση κάθε κατηγορήματος που εμφανίζεται στο  $L$ . Η τιμή του κόμβου είναι 1 αν το άτομο γείωσης (ground atom) είναι αληθές ή 0 στην αντίθετη περίπτωση.

2. Το  $M_{L,C}$  περιέχει ένα χαρακτηριστικό για κάθε πιθανή γείωση κάθε τύπου  $F_i$  στο  $L$ . Η τιμή αυτού του χαρακτηριστικού είναι 1 εάν ο τύπος γείωσης (ground formula) είναι αληθής ή 0 στην αντίθετη περίπτωση. Το βάρος του χαρακτηριστικού είναι το  $w_i$  που σχετίζεται με τον  $F_i$  στο  $L$ .

Η σύνταξη των τύπων σε ένα MLN είναι η καθιερωμένη σύνταξη της λογικής πρώτης τάξεως.

Οι ελεύθερες (δηλαδή οι μη ποσοτικοποιημένες) μεταβλητές αντιμετωπίζονται ως οικουμενικώς ποσοτικοποιημένες (universally quantified) στο εξώτερο επίπεδο του τύπου.

Ένα MLN μπορεί να θεωρηθεί σαν ένα πρότυπο για την κατασκευή δικτύων Markov. Λαμβάνοντας διαφορετικά σύνολα σταθερών θα προκύψουν διαφορετικά δίκτυα, τα οποία μπορεί να διαφέρουν κατά πολύ μεταξύ τους ως προς το μέγεθος τους αλλά όλα θα διαθέτουν συγκεκριμένες κανονικότητες ως προς την δομή και τις παραμέτρους τους οι οποίες θα καθορίζονται από το MLN (π.χ. όλες οι γειώσεις του ίδιου τύπου (formula) θα έχουν το ίδιο βάρος).

Ονομάζουμε κάθε ένα από αυτά τα δίκτυα, "δίκτυο Markov γείωσης" (ground Markov network) προκειμένου να τα ξεχωρίζουμε από το MLN πρώτης τάξεως.

Από τον ορισμό και τις εξισώσεις (3.1) και (3.2), η πιθανοτική κατανομή επί πιθανών κόσμων  $x$  που ορίζονται από το δίκτυο Markov γείωσης  $M_{L,C}$  δίνεται από την

$$P(X=x) = \frac{1}{Z} \exp \left( \sum_i w_i n_i(x) \right) = \frac{1}{Z} \prod_i \phi_i(x_{\{i\}})^{n_i(x)} \quad (3.3)$$

Όπου  $n_i(x)$  είναι ο αριθμός των αληθών γειώσεων του  $F_i$  στο  $x$ ,  $x_{\{i\}}$  είναι η κατάσταση (αληθοτιμές) των ατόμων που εμφανίζονται στον  $F_i$  και  $\phi_i(x_{\{i\}}) = e^{w_i}$ .

Να σημειωθεί ότι παρόλο που ορίσαμε τα MLNs σαν log-linear μοντέλα, θα μπορούσαν ισοδύναμα να ορισθούν εξίσου καλά ως γινόμενα εκθετικών συναρτήσεων, όπως δείχνει η δεύτερη παραπάνω ισότητα.

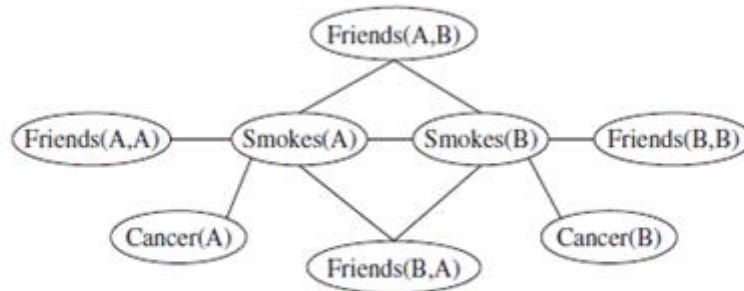
Αυτή θα είναι η πιο βολική προσέγγιση σε πεδία (domains) όπου έχουμε ταυτοχρόνως ισχυρούς και ασθενείς περιορισμούς (π.χ. εκεί όπου μερικοί τύποι ισχύουν μετά βεβαιότητας, αλλά που οδηγούν σε μηδενική πιθανότητα για κάποιους κόσμους).

Η γραφική δομή ενός  $M_{L,C}$  έχει ως εξής:

Μια ακμή μεταξύ δύο κόμβων του  $M_{L,C}$  υπάρχει αν και μόνο αν τα αντίστοιχα άτομα γείωσης (ground atoms) εμφανίζονται μαζί σε τουλάχιστον μία γείωση ενός τύπου του  $L$ .

Έτσι, τα άτομα σε κάθε τύπο γείωσης (ground formula) σχηματίζουν μία ακμή στο  $M_{L,C}$ .





**Figure (3.1)** Ground Markov network obtained by applying the last two formulae in table (3.1) to the constants Anna(A) and Bob(B).

Στο σχήμα 3.1 φαίνεται ο γράφος του δικτύου Markov γείωσης (ground Markov network) που ορίζεται από τους δύο τελευταίους τύπους του πίνακα 3.1 και τις σταθερές Anna και Bob.

Κάθε κόμβος σε αυτόν τον γράφο είναι ένα άτομο γείωσης- ground atom-(π.χ. Φίλοι (Anna, Bob)).

Ο γράφος περιέχει μια ακμή μεταξύ κάθε ζεύγους ατόμων που εμφανίζονται μαζί σε κάποια γείωση ενός από τους τύπους.

Μετά από όλα αυτά, το  $M_{L,C}$  μπορεί πλέον να χρησιμοποιηθεί για τον συμπερασμό της πιθανότητας η Anna και ο Bob να είναι φίλοι δεδομένων των καπνιστικών τους συνηθειών, της πιθανότητας ο Bob να πάσχει από καρκίνο δεδομένης της φιλίας του με την Anna, καθώς και εάν αυτή πάσχει από καρκίνο κλπ.

Κάθε κατάσταση του  $M_{L,C}$  αναπαριστά έναν πιθανό κόσμο.

Ένας πιθανός κόσμος αποτελείται από ένα σύνολο αντικειμένων, ένα σύνολο συναρτήσεων (απεικονίσεις από  $n$ -άδες αντικειμένων σε αντικείμενα) και ένα σύνολο σχέσεων που ισχύουν μεταξύ αυτών των αντικειμένων. Μαζί με μια ερμηνεία (interpretation), καθορίζουν την αληθοτιμή κάθε ενός ατόμου γείωσης.

Οι ακόλουθες υποθέσεις εξασφαλίζουν ότι το σύνολο των πιθανών κόσμων για το  $(L,C)$  είναι πεπερασμένος και ότι το  $M_{L,C}$  αναπαριστά μια μοναδική και καλώς ορισμένη πιθανοτική κατανομή επί αυτών των κόσμων ανεξαρτήτως της ερμηνείας και του πεδίου (domain).

Αυτές οι υποθέσεις είναι πολύ λογικές για τις περισσότερες πρακτικές εφαρμογές και απλοποιούν κατά πολύ την χρήση των MLNs.

Για τις απομένουσες περιπτώσεις, θα συζητήσουμε παρακάτω μέχρι σε ποιο βαθμό οι υποθέσεις αυτές μπορούν να "χαλαρώσουν".

### **Υπόθεση 1**

**Μοναδικά ονόματα** Διαφορετικές σταθερές αναφέρονται σε διαφορετικά αντικείμενα.

### **Υπόθεση 2**

**Περάτωση πεδίου (Domain closure)** Τα μόνα αντικείμενα μέσα στο πεδίο (domain) είναι εκείνα που μπορούν να παρασταθούν με την χρήση σταθερών και συναρτήσεων στο  $(L,C)$ .



### Υπόθεση 3

**Γνωστές συναρτήσεις** Για κάθε συνάρτηση που εμφανίζεται στο L, η τιμή της λαμβάνοντας ως ορίσματα κάθε πιθανή n-άδα είναι γνωστή και μάλιστα είναι μέλος του συνόλου C.

**Table 3.2** Construction of all groundings of a first-order formula under assumptions 1–3

---

```

function Ground( $F, C$ )
  inputs:  $F$ , a formula in first-order logic
            $C$ , a set of constants
  output:  $G_F$ , a set of ground formulae
  calls:  $CNF(F, C)$ , which converts  $F$  to conjunctive normal form, replacing
           existentially quantified formulae by disjunctions of their groundings over  $C$ 
 $F \leftarrow CNF(F, C)$ 
 $G_F = \emptyset$ 
for each clause  $F_j \in F$ 
   $G_j = \{F_j\}$ 
  for each variable  $x$  in  $F_j$ 
    for each clause  $F_k(x) \in G_j$ 
       $G_j \leftarrow (G_j \setminus F_k(x)) \cup \{F_k(c_1), F_k(c_2), \dots, F_k(c_{|C|})\}$ ,
      where  $F_k(c_i)$  is  $F_k(x)$  with  $x$  replaced by  $c_i \in C$ 
     $G_F \leftarrow G_F \cup G_j$ 
for each ground clause  $F_j \in G_F$ 
  repeat
    for each function  $f(a_1, a_2, \dots)$  all of whose arguments are constants
       $F_j \leftarrow F_j$  with  $f(a_1, a_2, \dots)$  replaced by  $c$ , where  $c = f(a_1, a_2, \dots)$ 
    until  $F_j$  contains no functions
  return  $G_F$ 

```

---

Αυτή η τελευταία υπόθεση μας επιτρέπει να αντικαταστήσουμε τις συναρτήσεις με τις τιμές τους όταν γειώνουμε τύπους.

Συνεπώς, τα μόνα άτομα γείωσης που πρέπει να λάβουμε υπόψη μας είναι αυτά που έχουν ως ορίσματα σταθερές.

Ο άπειρος αριθμός όρων (terms) που είναι κατασκευάσιμοι από όλες τις συναρτήσεις και τις σταθερές στο  $(L, C)$  (δηλαδή το σύμπαν Herbrand του  $(L, C)$ ) μπορεί να αγνοηθεί επειδή κάθε ένας από αυτούς τους όρους αντιστοιχεί σε μια γνωστή σταθερά του συνόλου  $C$  και τα άτομα που σχετίζονται με αυτούς αναπαρίστανται ήδη σαν τα άτομα που έχουν να κάνουν με τις αντίστοιχες σταθερές.

Συνεπώς, οι πιθανές γειώσεις ενός κατηγορήματος βρίσκονται με απλό τρόπο, αντικαθιστώντας κάθε μεταβλητή του κατηγορήματος με κάθε σταθερά του  $C$  και αντικαθιστώντας κάθε όρο της συνάρτησης του κατηγορήματος με την αντίστοιχη σταθερά.

Στον πίνακα 3.2 φαίνεται το πώς οι γειώσεις ενός τύπου προκύπτουν βάσει των υποθέσεων 1 ως 3.

Αν ένας τύπος περιέχει περισσότερες από μία προτάσεις (clauses), το βάρος του ισομοιράζεται μεταξύ των προτάσεων και το βάρος κάθε πρότασης ανατίθεται σε κάθε μία από τις γειώσεις της.

Η υπόθεση 1 (μοναδικά ονόματα) μπορεί να παραληφθεί εισάγοντας το κατηγορημα της ισότητας ( $\text{Equals}(x, y)$ , ή  $x = y$  για συντομία) και προσθέτοντας τα απαραίτητα αξιώματα στο MLN:

Η ισότητα είναι ανακλαστική, συμμετρική και μεταβατική.

Για κάθε μοναδιαίο (unary) κατηγορημα  $P$ ,  $\forall x \forall y x = y \Rightarrow (P(x) \Leftrightarrow P(y))$ .

Ομοίως για κατηγορήματα και συναρτήσεις μεγαλύτερης τάξης.

Το MLN που προκύπτει θα έχει ένα κόμβο για κάθε ζεύγος σταθερών, του οποίου η τιμή θα είναι 1 αν οι σταθερές αναπαριστούν το ίδιο αντικείμενο ή 0 στην αντίθετη περίπτωση.

Αυτοί οι κόμβοι θα συνδέονται μεταξύ τους, καθώς και με το υπόλοιπο δίκτυο με ακμές που αναπαριστούν τα προηγούμενα αξιώματα.

Να σημειώσουμε ότι αυτό μας επιτρέπει να κάνουμε πιθανοτικούς συμπερασμούς όσον αφορά την ισότητα δύο σταθερών.

Αυτό έχει χρησιμοποιηθεί με επιτυχία ως η βάση μιας προσέγγισης για την ταυτοποίηση αντικειμένων (βλέπε παράγραφο 3.6.5).

Αν ο αριθμός  $u$  των αγνώστων αντικειμένων είναι γνωστός, η υπόθεση 2 (περάτωση πεδίου) μπορεί να παραληφθεί, αν εισάγουμε  $u$  αυθαίρετες νέες σταθερές.

Αν το  $u$  είναι άγνωστο αλλά πεπερασμένο, η υπόθεση 2 μπορεί να παραληφθεί εαν εισάγουμε μια κατανομή επί του  $u$ , γειώσουμε το MLN για κάθε αριθμό αγνώστων αντικειμένων και υπολογίσουμε την πιθανότητα του τύπου σύμφωνα με την σχέση

$$P(F) = \sum_{u=0}^{u_{max}} P(u)P(F|M_{L,C}^u)$$

Όπου  $M_{L,C}^u$  είναι το MLN γείωσης (ground MLN) με  $u$  άγνωστα αντικείμενα.

Ένα άπειρο  $u$  απαιτεί να γίνει επέκταση των MLNs ως την περίπτωση  $|C| = \infty$ .

Έστω  $H_{L,C}$  το σύνολο των όρων γείωσης (ground terms) που μπορούν να κατασκευαστούν από τις συναρτήσεις του  $L$  και τις σταθερές των  $L$  και  $C$  (δηλαδή το σύμπαν Herbrand του  $(L,C)$ ).

Η υπόθεση 3 (γνωστές συναρτήσεις) μπορεί να παραληφθεί αν αντιμετωπίσουμε κάθε μέλος του  $H_{L,C}$  σαν μια επιπλέον σταθερά και εφαρμόσουμε την ίδια διαδικασία που χρησιμοποιήσαμε και για την αφαίρεση της υπόθεσης των μοναδικών ονομάτων.

Για παράδειγμα, με μια συνάρτηση  $G(x)$  και σταθερές  $A$  και  $B$ , το MLN πλέον θα περιέχει κόμβους για  $G(A) = A$ ,  $G(A) = B$  κλπ.

Αυτό οδηγεί σε έναν άπειρο αριθμό νέων σταθερών και έτσι απαιτείται η αντίστοιχη επέκταση των MLNs.

Αν όμως περιορίσουμε το επίπεδο της ένθεσης (nesting) μέχρι κάποιο μέγιστο, το προκύπτον MLN θα εξακολουθήσει να είναι πεπερασμένο.

Για να συνοψίσουμε, οι υποθέσεις 1 ως 3 μπορούν να παραληφθούν εφόσον το πεδίο είναι πεπερασμένο.

Πιστεύεται ότι είναι εφικτή η επέκταση των MLNs για άπειρα πεδία, αλλά αυτό είναι ένα θέμα κατά κύριο λόγο ερευνητικού ενδιαφέροντος.

Στα επόμενα θα συνεχίσουμε σύμφωνα με τις προϋποθέσεις 1-3, εκτός εάν αναφέρεται διαφορετικά.

Μια βάση γνώσεως πρώτης τάξεως μπορεί να μετασχηματιστεί σε MLN αναθέτοντας απλώς κάποιο βάρος σε κάθε τύπο.

Για παράδειγμα, οι προτάσεις και τα βάρη στις δύο τελευταίες στήλες του πίνακα 3.1 αποτελούν ένα MLN.

Σύμφωνα με αυτό το MLN, εκτός από την περίπτωση της ισότητας, ένας κόσμος όπου  $n$  άνθρωποι χωρίς φίλους είναι μη καπνιστές, είναι  $e^{(2.3)^n}$  φορές λιγότερο πιθανός από ότι ένας κόσμος όπου όλοι οι άνθρωποι χωρίς φίλους είναι καπνιστές.

Πρέπει να σημειωθεί ότι όλοι οι τύποι στον πίνακα 3.1 είναι ψευδείς για τον πραγματικό κόσμο ως οικουμενικώς ποσοτικοποιημένες λογικές δηλώσεις, εμπεριέχουν όμως σημαντικές πληροφορίες περί φιλίας και καπνιστικών συνηθειών όταν τις δούμε ως χαρακτηριστικά ενός δικτύου Markov.

Για παράδειγμα, είναι ευρέως γνωστό ότι οι έφηβοι φίλοι τείνουν να έχουν παρόμοιες καπνιστικές συνήθειες. Για την ακρίβεια, ένα MLN όπως αυτό του πίνακα 3.1 παρουσιάζει με συνοπτικό τρόπο έναν τύπο μοντέλου που είναι συγκερασμός αναλύσεων κοινωνικών δικτύων.

Είναι εύκολο να δούμε ότι τα MLNs υπάγουν κατ' ουσία όλα τα προτασιακά πιθανοτικά μοντέλα, όπως θα δούμε με περισσότερες λεπτομέρειες στα επόμενα.

Κάθε πιθανοτική κατανομή επί διακριτών ή πεπερασμένης ακρίβειας αριθμητικών μεταβλητών μπορούν να παρασταθούν ως ένα λογικό δίκτυο Markov.

Φυσικά, συμπαγή μοντέλα παραγόντων (factored models) όπως τα δίκτυα Markov και τα Bayesian δίκτυα μπορούν να παρασταθούν συμπαγώς με MLNs, ορίζοντας τύπους (formulae) για τους αντίστοιχους παράγοντες (factors) (αυθαίρετα χαρακτηριστικά για την περίπτωση των δικτύων Markov και καταστάσεις ενός κόμβου και των γονέων αυτού για την περίπτωση των Bayesian δικτύων).

Η λογική πρώτης τάξεως (με τις υποθέσεις 1-3) είναι η ειδική περίπτωση της Μαρκοβιανής λογικής που προκύπτει όταν όλα τα βάρη είναι ίσα μεταξύ τους και τείνουν στο άπειρο, όπως περιγράφεται παρακάτω.

Έστω KB μια ικανοποιήσιμη βάση γνώσεως, L το MLN που προκύπτει με την ανάθεση βαρών  $w$  σε κάθε τύπο της KB, C το σύνολο των σταθερών που εμφανίζονται στην KB,  $P_w(x)$  η πιθανότητα που ανατίθεται σε ένα σύνολο πιθανών κόσμων  $x$  από το  $M_{L,C}$ ,  $X_{KB}$  το σύνολο των κόσμων που ικανοποιούν την KB και F ένας αυθαίρετος τύπος σε λογική πρώτης τάξεως. Τότε:

1.

$$\begin{aligned} \forall x \in X_{KB} \lim_{w \rightarrow \infty} P_w(x) &= |X_{KB}|^{-1} \\ \forall x \notin X_{KB} \lim_{w \rightarrow \infty} P_w(x) &= 0 \end{aligned}$$

2. Για όλους τους τύπους F ισχύει ότι

$$KB \models F \text{ iff } \lim_{w \rightarrow \infty} P_w(F) = 1$$

Στο όριο όλων των ίσων απείρων βαρών το MLN αναπαριστά μια ομοιόμορφη κατανομή επί των κόσμων που ικανοποιούν την KB και όλες οι λογικά συνεπαγόμενες αναζητήσεις μπορούν να απαντηθούν υπολογίζοντας την πιθανότητα του τύπου της αναζήτησης (query formula) και ελέγχοντας αν ισούται με 1.

Ακόμα και όταν τα βάρη είναι άπειρα, η λογική πρώτης τάξεως " ενσωματώνεται" στην Μαρκοβιανή λογική με την εξής έννοια.

Ας υποθέσουμε χωρίς βλάβη της γενικότητας ότι όλα τα βάρη είναι μη αρνητικά.

Ένας τύπος με αρνητικό βάρος  $w$  μπορεί να αντικατασταθεί με την άρνηση του με βάρος  $-w$ .)

Αν η KB που αποτελείται από τους τύπους ενός MLN  $L$  (με την άρνηση αυτών αν τα βάρη τους είναι αρνητικά) είναι ικανοποιήσιμη, τότε για κάθε  $C$ , οι ικανοποιούσες αναθέσεις (satisfying assignments) είναι οι τρόποι της κατανομής που αναπαριστά το  $M_{L,C}$ .

Αυτό συμβαίνει επειδή οι τρόποι είναι οι κόσμοι  $x$  που έχουν μέγιστο

$\sum_i w_i p_i(x)$  (βλέπε (3.3)) και αυτή η έκφραση μεγιστοποιείται όταν όλες οι γειώσεις όλων των τύπων είναι αληθείς (δηλαδή, η KB ικανοποιείται).

Σε αντίθεση με μια KB πρώτης τάξεως όμως, ένα MLN μπορεί να παράγει χρήσιμα συμπεράσματα ακόμα και όταν περιέχει αντιθέσεις (contradictions).

Ένα MLN μπορεί επίσης να προκύψει κατόπιν συγχωνεύσεως πολλών KBs, ακόμα και αν είναι μερικώς ασύμβατα. Αυτό είναι εν δυνάμει χρήσιμο σε περιοχές όπως το the Semantic Web και η συνεργασία μαζών.

Έχει ενδιαφέρον να δούμε ένα απλό παράδειγμα για το πώς η Μαρκοβιανή λογική γενικεύει την λογική πρώτης τάξεως.

Ας θεωρήσουμε ένα MLN το οποίο περιέχει τον απλό τύπο  $\forall x R(x) \Rightarrow S(x)$  με βάρος  $w$  και  $C = \{A\}$ .

Αυτό οδηγεί σε τέσσερις πιθανούς κόσμους:

$\{\neg R(A), \neg S(A)\}$ ,  $\{\neg R(A), S(A)\}$ ,  $\{R(A), \neg S(A)\}$ , και  $\{R(A), S(A)\}$ .

Από την (3.3) παίρνουμε ότι  $P(\{R(A), \neg S(A)\}) = 1/(3e^w + 1)$  και ότι η πιθανότητα κάθε ενός από τους άλλους τρεις κόσμους είναι  $e^w/(3e^w + 1)$ .

(ο παρανομαστής είναι η συνάρτηση διαμερισμού  $Z$ , βλέπε παράγραφο 3.2)

Συνεπώς αν  $w > 0$ , η δράση του MLN θα είναι να δημιουργήσει τον κόσμο ο οποίος είναι πιθανότερο να είναι λιγότερο ασυνεπής με το  $\forall x R(x) \Rightarrow S(x)$  από ότι οι άλλοι τρεις κόσμοι.

Από τις παραπάνω πιθανότητες λαμβάνουμε ότι  $P(S(A)|R(A)) = 1/(1 + e^{-w})$ .

Όταν  $w \rightarrow \infty$  τότε  $P(S(A)|R(A)) \rightarrow 1$ , επανακτώντας την λογική συνέπεια.

Στην πράξη, έχει φανεί ότι είναι χρήσιμο να προσθέτουμε κάθε κατηγορήμα στο MLN σαν μία μοναδιαία πρόταση.

Με άλλα λόγια, για κάθε κατηγορήμα  $R(x_1, x_2, \dots)$  που εμφανίζεται στο MLN προσθέτουμε τον τύπο  $\forall x_1, x_2, \dots R(x_1, x_2, \dots)$  με κάποιο βάρος  $w_R$ .

Το βάρος μιας μοναδιαίας πρότασης μπορεί (χονδρικά) να αποδώσει την περιθώρια κατανομή του αντίστοιχου κατηγορήματος, αφήνοντας τα βάρη των μη μοναδιαίων προτάσεων ελεύθερα για την μοντελοποίηση μόνο ανεξαρτησιών μεταξύ κατηγορημάτων.

Όταν κατασκευάζουμε με το χέρι ένα MLN ή ερμηνεύουμε κάποιο που είναι ήδη εκμαθημένο, είναι χρήσιμο να έχουμε μια διαισθητική κατανόηση των βαρών.

Το βάρος ενός τύπου  $F$  είναι απλά οι λογαριθμικές πιθανότητες μεταξύ ενός κόσμου όπου ο  $F$  είναι αληθής και ενός κόσμου που είναι ψευδής, εκτός από την περίπτωση της ισότητας.

Αν όμως ο  $F$  μοιράζεται μεταβλητές με άλλους τύπους (όπως δηλαδή γίνεται συνήθως) ίσως να μην είναι δυνατό να διατηρήσουμε απαράλλακτες τις αληθοτιμές αυτών των τύπων καθώς αντιστρέφουμε αυτές του  $F$ .

Σε αυτή την περίπτωση δεν υπάρχει πλέον η ένα-προς-ένα συσχέτιση μεταξύ βαρών και πιθανοτήτων των τύπων.

Παρ' όλα αυτά όμως, οι πιθανότητες όλων των τύπων καθορίζουν με συλλογικό τρόπο όλα τα βάρη, εάν τις δούμε σαν περιορισμούς επί μιας κατανομής μεγίστης εντροπίας ή αν τις χειριστούμε σαν εμπειρικές πιθανότητες προκειμένου να μάθουμε τα βάρη μεγίστης πιθανοφάνειας (αυτά τα δύο είναι ισοδύναμα).

Άρα ένας καλός τρόπος για να καθορίζουμε τα βάρη ενός MLN είναι να γράψουμε την πιθανότητα με την οποία κάθε τύπος πρέπει να ισχύει, να τις αντιμετωπίσουμε σαν εμπειρικές συχνότητες και να μάθουμε τα βάρη από αυτές χρησιμοποιώντας τον αλγόριθμο της παραγράφου 3.8.

Αντιστρόφως, τα βάρη ενός ήδη γνωστού (μαθημένου) MLN μπορούμε να τα δούμε σαν να κωδικοποιούν συλλογικά τις εμπειρικές πιθανότητες του τύπου.

Το μέγεθος των δικτύων Markov γείωσης μπορεί να ελαττωθεί σημαντικά έχοντας γράψει τις σταθερές και τις μεταβλητές και γειώνοντας μόνο μεταβλητές σε σταθερές του ίδιου τύπου.

Ακόμη και σε αυτήν την περίπτωση, το μέγεθος του δικτύου μπορεί να είναι πολύ μεγάλο.

Ευτυχώς, πολλοί συμπερασμοί δεν απαιτούν την γείωση ολόκληρου του δικτύου όπως θα δούμε στην παράγραφο 3.7.

### 3.5 Προσεγγίσεις SRL

Εξαιτίας της απλότητας και της γενικότητας της Μαρκοβιανής λογικής, πολλές αναπαραστάσεις που χρησιμοποιούνται στην SRL μπορούν με ευκολία να απεικονισθούν σε αυτήν.

Αυτό ακριβώς θα προσπαθήσουμε να κάνουμε σε αυτήν την παράγραφο για ένα αντιπροσωπευτικό δείγμα αυτού του είδους των προσεγγίσεων.

Σκοπός δεν είναι να αναδείξουμε τις πράγματι πάρα πολλές λεπτομέρειες τους αλλά να αναδείξουμε κάποιο είδος σαφούς δομής το πεδίου (field).

Επιπλέον, μετατρέποντας αυτού του είδους τις αναπαραστάσεις σε Μαρκοβιανή λογική θα δούμε ότι προκύπτουν νέες δυνατότητες και πλεονεκτήματα τα οποία και θα συζητήσουμε.

#### 3.5.1 Κατασκευή μοντέλων βασιζόμενη σε γνώση

Η κατασκευή μοντέλων βασιζόμενη σε γνώση (Knowledge-Based Model Construction-KBMC) είναι ένας συνδυασμός λογικού προγραμματισμού και Bayesian δικτύων.

Όπως και στην Μαρκοβιανή λογική, οι κόμβοι στην KBMC αναπαριστούν κατηγορήματα γειώσεως.

Με δεδομένη μια Horn KB, η KBMC απαντά σε μια αναζήτηση (query) βρίσκοντας όλες τις δυνατές αποδείξεις της αναζήτησης που μας οδηγούν αλυσιδωτά προς τα

πίσω (backward-chaining proofs) καθώς και τα αποδεικτικά κατηγορήματα (evidence predicates) αλλήλων κατασκευάζοντας ένα Bayesian δίκτυο επί των κατηγορημάτων γείωσης των αποδείξεων και εκτελώντας συμπερασμό επί αυτού του δικτύου.

Οι γονείς ενός κόμβου-κατηγορήματος του δικτύου είναι ντετερμινιστικοί κόμβοι AND που αναπαριστούν τα σώματα των προτάσεων που έχουν αυτόν τον κόμβο ως κεφαλή.

Η δεσμευμένη πιθανότητα του κόμβου σύμφωνα με τα παραπάνω καθορίζεται μια συνάρτηση συνδυασμού (π.χ. noisy OR, λογιστική αναδρομή, πίνακας αυθαίρετης δεσμευμένης πιθανότητας (CPT)).

Η Μαρκοβιανή λογική γενικεύει την KBMC επιτρέποντας αυθαίρετους τύπους (όχι απλώς προτάσεις Horn) και συμπερασμό προς οποιαδήποτε κατεύθυνση.

Επίσης παρακάμπτει το πρόβλημα αποφυγής κύκλων εντός Bayesian δικτύων που κατασκευάζονται από KBMC και καθιστά περιττή την ανάγκη για ad hoc συναρτήσεις συνδυασμού για προτάσεις με την ίδια συνέπεια.

Ένα μοντέλο KBMC μπορεί να μεταφραστεί σε Μαρκοβιανή λογική γράφοντας ένα σύνολο τύπων για κάθε κατηγορήμα πρώτης τάξεως  $P_k(\dots)$  του πεδίου.

Κάθε τύπος είναι μια σύζευξη που περιέχει  $P_k(\dots)$  και ένα λέκτημα για κάθε γονέα του  $P_k(\dots)$  ως συνέπεια.

Δημιουργούμε την άρνηση (negation) κάποιου υποσυνόλου αυτών των λεκτημάτων. Υπάρχει ένας τύπος για κάθε δυνατό συνδυασμό θετικών και αρνητικών λεκτημάτων. Το βάρος του τύπου είναι  $w = \log[p/(1-p)]$ , όπου  $p$  είναι η δεσμευμένη πιθανότητα του κατηγορήματος-απογόνου όταν η αντίστοιχη σύζευξη κατηγορημάτων-γονέων είναι αληθής σύμφωνα με την συνάρτηση συνδυασμού που χρησιμοποιείται.

Εάν η συνάρτηση συνδυασμού είναι λογιστική αναδρομή, μπορεί να αναπαρασταθεί χρησιμοποιώντας μόνο έναν γραμμικό αριθμό τύπων, εκμεταλλευόμενοι το γεγονός ότι το μοντέλο μιας λογιστικής αναδρομής είναι ένα (δεσμευμένο) δίκτυο Markov με μια δυαδική ακμή μεταξύ κάθε προβλεπτικού παράγοντα και της απόκρισης.

Η Noisy OR μπορεί επίσης να αναπαρασταθεί με έναν γραμμικό αριθμό γονέων.

### 3.5.2 Άλλες προσεγγίσεις λογικού προγραμματισμού

Τα στοχαστικά λογικά προγράμματα (Stochastic Logic Programs-SLPs) είναι συνδυασμός λογικού προγραμματισμού και log-linear μοντέλων.

Οι Puech και Muggleton έδειξαν ότι τα SLPs είναι μια ειδική περίπτωση KBMC και κατά συνέπεια μπορούν να μετατραπούν σε Μαρκοβιανή λογική με τον ίδιο τρόπο.

Όπως και στην Μαρκοβιανή λογική, τα SLPs έχουν ένα συντελεστή για κάθε πρόταση αλλά αναπαριστούν κατανομές επί δέντρων αποδείξεως Prolog και όχι επί κατηγορημάτων. Το δεύτερο μπορεί να επιτευχθεί με περιθωριοποίηση.

Παρόμοια σχόλια μπορούν να ειπωθούν για ένα πλήθος άλλων αναπαραστάσεων οι οποίες είναι κατ' ουσίαν ισοδύναμες με τα SLPs, όπως η λογική ανεξάρτητης επιλογής και το PRISM.

Το MACCENT είναι ένα σύστημα το οποίο μαθαίνει log-linear μοντέλα με χαρακτηριστικά πρώτης τάξεως. Κάθε χαρακτηριστικό (feature) είναι μια σύζευξη μιας κλάσεως και μιας αναζήτησης Prolog (Prolog query), δηλαδή μιας πρότασης με κενή κεφαλή.

Μια ουσιαστική διαφορά μεταξύ MACCENT και Μαρκοβιανής λογικής είναι ότι το MACCENT είναι ένα σύστημα κατηγοριοποίησης (δηλαδή, προβλέπει την δεσμευμένη κατανομή της κλάσεως ενός αντικειμένου δεδομένων των ιδιοτήτων



του), ενώ ένα MLN αναπαριστά την συνολική από κοινού κατανομή ενός συνόλου κατηγορημάτων.

Όπως κάθε προσέγγιση πιθανοτικής εκτίμησης, η Μαρκοβιανή λογική μπορεί να χρησιμοποιηθεί και για κατηγοριοποίηση εισάγοντας απλώς τις κατάλληλες δεσμευμένες αναζητήσεις (conditional queries).

Συγκεκριμένα, ένα μοντέλο MACCENT μπορεί να μετατραπεί σε Μαρκοβιανή λογική ορίζοντας απλώς ένα κατηγορημα κλάσεως (όπως στη παράγραφο 3.6.1), προσθέτοντας τα ανάλογα χαρακτηριστικά με τα βάρη τους στο MLN και προσθέτοντας έναν τύπο με άπειρο βάρος, δηλώνοντας έτσι ότι κάθε αντικείμενο πρέπει να έχει ακριβώς μία κλάση.

Το MACCENT μπορεί να χρησιμοποιήσει ντετερμινιστική γνώση υποβάθρου με την μορφή προτάσεων Prolog. Αυτές μπορούν να προστεθούν στο MLN σαν τύποι με άπειρα βάρη.

Επιπλέον, η Μαρκοβιανή λογική επιτρέπει την αβέβαιη γνώση υποβάθρου (μέσω τύπων με άπειρα βάρη).

Όπως περιγράφηκε στην παράγραφο 3.6.1, τα MLNs μπορούν να χρησιμοποιηθούν για συλλεκτική κατηγοριοποίηση, όπου οι κλάσεις διαφορετικών αντικειμένων μπορεί να είναι αλληλοεξαρτώμενες.

Το MACCENT, το οποίο απαιτεί το κάθε αντικείμενο να αναπαρίσταται με μια ξεχωριστή Prolog KB, δεν έχει αυτήν την δυνατότητα.

Ο δεσμευμένος λογικός προγραμματισμός είναι μια προέκταση του λογικού προγραμματισμού, όπου οι μεταβλητές δεσμεύονται αντί να περιορίζονται σε συγκεκριμένες τιμές κατά την διάρκεια του συμπερασμού.

Τα πιθανοτικά CLP γενικεύουν τα SLPs σε CLP και τα CLP(BN) συνδυάζουν CLP με Bayesian δίκτυα.

Σε αντίθεση με την Μαρκοβιανή λογική, οι περιορισμοί σε CLP(BN) είναι ισχυροί (Δηλαδή δεν μπορούν να παραβιαστούν και επιπλέον ορίζουν την μορφή της πιθανοτικής κατανομής)

### 3.5.3 Πιθανοτικά σχεσιακά μοντέλα

Τα πιθανοτικά σχεσιακά μοντέλα (Probabilistic Relational Models-PRMs) είναι ένας συνδυασμός συστημάτων που βασίζονται σε πλαίσια (frame-based) και Bayesian δικτύων.

Τα PRMs μπορούν να μετατραπούν σε Μαρκοβιανή λογική ορίζοντας ένα κατηγορημα  $S(x,v)$  για κάθε (προτασιακή ή σχεσιακή) ιδιότητα κάθε κλάσεως, όπου  $S(x,v)$  σημαίνει “Η τιμή της ιδιότητας  $S$  στο αντικείμενο  $x$  είναι  $v$ .”

Ένα PRM μεταφράζεται σε ένα MLN γράφοντας έναν τύπο για κάθε γραμμή καθενός CPT και για κάθε τιμή της ιδιότητας παιδιού.

Ο τύπος είναι μια σύζευξη λεκτημάτων που δείχνουν τις τιμές των γονέων (parent values) και ένα λέκτημα που δείχνει την τιμή παιδιού (child value). Το δε βάρος του είναι ο λογάριθμος του  $P(x|Parents(x))$ , το οποίο είναι η αντίστοιχη είσοδος στο CPT. Επιπλέον, το MLN περιέχει τύπους με άπειρα βάρη που δηλώνει ότι κάθε ιδιότητα πρέπει να λάβει ακριβώς μία τιμή.

Αυτή η προσέγγιση διαχειρίζεται όλους τους τύπους αβεβαιότητας στα PRMs (αβεβαιότητα ως προς την ιδιότητα, αβεβαιότητα ως προς την αναφορά, αβεβαιότητα ως προς την ύπαρξη).

Όπως επισημαίνει ο Taskar και η ερευνητική ομάδα αυτού, η ανάγκη της αποφυγής κύκλων μέσα σε PRMs προκαλεί σημαντικές δυσκολίες ως προς αναπαράσταση και τους υπολογισμούς.

Ο συμπερασμός σε PRMs γίνεται δημιουργώντας το πλήρες δίκτυο γειώσης, το οποίο περιορίζει την ιδιότητα κλιμάκωσης.

Τα PRMs απαιτούν τον προσδιορισμό ενός πλήρους δεσμευμένου μοντέλου (conditional model) για την κάθε ιδιότητα κάθε κλάσεως και αυτό είναι κάτι που για μεγάλους και πολύπλοκους τομείς μπορεί να είναι ιδιαίτερα επίπονο.

Αντιθέτως, η Μαρκοβιανή λογική δημιουργεί μια πλήρη από κοινού κατανομή από οποιοδήποτε αριθμό ιδιοτήτων πρώτης τάξεως ο χρήστης επιλέξει να προσδιορίσει.

### 3.5.4 Σχεσιακά δίκτυα Markov

Τα σχεσιακά δίκτυα Markov (Relational Markov Networks-RMN) χρησιμοποιούν συζευκτικές αναζητήσεις βάσεων δεδομένων σαν πρότυπα κλίκα.

Δεν παρέχουν κάποιου είδους γλώσσα για τον προσδιορισμό χαρακτηριστικών.

Σαν αποτέλεσμα, εξ αρχής τα RMNs απαιτούν ένα χαρακτηριστικό για κάθε πιθανή κατάσταση μίας κλίκα καθιστώντας τες εκθετικές ως προς το μέγεθος τους και περιορίζοντας την πολυπλοκότητα των εξαρτήσεων που μπορούν να μοντελοποιήσουν.

Η Μαρκοβιανή λογική μας παρέχει την λογική πρώτης τάξεως σαν μια ισχυρή γλώσσα για τον προσδιορισμό χαρακτηριστικών.

Ο προσδιορισμός των χαρακτηριστικών, προσδιορίζει με έμμεσο τρόπο και τις κλίκες, οι οποίες μπορεί να είναι πολύ μεγάλες όσο ο αριθμός των σχετικών χαρακτηριστικών (π.χ. τύποι- formulae) είναι εύκολος στον χειρισμό του.

Επιπλέον, η Μαρκοβιανή λογική γενικεύει τα RMNs επιτρέποντας την ύπαρξη αβεβαιότητας επί αυθαίρετων σχέσεων (όχι απλώς ιδιότητες μεμονωμένων αντικειμένων).

Τα RMNs εκπαιδεύονται μεροληπτικά και δεν προσδιορίζουν μια πλήρη από κοινού κατανομή για τις μεταβλητές του μοντέλου.

Η μεροληπτική εκπαίδευση των MLNs είναι εύκολο να κατανοηθεί.

Τα RMNs χρησιμοποιούν προσέγγιση MAP μαζί με μετάδοση πίστωσης για να κάνουν συμπερασμό και αυτό κάνει την μάθηση αρκετά αργή παρ' όλο το απλοποιημένο μεροληπτικό περιβάλλον.

Τόσο η βελτιστοποίηση ψευδο-πιθανοφάνειας (pseudo-likelihood optimization) όσο και η μεροληπτική εκπαίδευση (discriminative training) περιγράφονται σαν κατά πολύ γρηγορότερες στις εργασίες των Singla και Domingos.

Μέχρι και σήμερα, δεν έχουν προταθεί αλγόριθμοι εκμάθησης δομής για RMNs.

Η δομή MLN μπορεί να "διδασχθεί" είτε χρησιμοποιώντας καθιερωμένες τεχνικές επαγωγικού λογικού προγραμματισμού (ILP), όπως θα δείξουμε στα επόμενα αυτού του κεφαλαίου είτε με απευθείας βελτιστοποίηση ψευδο-πιθανοφάνειας, όπως περιγράφεται από την εργασία των Kok και Domingos.



### 3.5.5 Λογιστική αναδρομή δομής

Στην λογιστική αναδρομή δομής (Structural Logistic Regression-SLR), οι προβλέπτες (predictors) είναι η έξοδος (δηλαδή το αποτέλεσμα) αναζητήσεων SQL (SQL queries) επί των δεδομένων εισόδου (input data).

Με τον ίδιο τρόπο που μπορούμε να δούμε ένα μοντέλο λογιστικής αναδρομής σαν ένα μεροληπτικώς εκπαιδευμένο δίκτυο Markov, έτσι μπορούμε να δούμε και ένα μοντέλο SLR σαν ένα μεροληπτικώς εκπαιδευμένο MLN.

### 3.5.6 Δίκτυα σχεσιακής εξάρτησης

Σε ένα δίκτυο σχεσιακής εξάρτησης (Relational Dependency Network RDN), η πιθανότητα κάθε κόμβου η οποία περιορίζεται από την αντίστοιχη περιοχή κάλυψης Markov (Markov blanket) δίνεται από ένα δέντρο αποφάσεων.

Κάθε RDN έχει ένα αντίστοιχο MLN κατά τον ίδιο τρόπο που κάθε δίκτυο εξάρτησης έχει και ένα αντίστοιχο δίκτυο Markov το οποίο δίνεται από την στάσιμη κατανομή ενός δειγματολήπτη Gibbs ο οποίος δρά επί αυτού.

### 3.5.7 Επικαλύψεις και πιθανοτικά μοντέλα σχέσεων οντοτήτων

Μεγάλα γραφικά μοντέλα με επαναλαμβανόμενες δομές, συχνά αναπαρίστανται με έναν πιο συμπαγή τρόπο χρησιμοποιώντας επικαλύψεις (plates).

Η Μαρκοβιανή λογική επιτρέπει τον καθορισμό των επικαλύψεων με την χρήση της οικουμενικής ποσοτικοποίησης (universal quantification).

Επιπλέον, επιτρέπει την αναπαράσταση των ξεχωριστών οντοτήτων και των μεταξύ τους σχέσεων με σαφήνεια καθώς και την συμπαγή γραφή ανεξαρτησιών με βάση το περιεχόμενο αντί να αφήνονται ως υπόνοιες στα μοντέλα των κόμβων.

Πρόσφατα, ο Heckerman και η ομάδα του πρότειναν τα πιθανοτικά μοντέλα σχέσεων οντοτήτων (probabilistic entity relationship (ER) models).

Πρόκειται για μία γλώσσα που βασίζεται στα μοντέλα ER και που συνδυάζει τα χαρακτηριστικά των επικαλύψεων (plates) και των PRMs.

Αυτή η γλώσσα μπορεί να απεικονιστεί σε Μαρκοβιανή λογική με τον ίδιο τρόπο που τα μοντέλα ER μπορούν να απεικονιστούν σε λογική πρώτης τάξεως.

Τα πιθανοτικά μοντέλα ER επιτρέπουν στις λογικές εκφράσεις να δρουν ως περιορισμοί πάνω στον τρόπο με τον οποίο θα κατασκευαστούν τα δίκτυα γείωσης. Θα πρέπει όμως οι αληθοτιμές αυτών των εκφράσεων να είναι εξ αρχής γνωστές.

Η Μαρκοβιανή λογική επιτρέπει την ύπαρξη αβεβαιότητας επί όλων των λογικών εκφράσεων.

### 3.5.8 BLOG

Ο Milch και η ερευνητική ομάδα αυτού πρότειναν μία γλώσσα, η οποία ονομάζεται BLOG (Bayesian Logic) και οποία είναι σχεδιασμένη έτσι ώστε να μην χρειάζεται να γίνονται οι υποθέσεις μοναδικών ονομάτων και περάτωσης πεδίου.

Ένα πρόγραμμα BLOG προσδιορίζει διαδικαστικά το πώς να δημιουργηθεί ένας πιθανός κόσμος και δεν επιτρέπει εύκολα την ενσωμάτωση αυθαίρετης γνώσης πρώτης τάξεως.

Επίσης, καθορίζει μόνο την δομή του μοντέλου, αφήνοντας τις παραμέτρους να καθοριστούν με εξωτερικό τρόπο.

Τα μοντέλα BLOG είναι προσανατολισμένοι γράφοι και γι' αυτό απαιτείται να αποφεύγουμε τους κύκλους, καθότι ουσιαστικά κάνουν την κατασκευή τους πιο πολύπλοκη.

Είδαμε σε προηγούμενες παραγράφους πώς μπορούμε να αφαιρέσουμε τις υποθέσεις μοναδικών ονομάτων και περάτωσης πεδίου σε Μαρκοβιανή λογική. Όταν υπάρχουν άγνωστα αντικείμενα πολλαπλών τύπων (multiple types), εισάγουμε μια τυχαία μεταβλητή για τον αριθμό των αντικειμένων κάθε τύπου (type).

Ο συμπερασμός για τις ιδιότητες ενός αντικειμένου (αντί για τις παρατηρήσεις αυτού) γίνεται έχοντας μεταβλητές τόσο για τα αντικείμενα όσο και για τις παρατηρήσεις του (π.χ. τόσο για βιβλία όσο και για τις αντίστοιχες αναφορές σε αυτά).

Μέχρι σήμερα, δεν έχουν προταθεί για την BLOG αλγόριθμοι μάθησης ή πρακτικοί αλγόριθμοι συμπερασμού.

## 3.6 Εργασίες SRL

Πολλές εργασίες SRL μπορούν να μορφοποιηθούν σε Μαρκοβιανή λογική, καθιστώντας έτσι εφικτό να δούμε το πώς σχετίζονται μεταξύ τους και να αναπτύξουμε αλγορίθμους οι οποίοι να είναι ταυτοχρόνως εφαρμόσιμοι σε όλες.

Σε αυτήν την παράγραφο θα δείξουμε πώς γίνεται αυτό μέσω πέντε χαρακτηριστικών εργασιών: την συλλογική κατηγοριοποίηση, την πρόβλεψη συνδέσμου, την συμπλεγματοποίηση η οποία βασίζεται σε συνδέσμους, την μοντελοποίηση κοινωνικών δικτύων και την ταυτοποίηση αντικειμένων.

### 3.6.1 Συλλογική κατηγοριοποίηση

Σκοπός της απλής κατηγοριοποίησης είναι να προβλεφθεί η κλάση ενός αντικειμένου σύμφωνα με τις ιδιότητες που αυτό διαθέτει.

Η συλλογική κατηγοριοποίηση λαμβάνει επίσης υπόψη της τις κλάσεις σχετιζομένων αντικειμένων.

Οι ιδιότητες μπορούν να αναπαρασταθούν σε Μαρκοβιανή λογική σαν κατηγορήματα της μορφής  $A(x,v)$ , όπου  $A$  είναι μια ιδιότητα,  $x$  είναι ένα αντικείμενο και  $v$  είναι η τιμή της  $A$  στο  $x$ .

Η κλάση είναι μια προσδιορισμένη ιδιότητα  $C$ , αναπαριστώμενη ως  $C(x,v)$ , όπου  $v$  είναι η τάξη του  $x$ .

Η κατηγοριοποίηση πλέον ανάγεται απλώς στο πρόβλημα συμπερασμού της αληθοτιμής του  $C(x,v)$  για όλα τα  $x$  και  $v$  του ενδιαφέροντος μας, έχοντας ως δεδομένα όλα τα γνωστά  $A(x,v)$ .

Η απλή κατηγοριοποίηση είναι η ειδική περίπτωση όπου τα  $C(x_i,v)$  και  $C(x_j,v)$  είναι ανεξάρτητα για όλα τα  $x_i$  και  $x_j$ , δεδομένων των γνωστών  $A(x,v)$ .

Στην συλλογική κατηγοριοποίηση, η περιοχή κάλυψης Markov των  $C(x_i,v)$  περιλαμβάνει άλλα  $C(x_j,v)$  ακόμα και μετά την επιβολή περιορισμών επί των γνωστών  $A(x,v)$ .

Οι σχέσεις μεταξύ των αντικειμένων αναπαριστώνται με κατηγορήματα της μορφής  $R(x_i, x_j)$ .

Ένα πλήθος γενικεύσεων που παρουσιάζουν ενδιαφέρον είναι πλέον προφανές.

Για παράδειγμα, τα  $C(x_i,v)$  και  $C(x_j,v)$  ίσως να είναι εμμέσως εξαρτώμενα μέσω αγνώστων κατηγορημάτων. Ενδεχομένως συμπεριλαμβανομένων και των ιδίων των κατηγορημάτων  $R(x_i, x_j)$

### 3.6.2 Πρόβλεψη συνδέσμου

Σκοπός της πρόβλεψης συνδέσμου είναι να καθοριστεί αν υπάρχει σχέση μεταξύ δύο αντικειμένων του ενδιαφέροντος μας (π.χ. αν η Anna είναι επιβλέπουσα της διδακτορικής διατριβής του Bob) από τις ιδιότητες αυτών των αντικειμένων και ενδεχόμενες άλλες μεταξύ τους σχέσεις που να είναι γνωστές.

Η μορφοποίηση αυτού του προβλήματος σε Μαρκοβιανή λογική είναι ίδια με αυτήν της συλλογικής κατηγοριοποίησης, με την μόνη διαφορά ότι σκοπός τώρα είναι να συμπερασθεί η τιμή του  $R(x_i, x_j)$  για όλα τα ζεύγη αντικειμένων του ενδιαφέροντος μας αντί του  $C(x, v)$ .

Η εργασία που χρησιμοποιείται στα πειράματά μας είναι ένα παράδειγμα πρόβλεψης συνδέσμου (βλέπε παράγραφο 3.9).

### 3.6.3 Συμπλεγματοποίηση βασιζόμενη σε συνδέσμους

Σκοπός της συμπλεγματοποίησης είναι η ομαδοποίηση αντικειμένων με παρόμοιες ιδιότητες.

Στην συμπλεγματοποίηση που βασίζεται σε μοντέλα, υποθέτουμε ένα γενικευμένο μοντέλο  $P(X) = \sum_C P(C) P(X|C)$ , όπου  $X$  είναι ένα αντικείμενο, το  $C$  κινείται (λαμβάνει τιμές επί των συμπλεγμάτων και  $P(C|X)$  είναι ο βαθμός της ιδιότητας μέλους του  $X$  στο σύμπλεγμα  $C$ .

Στην συμπλεγματοποίηση που βασίζεται σε συνδέσμους, τα αντικείμενα συμπλεγματοποιούνται σύμφωνα με τους συνδέσμους τους (π.χ. αντικείμενα που συνδέονται πιο στενά είναι πιο πιθανό να ανήκουν στο ίδιο σύμπλεγμα) αλλά ίσως και σύμφωνα με τις ιδιότητες τους (σύμφωνα με την εργασία του Flake και της ομάδας αυτού).

Αυτό το πρόβλημα μπορεί να μορφοποιηθεί σε Μαρκοβιανή λογική διατυπώνοντας ένα μη παρατηρημένο (unobserved) κατηγορήματα  $C(x,v)$ , με την έννοια ότι “το  $x$  ανήκει στο σύμπλεγμα  $v$ ” και έχοντας τύπους στο MLN που να σχετίζονται με αυτό το κατηγορήματα καθώς και τα παρατηρημένα (observed) κατηγορήματα (π.χ.  $R(x_i, x_j)$  για τους συνδέσμους και  $A(x, v)$  για τις ιδιότητες).

Συνεπώς μπορεί πλέον να εκτελεστεί η συμπλεγματοποίηση η οποία βασίζεται σε συνδέσμους απλά μαθαίνοντας τις παραμέτρους του MLN και με τις ιδιότητες μέλους του συμπλέγματος (cluster memberships) να δίνονται από τις πιθανότητες των κατηγορημάτων  $C(x,v)$  τα οποία με την σειρά τους περιορίζονται από τα παρατηρημένα κατηγορήματα.

### 3.6.4 Μοντελοποίηση κοινωνικών δικτύων

Τα κοινωνικά δίκτυα είναι γράφοι των οποίων οι κόμβοι αναπαριστούν κοινωνικούς δράστες (π.χ. ανθρώπους) ενώ οι ακμές αναπαριστούν τις μεταξύ τους σχέσεις (π.χ. φιλία).

Η ανάλυση κοινωνικών δικτύων ασχολείται με την κατασκευή μοντέλων τα οποία συσχετίζουν τις ιδιότητες και τις σχέσεις (links) των δραστών.

Για παράδειγμα, η πιθανότητα δύο δράστες να συνδέονται μεταξύ τους μπορεί να εξαρτάται από τις ομοιότητες των ιδιοτήτων τους και αντιστρόφως δύο δράστες που συνδέονται είναι πιο πιθανό να έχουν κάποιες συγκεκριμένες ιδιότητες.

Αυτά τα μοντέλα είναι συνήθως δίκτυα Markov και μπορούν να αναπαρασταθούν με ακρίβεια με τύπους της μορφής  $\forall x \forall y \forall v R(x, y) \Rightarrow (A(x, v) \Leftrightarrow A(y, v))$ , όπου  $x$  και  $y$  είναι οι δράστες,  $R(x, y)$  είναι μια μεταξύ αυτών σχέση, το  $A(x, v)$  αναπαριστά κάποια ιδιότητα του  $x$  και το βάρος του τύπου δείχνει το πόσο ισχυρή είναι η συσχέτιση μεταξύ της σχέσης (relation) και της ομοιότητας των ιδιοτήτων (attribute similarity).

Για παράδειγμα, ένα μοντέλο που δηλώνει ότι οι φίλοι τείνουν να έχουν παρόμοιες καπνιστικές συνήθειες μπορεί να αναπαρασταθεί με τον τύπο

$\forall x \forall y \text{ Φίλοι}(x, y) \Rightarrow (\text{Καπνίζει}(x) \Leftrightarrow \text{Καπνίζει}(y))$  (βλέπε πίνακα 3.1).

Η λογική Markov, εκτός του ότι μπορεί να περιλαμβάνει ήδη υπάρχοντα μοντέλα κοινωνικών δικτύων, επιτρέπει επίσης να ορισθούν και ακόμα πιο "πλούσια" δίκτυα (π.χ. γράφοντας τύπους οι οποίοι να περιλαμβάνουν πολλαπλούς τύπους σχέσεων και πολλαπλές ιδιότητες, καθώς και πιο πολύπλοκες εξαρτήσεις μεταξύ αυτών).

### 3.6.5 Ταυτοποίηση αντικειμένων

Η ταυτοποίηση αντικειμένων είναι το πρόβλημα κατά το οποίο θέλουμε να προσδιορίσουμε ποιές εγγραφές της βάσης δεδομένων αναφέρονται στην ίδια οντότητα του πραγματικού κόσμου (π.χ. ποιές καταχωρήσεις σε μια βάση δεδομένων βιβλιογραφίας αναπαριστούν την ίδια δημοσίευση).

Αυτό το πρόβλημα είναι κρίσιμης σημασίας για πολλές εταιρείες, κρατικές υπηρεσίες καθώς και επιστημονικές εργασίες ευρείας κλίμακας.

Ένας τρόπος για την αναπαράσταση της σε Μαρκοβιανή λογική είναι να αφαιρεθεί η υπόθεση μοναδικών ονομάτων όπως περιγράφηκε στην παράγραφο 3.4, δηλαδή ορίζοντας ένα κατηγορήμα  $\text{Equals}(x, y)$  (ή  $x = y$  για συντομία) εννοώντας ότι "το  $x$  αναπαριστά την ίδια οντότητα πραγματικού κόσμου με το  $y$ ".

Αυτό το κατηγορήμα εφαρμόζεται τόσο στις εγγραφές όσο και στα πεδία τους (π.χ. "ICML" = "Intl. Conf. on Mach. Learn.").

Οι εξαρτήσεις μεταξύ ταιριάσματος εγγραφών και ταιριάσματος πεδίων μπορούν τότε να παρασταθούν με τύπους της μορφής  $\forall x \forall y \ x = y \Leftrightarrow f_i(x) = f_i(y)$ , όπου  $x$  και  $y$  είναι εγγραφές, ενώ  $f_i(x)$  είναι μια συνάρτηση που επιστρέφει την τιμή του  $i$ -οστού πεδίου της εγγραφής  $x$ .

Αυτή η προσέγγιση έχει εφαρμοστεί με επιτυχία στην απαλοιφή διπλοτύπων από την βάση δεδομένων Coqa η οποία περιλαμβάνει τις δημοσιεύσεις που αφορούν την επιστήμη υπολογιστών.

Επειδή αυτή η προσέγγιση επιτρέπει την διάδοση πληροφορίας από μια απόφαση ταιριάσματος (π.χ. μια γείωση του  $x = y$ ) σε μια άλλη απόφαση ταιριάσματος μέσω πεδίων που εμφανίζονται και στα δύο ζεύγη, μπορεί και να εκτελεί με αποτελεσματικό τρόπο συλλογική ταυτοποίηση αντικειμένων. Μάλιστα, στην περίπτωση των πειραμάτων που θα αναφέρουμε παρακάτω ξεπέρασε κατά πολύ ως προς την απόδοση (outperformed) την παραδοσιακή μέθοδο λήψης απόφασης ταιριάσματος όπου η κάθε απόφαση λαμβάνεται ανεξάρτητα από όλες τις άλλες.

Για παράδειγμα, το ταίριασμα (matching) δύο αναφορών μπορεί να μας επιτρέψει να προσδιορίσουμε ότι η "ICML" και η "MLC" αναπαριστούν την ίδια συνεδρίαση, το οποίο με την σειρά του μπορεί να μας βοηθήσει στο ταίριασμα ενός άλλου ζεύγους αναφορών όπου η μία να περιέχει την "ICML" και η άλλη την "MLC."

Η Μαρκοβιανή λογική επιτρέπει επίσης την εύκολη προσάρτηση επιπλέον πληροφοριών σε ένα σύστημα απαλοιφής διπλοτύπων.

Για παράδειγμα, μπορούμε να συμπεριλάβουμε το μεταβατικό κλείσιμο (transitive closure) προσθέτοντας τον τύπο  $\forall x \forall y \forall z \ x = y \wedge y = z \Rightarrow x = z$ , με ένα βάρος το οποίο θα προκύψει από τα δεδομένα.

### 3.7 Συμπερασμός

Σε αυτήν την παράγραφο θα δείξουμε πως μπορεί να γίνει συμπερασμός σε Μαρκοβιανή λογική.

Η Μαρκοβιανή λογική μπορεί να απαντήσει αυθαίρετες αναζητήσεις της μορφής "Ποια είναι η πιθανότητα ο τύπος  $F_1$  να ισχύει εάν γνωρίζουμε ότι ισχύει ο τύπος  $F_2$  ?"

Εάν οι  $F_1$  και  $F_2$  είναι δύο τύποι λογικής πρώτης τάξεως, το  $C$  είναι ένα πεπερασμένο σύνολο σταθερών το οποίο περιλαμβάνει οποιαδήποτε σταθερά που εμφανίζεται στον  $F_1$  ή στον  $F_2$  και  $L$  είναι ένα MLN, τότε

$$\begin{aligned} P(F_1|F_2, L, C) &= P(F_1|F_2, M_{L,C}) \\ &= \frac{P(F_1 \wedge F_2|M_{L,C})}{P(F_2|M_{L,C})} \\ &= \frac{\sum_{x \in \mathcal{X}_{F_1} \cap \mathcal{X}_{F_2}} P(X=x|M_{L,C})}{\sum_{x \in \mathcal{X}_{F_2}} P(X=x|M_{L,C})} \end{aligned} \quad (3.4)$$

Όπου  $\mathcal{X}_{F_i}$  είναι το σύνολο των κόσμων όπου ο  $F_i$  ισχύει και το  $P(x|M_{L,C})$  δίνεται από την σχέση (3.3).

Οι συνήθεις δεσμευμένες αναζητήσεις σε γραφικά μοντέλα είναι η ειδική περίπτωση της (3.4) όπου όλα τα κατηγορήματα στους  $F_1$ ,  $F_2$ , και  $L$  είναι μηδενικών ορισμάτων και οι τύποι είναι συζεύξεις.

Η ερώτηση για το αν μια βάση γνώσεως KB συνεπάγεται έναν τύπο F σε λογική πρώτης τάξεως είναι ίδια με την ερώτηση εάν  $P(F|L_{KB}, C_{KB,F})=1$ , όπου  $L_{KB}$  είναι το MLN που προκύπτει αν αναθέσουμε άπειρα βάρη σε όλους τους τύπους της KB και  $C_{KB,F}$  είναι το σύνολο όλων των σταθερών που εμφανίζονται στην KB ή στον F.

Η ερώτηση απαντάται υπολογίζοντας την  $P(F|L_{KB}, C_{KB,F})$  μέσω της εξίσωσης (3.4) με την συνθήκη  $F_2 = \text{Αληθής}$ .

Ο υπολογισμός της (3.4) θα είναι γενικά δύσκολη υπόθεση.

Αφού ο συμπερασμός στην λογική Markov εμπεριέχει πιθανοτικό συμπερασμό, ο οποίος είναι #P-complete και λογικό συμπερασμό σε πεπερασμένα πεδία, ο οποίος είναι NP-complete, δεν μπορούμε να περιμένουμε κάποιο καλύτερο αποτέλεσμα.

Παρ' όλα αυτά όμως, αρκετές από τις πραγματικά πάρα πολλές τεχνικές που χρησιμοποιούνται για επαρκή συμπερασμό μπορούν σε κάθε περίπτωση να εφαρμοστούν στην λογική Markov.

Επειδή η Μαρκοβιανή λογική επιτρέπει την κομψή κωδικοποίηση γνώσης, συμπεριλαμβανομένων των εξαρτήσεων από συγκεκριμένα συμφραζόμενα, ο συμπερασμός με αυτήν μπορεί σε αρκετές περιπτώσεις να είναι πιο επαρκής σε σχέση με τον συμπερασμό σε ένα συνηθισμένο γραφικό μοντέλο για το ίδιο ακριβώς πεδίο.

Από την πλευρά της λογικής, η πιθανοτική σημασιολογία της Μαρκοβιανής λογικής επιτρέπει τον προσεγγιστικό συμπερασμό, με αποτελεσματικότητα και στα αντίστοιχα δυνατά κέρδη.

Θεωρητικά, η  $P(F_1|F_2, L, C)$  μπορεί να προσεγγιστεί χρησιμοποιώντας έναν αλγόριθμο MCMC ο οποίος θα απορρίπτει όλες τις κινήσεις προς τις καταστάσεις όπου ο  $F_2$  δεν ισχύει και που θα μετράει τον αριθμό των δειγμάτων για τα οποία ο  $F_1$  ισχύει.

Ακόμα και αυτό όμως ενδέχεται να είναι πολύ αργό για αυθαίρετους τύπους.

Αντί αυτού, παρέχουμε έναν αλγόριθμο συμπερασμού για την περίπτωση που οι  $F_1$  και  $F_2$  είναι συζεύξεις λεκτημάτων γείωσης.

Αν και είναι λιγότερο γενικό από την εξίσωση (3.4), αυτός είναι ο συνηθέστερος τύπος αναζήτησης που χρησιμοποιείται στην πράξη και ο αλγόριθμος που παρέχουμε την απαντά πιο αποτελεσματικά από ότι θα έκανε μια απευθείας εφαρμογή της (3.4).

Η εξερεύνηση του lifted reasoning (όπου οι αναζητήσεις που περιέχουν μεταβλητές, απαντώνται χωρίς αυτές να γειωθούν) είναι μια σημαντική κατεύθυνση για μελλοντική έρευνα (βλέπε εργασία των Jaeger και Poole για κάποια αρχικά αποτελέσματα).

Ο αλγόριθμος εξελίσσεται σε δύο φάσεις, εντελώς ανάλογα με αυτό που γίνεται και κατά την κατασκευή μοντέλων η οποία βασίζεται σε ήδη υπάρχουσα γνώση.

Η πρώτη φάση επιστρέφει ως αποτέλεσμα το ελάχιστο υποσύνολο  $M$  του δικτύου γείωσης Markov (ground Markov network) το οποίο απαιτείται για τον υπολογισμό της  $P(F_1|F_2, L, C)$ .

Ο αλγόριθμος για αυτό φαίνεται στον πίνακα 3.3.



**Table 3.3** Network construction for inference in Markov logic

---

```

function ConstructNetwork( $F_1, F_2, L, C$ )
  inputs:  $F_1$ , a set of ground atoms with unknown truth-values (the “query”)
            $F_2$ , a set of ground atoms with known truth-values (the “evidence”)
            $L$ , a Markov logic network
            $C$ , a set of constants
  output:  $M$ , a ground Markov network
  calls:  $MB(q)$ , the Markov blanket of  $q$  in  $M_{L,C}$ 
   $G \leftarrow F_1$ 
  while  $F_1 \neq \emptyset$ 
    for all  $q \in F_1$ 
      if  $q \notin F_2$ 
         $F_1 \leftarrow F_1 \cup (MB(q) \setminus G)$ 
         $G \leftarrow G \cup MB(q)$ 
       $F_1 \leftarrow F_1 \setminus \{q\}$ 
  return  $M$ , the ground Markov network composed of all nodes in  $G$ , all arcs between
  them in  $M_{L,C}$ , and the features and weights on the corresponding cliques

```

---

Το μέγεθος του δικτύου που επιστρέφεται μπορεί να ελαττωθεί και άλλο και κατά αυτόν τον τρόπο να επιταχυνθεί ο αλγόριθμος, αν παρατηρήσουμε ότι οποιοσδήποτε τύπος γείωσης ο οποίος καθίσταται αληθής από τις αποδείξεις μπορεί να αγνοηθεί και οι αντίστοιχες καμπύλες να αφαιρεθούν από το δίκτυο.

Στην χειρότερη περίπτωση, το δίκτυο περιέχει  $O(|C|^a)$  κόμβους, όπου  $a$  είναι το μεγαλύτερο όρισμα κατηγορήματος στον τομέα αλλά στην πράξη μπορεί να είναι ακόμα μικρότερο.

Στην δεύτερη φάση γίνεται συμπερασμός σε αυτό το δίκτυο, με τους κόμβους στον  $F_2$  να έχουν τις τιμές που έχουν στον  $F_2$ .

Η δική μας υλοποίηση χρησιμοποιεί δειγματοληψία Gibbs αλλά μπορεί να χρησιμοποιηθεί οποιαδήποτε μέθοδος συμπερασμού.

Το βασικό βήμα Gibbs αποτελείται από την δειγματοληψία ενός ατόμου γείωσης δεδομένης της περιοχής κάλυψης Markov (Markov blanket) αυτού.

Η περιοχή κάλυψης Markov ενός ατόμου γείωσης είναι το σύνολο των κατηγορημάτων γείωσης που εμφανίζονται σε κάποια γείωση ενός τύπου με αυτό.

Η πιθανότητα ενός ατόμου γείωσης  $X_i$  όταν η περιοχή κάλυψης Markov αυτού είναι σε κατάσταση  $b_i$  είναι

$$P(X_i = x_i | B_i = b_i) = \frac{\exp(\sum_{f_i \in F_i} w_i f_i(X_i = x_i, B_i = b_i))}{\exp(\sum_{f_i \in F_i} w_i f_i(X_i = 0, B_i = b_i)) + \exp(\sum_{f_i \in F_i} w_i f_i(X_i = 1, B_i = b_i))} \quad (3.5)$$

Όπου  $F_i$  είναι το σύνολο των τύπων γείωσης στα οποία εμφανίζεται το  $X_i$  και  $f_i(X_i = x_i, B_i = b_i)$  είναι η τιμή (0 ή 1) του χαρακτηριστικού που αντιστοιχεί στον  $i$ -οστό τύπο γείωσης όταν  $X_i = x_i$  και  $B_i = b_i$ .

Για σύνολα ατόμων στα οποία ακριβώς ένα (και μόνο ένα) άτομο είναι αληθές για οποιοδήποτε κόσμο (π.χ. οι πιθανές τιμές κάποιας ιδιότητας), μπορεί να χρησιμοποιηθεί εμπόδιση (blocking) (δηλαδή ένα άτομο να καθίσταται αληθές και

όλα τα υπόλοιπα να καθίστανται ψευδή σε ένα βήμα με δεσμευμένη δειγματοληψία από την συλλογική τους περιοχή κάλυψης Markov- collective Markov blanket).

Η υπολογισμένη πιθανότητα μιας σύζευξης λεκτημάτων γείωσης είναι το κλάσμα των δειγμάτων για τα οποία τα λεκτήματα γείωσης είναι αληθή, αφού πρώτα έχει γίνει σύγκλιση της αλυσίδας Markov.

Επειδή η κατανομή πιθανότατα θα έχει πολλούς τρόπους (modes), "τρέχουμε" την αλυσίδα Markov πολλές φορές.

Όταν το MLN είναι σε προτασιακή μορφή (clausal form), μεγιστοποιούμε τον χρόνο burn-in ξεκινώντας κάθε "τρέξιμο" από έναν τρόπο (mode) που βρέθηκε με χρήση του MaxWalkSat, ενός αλγόριθμου τοπικής αναζήτησης που χρησιμοποιείται για το πρόβλημα ικανοποίησης με βάρη (weighted satisfiability problem), δηλαδή την εύρεση μιας ανάθεσης αληθοτιμών η οποία να μεγιστοποιεί το άθροισμα των βαρών των προτάσεων που ικανοποιούνται.

Όταν υπάρχουν ισχυροί περιορισμοί (προτάσεις με άπειρο βάρος), ο MaxWalkSat βρίσκει περιοχές που να τους ικανοποιούν και τότε ο δειγματολήπτης Gibbs δειγματοληπτεί από αυτές τις περιοχές προκειμένου να καταλήξει σε πιθανοτικά εκτιμώμενα.

### 3.8 Μάθηση

Τα βάρη του MLN μαθαίνονται από μία ή περισσότερες σχεσιακές βάσεις δεδομένων.

Για συντομία, η αντιμετώπιση παρακάτω γίνεται για μια βάση δεδομένων αλλά μπορεί να γίνει πολύ εύκολα γενίκευση για πολλές βάσεις δεδομένων)

Κάνουμε υπόθεση κλειστού κόσμου:

Δηλαδή, αν ένα άτομο γείωσης δεν βρίσκεται στην βάση δεδομένων υποθέτουμε ότι είναι ψευδές.

Αν υπάρχουν  $n$  πιθανά άτομα γείωσης, μια βάση δεδομένων μπορεί να παρασταθεί σαν ένα διάνυσμα  $x = (x_1, \dots, x_1, \dots, x_n)$ , όπου  $x_i$  είναι η αληθοτιμή του  $i$ -οστού ατόμου γείωσης ( $x_i = 1$  αν το άτομο εμφανίζεται στην βάση δεδομένων και  $x_i = 0$  αν δεν εμφανίζεται).

Δεδομένης μιας βάσης δεδομένων, τα βάρη ενός MLN μπορούν να μαθευτούν χρησιμοποιώντας την παρακάτω καθιερωμένη μέθοδο.

Αν ο  $i$ -οστός τύπος έχει  $n_i(x)$  αληθείς γειώσεις στα δεδομένα  $x$ , τότε από την εξίσωση (3.3) η παράγωγος της λογαριθμικής πιθανοφάνειας ως προς το βάρος της είναι

$$\frac{\partial}{\partial w_i} \log P_w(X=x) = n_i(x) - \sum_{x'} P_w(X=x') n_i(x') \quad (3.6)$$

Όπου το άθροισμα γίνεται επί όλων των δυνατών βάσεων δεδομένων  $x'$

και  $P_w(X=x')$  είναι η  $P(X=x')$  υπολογισμένη χρησιμοποιώντας το τρέχον διάνυσμα βαρών  $w = (w_1, \dots, w_i, \dots)$ .

Με άλλα λόγια, το  $i$ -οστό στοιχείο της κλίσης είναι η διαφορά μεταξύ του αριθμού των αληθών γειώσεων του  $i$ -οστού τύπου των δεδομένων και της προσδοκίας αυτού σύμφωνα με το τρέχον μοντέλο.



Δυστυχώς, η καταμέτρηση του αριθμού των αληθών γειώσεων ενός τύπου σε μια βάση δεδομένων είναι δύσκολη ακόμα και όταν ο τύπος είναι μια απλή πρόταση, όπως δηλώνεται στο αμέσως επόμενο.

Η καταμέτρηση του αριθμού των αληθών γειώσεων μιας πρότασης πρώτης τάξεως σε μια βάση δεδομένων είναι #P-complete ως προς το μήκος της πρότασης.

Σε μεγάλα πεδία, ο αριθμός των αληθών γειώσεων ενός τύπου μπορεί να μετρηθεί κατά προσέγγιση εκτελώντας ομοιόμορφη δειγματοληψία των γειώσεων του τύπου και ελέγχοντας αν είναι αληθείς για τα δεδομένα.

Σε μικρότερα πεδία, καθώς και στα πειράματα που θα δείξουμε παρακάτω, χρησιμοποιούμε έναν αποδοτικό αναδρομικό αλγόριθμο προκειμένου να βρούμε τον ακριβή αριθμό.

Ένα δεύτερο πρόβλημα με την εξίσωση (3.6) είναι ότι ο υπολογισμός του αναμενόμενου αριθμού των αληθών γειώσεων είναι επίσης πολύ δύσκολος καθώς απαιτεί συμπερασμό επί του μοντέλου.

Επιπλέον, οι αποδοτικές μέθοδοι βελτιστοποίησης επίσης απαιτούν τον υπολογισμό της λογαριθμικής πιθανοφάνειας (3.3) και κατά συνέπεια και της συνάρτησης διαμερισμού  $Z$ .

Αυτό μπορεί να γίνει κατά προσέγγιση χρησιμοποιώντας έναν εκτιμητή μέγιστης πιθανοφάνειας Monte Carlo (Monte Carlo Maximum Likelihood Estimator - MC-MLE).

Στα πειράματα μας όμως, η δειγματοληψία Gibbs που χρησιμοποιήθηκε για τον υπολογισμό της MC-MLEs και των κλίσεων δεν συνέκλινε σε κάποιο λογικό χρόνο και η χρήση των δειγμάτων από τις αλυσίδες που δεν είχαν συγκλίνει, έδιναν φτωχά αποτελέσματα.

Μια πιο αποδοτική εναλλακτική που χρησιμοποιείται ευρέως σε περιοχές όπως είναι οι χωρικές στατιστικές, η μοντελοποίηση κοινωνικών δικτύων και η επεξεργασία γλώσσας είναι η βελτιστοποίηση της ψευδο-πιθανοφάνειας

$$P_w^*(X=x) = \prod_{l=1}^n P_w(X_l=x_l | MB_x(X_l)) \quad (3.7)$$

Όπου  $MB_x(X_l)$  είναι η κατάσταση της περιοχής κάλυψης Markov του  $X_l$  στα δεδομένα.

Η κλίση της ψευδο-πιθανοφάνειας είναι

$$\frac{\partial}{\partial w_i} \log P_w^*(X=x) = \sum_{l=1}^n [n_i(x) - P_w(X_l=0 | MB_x(X_l)) n_i(x_{[X_l=0]}) - P_w(X_l=1 | MB_x(X_l)) n_i(x_{[X_l=1]})] \quad (3.8)$$

Όπου  $n_i(x_{[X_i=0]})$  είναι ο αριθμός των αληθών γειώσεων του  $i$ -οστού τύπου όταν επιβάλουμε να είναι  $X_i = 0$  ενώ αφήσουμε τα υπόλοιπα δεδομένα άθικτα. Όμοια ισχύουν και για το  $n_i(x_{[X_i=1]})$ .

Ο υπολογισμός αυτής της έκφρασης ή της (3.7) δεν απαιτεί συμπερασμό επί του μοντέλου.

Βελτιστοποιούμε την λογαριθμική ψευδο-πιθανοφάνεια (pseudo-log-likelihood), χρησιμοποιώντας τον αλγόριθμο περιορισμένης μνήμης BFGS.

Ο υπολογισμός μπορεί να γίνει πιο αποδοτικός με διάφορους τρόπους:

-Το άθροισμα στην (3.8) μπορεί να επιταχυνθεί κατά πολύ αν αγνοήσουμε τα κατηγορήματα που δεν εμφανίζονται στον  $i$ -οστό τύπο.

-Οι καταμετρήσεις των  $n_i(x)$ ,  $n_i(x_{[X_i=0]})$  και  $n_i(x_{[X_i=1]})$  δεν αλλάζουν με τα βάρη και χρειάζεται να υπολογιστούν μόνο μία φορά (σε αντίθεση με αυτό που πρέπει να γίνεται σε κάθε επανάληψη του BFGS).

-Οι τύποι γείωσης των οποίων οι αληθοτιμές παραμένουν ανεπηρέαστες από την αλλαγή της αληθοτιμής οποιουδήποτε λεκτήματος μπορούν να αγνοηθούν αφού τότε ισχύει ότι  $n_i(x) = n_i(x_{[X_i=0]}) = n_i(x_{[X_i=1]})$ .

Συγκεκριμένα, αυτό ισχύει για κάθε πρόταση η οποία περιέχει τουλάχιστον δύο αληθή λεκτήματα. Έτσι είναι άλλωστε και η πλειοψηφία των προτάσεων γείωσης.

Προκειμένου να αντιμετωπιστεί το πρόβλημα του overfitting, ποινικοποιούμε την ψευδο-πιθανοφάνεια με μια Gaussian μπροστά από κάθε βάρος.

Όταν γνωρίζουμε a priori ποια κατηγορήματα θα είναι αποδεικτικά, τα βάρη των MLN μπορούν να μαθευτούν ξεχωριστά.

Οι τεχνικές ILP μπορούν να χρησιμοποιηθούν προκειμένου να μάθουμε επιπλέον προτάσεις, να διωλίσουμε αυτές που ήδη υπάρχουν στο MLN ή να μάθουμε ένα MLN εκ του μηδενός.

Εδώ χρησιμοποιούμε συστήματα CLAUDIEN για αυτόν τον σκοπό.

Σε αντίθεση με τα περισσότερα ILP συστήματα, τα οποία μαθαίνουν προτάσεις Horn, το CLAUDIEN μπορεί να μάθει αυθαίρετες προτάσεις πρώτης τάξεως καθιστώντας το ιδιαίτερας ταιριαστό με την Μαρκοβιανή λογική.

Επίσης, κατασκευάζοντας μια συγκεκριμένη πόλωση γλώσσας μπορούμε να κατευθύνουμε το CLAUDIEN να αναζητήσει διυλίσεις (refinements) της δομής του MLN.

Εναλλακτικά, η δομή MLN μπορεί να μαθευτεί με απευθείας βελτιστοποίηση της ψευδο-πιθανοφάνειας.

### 3.9 Πειράματα

Οι αλγόριθμοι που περιγράφηκαν στα παραπάνω δοκιμάστηκαν πάνω σε μια βάση δεδομένων η οποία περιγράφει το τμήμα της επιστήμης υπολογιστών και μηχανικής του πανεπιστημίου της Washington (Department of Computer Science and Engineering at the University of Washington - UW-CSE).

Το πεδίο αποτελείται από 12 κατηγορήματα και 2707 σταθερές που χωρίζονται σε 10 τύπους.

Αυτοί οι τύποι περιλαμβάνουν:

δημοσίευση (342 σταθερές), πρόσωπο (442), μάθημα (176), project (153), ακαδημαϊκή περίοδος (20) κλπ.

Τα κατηγορήματα περιλαμβάνουν:

Καθηγητής (πρόσωπο), Φοιτητής (πρόσωπο), Περιοχή (x,περιοχή) με το x να λαμβάνει τιμές από δημοσιεύσεις, πρόσωπα, μαθήματα και projects,

Συγγραφέας Του (έκδοση, πρόσωπο), Υπό Την Επίβλεψη Του (πρόσωπο, πρόσωπο),

Έτη στο Πρόγραμμα (πρόσωπο, χρόνια), Επίπεδο Μαθήματος (μάθημα, επίπεδο),

Διδάσκεται Από (μάθημα, πρόσωπο, περίοδος), Βοηθός Καθηγητή (μάθημα, πρόσωπο, περίοδος) κλπ.

Επιπλέον υπάρχουν 10 κατηγορήματα ισότητας:

Ίδιο Πρόσωπο (πρόσωπο, πρόσωπο), Ίδιο Μάθημα (μάθημα, μάθημα) κλπ. τα οποία έχουν πάντα γνωστές και σταθερές τιμές και που είναι αληθή αν και μόνο αν τα δύο ορίσματα είναι η ίδια σταθερά.

Με την χρήση μεταβλητών που ανήκουν σε διάφορους τύπους, το συνολικό πλήθος των δυνατών ατόμων γείωσης (n στην παράγραφο 3.8) ήταν 4,106,841.

Η βάση δεδομένων περιείχε συνολικά 3380 n-άδες (δηλαδή υπήρχαν 3380 αληθή άτομα γείωσης).

Η συγκεκριμένη βάση δεδομένων προέκυψε από σελίδες του website του τμήματος ([www.cs.washington.edu](http://www.cs.washington.edu)).

Οι Εκδόσεις και οι σχέσεις Συγγραφέας Του προέκυψαν με εξαγωγή από την βάση δεδομένων BibServ ([www.bibserv.org](http://www.bibserv.org)) όλων των εγγραφών με πεδία συγγραφέα που περιείχαν τα ονόματα τουλάχιστον δύο μελών του τμήματος (με την μορφή “επώνυμο, όνομα” ή “επώνυμο, αρχικό”).

Η βάση γνώσεως αποκτήθηκε μέσω παράκλησης σε τέσσερεις εθελοντές να παρέχουν ένα σύνολο τύπων σε λογική πρώτης τάξεως που να περιγράφουν τον τομέα. Οι εθελοντές ήταν μέλη του τμήματος και για αυτό μπορούσαν να περιγράψουν με ακρίβεια τον τομέα.

Η συγχώνευση όλων των παραπάνω έδωσε μια KB που περιλάμβανε 96 τύπους.

Ολόκληρη η KB, οι οδηγίες προς τους εθελοντές, η βάση δεδομένων και οι ρυθμίσεις των παραμέτρων των αλγορίθμων βρίσκονται online στην διεύθυνση <http://www.cs.washington.edu/ai/mln>.

Τύποι στην KB περιλαμβάνουν δηλώσεις όπως:

Οι φοιτητές δεν είναι καθηγητές, κάθε φοιτητής έχει το πολύ ένα σύμβουλο, αν ένας φοιτητής είναι συγγραφέας μίας δημοσίευσης τότε αυτομάτως είναι και ο σύμβουλος, το πολύ ένας συγγραφέας μιας δεδομένης δημοσίευσης είναι καθηγητής, οι φοιτητές που βρίσκονται στο πρώτο στάδιο του διδακτορικού τους δεν έχουν σύμβουλο κλπ.

Αυτές οι δηλώσεις δεν είναι πάντοτε αληθείς αλλά είναι τυπικά αληθείς.

Για εκπαιδευτικούς και δοκιμαστικούς σκοπούς η βάση δεδομένων χωρίστηκε σε πέντε υποβάσεις, μία για κάθε περιοχή: τεχνητή νοημοσύνη, γραφική, γλώσσες προγραμματισμού, συστήματα και θεωρία.

Οι καθηγητές και τα μαθήματα ανατέθηκαν χειροκίνητα σε περιοχές (areas) και άλλες σταθερές ανατέθηκαν με επαναληπτικό τρόπο στην περιοχή (area) που συναντούσαμε πιο συχνά, μεταξύ άλλων σταθερών με τις οποίες εμφανίζονταν μαζί σε κάποια n-άδα.

Κατόπιν, κάθε n-άδα ανατέθηκε στην περιοχή που αντιστοιχούσε στις μεταβλητές που περιείχε.

Οι n-άδες που περιείχαν σταθερές που αντιστοιχούσαν σε περισσότερες από μία περιοχές δεν λήφθηκαν υπόψη ώστε να αποφευχθεί το φαινόμενο "train-test contamination".

Οι υποβάσεις περιείχαν κατά μέσο όρο 521 αληθή άτομα γείωσης από τα 58,457 δυνατά.

Εκτελέστηκαν δοκιμές τύπου leave-one-out κατά περιοχή.

Σκοπός της εργασίας δοκιμής ήταν να προβλεφθεί το κατηγορήμα Συμβουλευεται Από (x, y) δεδομένων:

(α) όλων των άλλων (All Info)

(b) όλων των άλλων εκτός από Φοιτητής (x) και Καθηγητής (x) (Partial Info).

Μετρήθηκε και στις δύο περιπτώσεις η μέση δεσμευμένη λογαριθμική πιθανοφάνεια όλων των δυνατών γειώσεων του κατηγορήματος Συμβουλευεται Από (x, y) επί όλων των περιοχών, σχεδιάστηκαν οι καμπύλες σε άξονες ακρίβεια-ανάκληση και υπολογίστηκε η περιοχή κάτω από την καμπύλη.

Αυτή η εργασία είναι μια περίπτωση πρόβλεψης δεσμού, ένα πρόβλημα που υπήρξε αντικείμενο μεγάλου ενδιαφέροντος στην στατιστική σχεσιακή μάθηση (βλέπε παράγραφο 3.6).

Όλες οι KBs μετατράπηκαν σε προτασιακή μορφή ενώ τα χρονικά αποτελέσματα έγιναν σε έναν Pentium 4 στα 2.8Ghz.

### 3.9.1 Συστήματα

Προκειμένου να αποτιμηθεί η Μαρκοβιανή λογική, η οποία όπως είδαμε χρησιμοποιεί λογική και πιθανότητες για συμπερασμό, θελήσαμε να την συγκρίνουμε με μεθόδους που χρησιμοποιούν μόνο λογική ή μόνο πιθανότητες.

Επίσης έχει ενδιαφέρον να γίνει αναφορά στην αυτόματη επαγωγή προτάσεων χρησιμοποιώντας τεχνικές ILP.

Σε αυτήν την παράγραφο δίνονται λεπτομέρειες σχετικά με αυτές τις συγκρίσεις.

### 3.9.1.1 Λογική

Μια σημαντική ερώτηση που θέλαμε να απαντηθεί από τα πειράματα είναι το αν η προσθήκη πιθανότητας σε μια λογική KB βελτιώνει την ικανότητα της να μοντελοποιεί τον τομέα.

Για να το κάνουμε αυτό πρέπει να παρατηρήσουμε τα αποτελέσματα των αναζητήσεων που απαντήθηκαν χρησιμοποιώντας συμπερασμό ο οποίος βασίζεται μόνο στην λογική, κάτι τέτοιο όμως περιπλέκεται από το γεγονός ότι ο υπολογισμός της λογαριθμικής πιθανοφάνειας αλλά και της περιοχής κάτω από την καμπύλη ακρίβειας-ανάκλησης απαιτεί πιθανότητες πραγματικής τιμής (real-valued probabilities) ή τουλάχιστον κάποιο μέτρο "εμπιστοσύνης" για την αλήθεια (truth) που μπορεί να φέρει κάθε άτομο γείωσης το οποίο δοκιμάζεται.

Εξαιτίας αυτού χρησιμοποιήθηκε η παρακάτω προσέγγιση.

Για μια γνωστή βάση γνώσεως KB και ένα γνωστό σύνολο αποδεικτικών ατόμων (evidence atoms) E, έστω  $X_{KB \cup E}$  το σύνολο των κόσμων που ικανοποιούν την ένωση  $KB \cup E$ .

Τότε, η πιθανότητα ενός ατόμου αναζήτησης (query atom) q ορίζεται από την σχέση

$$P(q) = \frac{|X_{KB \cup E \cup q}|}{|X_{KB \cup E}|}$$

Δηλαδή από το μέρος του συνόλου  $X_{KB \cup E}$  στο οποίο το q είναι αληθές.

Ένα ακόμα πιο σοβαρό πρόβλημα προκύπτει αν η KB δεν είναι συνεπής (κάτι το οποίο πράγματι συνέβη στην περίπτωση της KB που συλλέχθηκε με την βοήθεια εθελοντών).

Σε αυτή την περίπτωση, ο παρανομαστής της  $P(q)$  είναι μηδέν.

Επίσης να υπενθυμίσουμε ότι μια ασυνεπής KB συνεπάγεται επουσιωδώς οποιαδήποτε αυθαίρετο τύπο.

Για να αντιμετωπίσουμε αυτό το πρόβλημα, επαναπροσδιορίζουμε το  $X_{KB \cup E}$  ώστε να είναι το σύνολο των κόσμων το οποίο ικανοποιεί τον μεγαλύτερο δυνατό αριθμό προτάσεων γείωσης.

Προκειμένου να εκτελέσουμε δειγματοληψία από αυτό το σύνολο, χρησιμοποιούμε δειγματοληψία Gibbs και με κάθε αλυσίδα να ξεκινάει από έναν τρόπο (mode) ο οποίος βρέθηκε με χρήση του αλγορίθμου WalkSat.

Σε κάθε βήμα Gibbs, το βήμα λαμβάνεται με πιθανότητα ίση με:

1 , αν η νέα κατάσταση ικανοποιεί περισσότερες προτάσεις από την ήδη υπάρχουσα κατάσταση (αφού αυτό σημαίνει ότι η παρούσα κατάσταση θα έπρεπε να έχει μηδενική πιθανότητα)

0.5 , αν η νέα κατάσταση ικανοποιεί τον ίδιο αριθμό προτάσεων (αφού τότε η νέα και η παλιά κατάσταση είναι ισοπίθανες)

0 , αν η νέα κατάσταση ικανοποιεί λιγότερες προτάσεις

Κατόπιν, χρησιμοποιούμε μόνο τις καταστάσεις με τον μεγαλύτερο αριθμό ικανοποιούμενων προτάσεων προκειμένου να υπολογίσουμε τις πιθανότητες.

Να σημειωθεί ότι κάτι τέτοιο είναι ισοδύναμο με το να χρησιμοποιηθεί ένα MLN το οποίο κατασκευάζεται από την KB και με όλα τα βάρη ίσα και άπειρα.

### 3.9.1.2 Πιθανότητα

Η άλλη ερώτηση που θέλαμε να απαντηθεί με την βοήθεια πειραμάτων είναι αν τα υπάρχοντα (προτασιακά) πιθανοτικά μοντέλα είναι ήδη αρκετά ισχυρά ώστε να μπορούν να χρησιμοποιηθούν σε σχεσιακά πεδία (relational domains) χωρίς να χρειάζονται τις επιπλέον δυνατότητες αναπαράστασης που παρέχονται από τα MLNs. Προκειμένου να χρησιμοποιηθούν τέτοια μοντέλα, ο τομέας πρέπει πρώτα να προτασιοποιηθεί ορίζοντας χαρακτηριστικά (features) τα οποία περικλείουν χρήσιμες πληροφορίες για αυτόν.

Η δημιουργία καλών ιδιοτήτων (attributes) για εκπαιδευόμενους προτασιακά σε αυτόν τον ιδιαίτερα σχετικιστικό τομέα είναι ένα δύσκολο πρόβλημα.

Με τον συμβιβασμό όμως μεταξύ προσάρτησης όσο το δυνατόν περισσότερο σχετικής πληροφορίας και της αποφυγής ιδιαίτερας μεγάλων διανυσμάτων χαρακτηριστικών (feature vectors), ορίστηκαν δύο σύνολα προτασιακών ιδιοτήτων (propositional attributes) : το σύνολο τάξεως-1(order-1) και το σύνολο τάξεως-2 (order-2).

Το σύνολο order-1 έχει να κάνει με τα χαρακτηριστικά μεμονωμένων σταθερών που υπάρχουν στο κατηγορημα αναζήτησης (query predicate) ενώ το σύνολο order-2 έχει να κάνει με τα χαρακτηριστικά των σχέσεων μεταξύ των σταθερών που υπάρχουν στο κατηγορημα αναζήτησης.

Για τις ιδιότητες του συνόλου order-1, ορίστηκε μια μεταβλητή για κάθε ζεύγος (a, b), όπου a είναι ένα όρισμα του κατηγορήματος αναζήτησης και b είναι ένα όρισμα κάποιου κατηγορήματος που έχει την ίδια τιμή με το a.

Η μεταβλητή αυτή είναι το μέρος των αληθών γειώσεων αυτού του κατηγορήματος στα δεδομένα.

Μερικά παραδείγματα ιδιοτήτων πρώτης τάξεως για το Συμβουλευεται Από (Matt, Pedro) είναι:

Αν ο Pedro είναι φοιτητής, το κλάσμα των δημοσιεύσεων που έχουν δημοσιευθεί από τον Pedro, το κλάσμα των μαθημάτων για τα οποία ο Matt ήταν βοηθός καθηγητή κλπ.

Οι ιδιότητες του συνόλου order-2 ορίστηκαν ως εξής:

Για ένα δεδομένο (γειωμένο) κατηγορήμα αναζήτησης  $Q(q_1, q_2, \dots, q_k)$ , θεωρούμε όλα τα σύνολα  $k$  κατηγορημάτων και όλες τις αναθέσεις των σταθερών  $q_1, q_2, \dots, q_k$  σαν ορίσματα για τα  $k$  κατηγορήματα, με ακριβώς μια σταθερά για κάθε κατηγορήμα (για οποιαδήποτε σειρά).

Για παράδειγμα, αν ο  $Q$  Συμβουλευεται Από (Matt, Pedro) τότε ένα τέτοιο πιθανό σύνολο θα ήταν το {Βοηθός Μαθήματος ( \_, Matt, \_), Διδάσκεται Από ( \_, Pedro, \_)}.

Αυτό σχηματίζει  $2^k$  ιδιότητες του παραδείγματος, με κάθε μια να αντιστοιχεί σε μια συγκεκριμένη αληθή ανάθεση των  $k$  κατηγορημάτων.

Η τιμή μιας ιδιότητας (attribute) είναι ο αριθμός των φορών που το σύνολο των κατηγορημάτων έχει την συγκεκριμένη αληθο-ανάθεση (εντός των εκπαιδευτικών δεδομένων), όταν τα ορίσματα στα οποία δεν έχει γίνει ανάθεση περιέχουν όλα τις ίδιες σταθερές.

Για παράδειγμα, ας θεωρήσουμε ότι συμπληρώνουμε τα προηγούμενα κενά ορίσματα με “CSE546” και “Χειμερινό\_0304.”

Το προκύπτον σύνολο { Βοηθός Μαθήματος (CSE546, Matt, Autumn\_0304), Διδάσκεται Από (CSE546, Pedro, Χειμερινό\_0304)}

έχει κάποια αληθο-ανάθεση για τα εκπαιδευτικά δεδομένα, π.χ. {Αληθές, Αληθές}, {Ψευδές, Ψευδές}, ...

Μια ιδιότητα είναι το πλήθος τέτοιων συνόλων σταθερών τα οποία δημιουργούν την αληθο-ανάθεση { Αληθές, Αληθές }, μια άλλη για την αληθο-ανάθεση { Αληθές, Ψευδές } κοκ.

Μερικά παραδείγματα ιδιοτήτων δευτέρας τάξεως που δημιουργήθηκαν για την αναζήτηση Συμβουλευεται Από (Matt, Pedro) είναι:

Πόσο συχνά είναι ο Matt βοηθός καθηγητή για ένα μάθημα που δίδαξε ο Pedro (καθώς και πόσες φορές δεν ήταν βοηθός καθηγητή), πόσες δημοσιεύσεις συνέγραψαν ο Pedro και ο Matt κλπ.

Οι 28 προκύπτουσες ιδιότητες πρώτης τάξεως και οι 120 ιδιότητες δευτέρας τάξεως (για την περίπτωση All Info) χωρίστηκαν και μοιράστηκαν σε πέντε κάδους ίσης συχνότητας (βάσει του εκπαιδευτικού συνόλου).

Χρησιμοποιήθηκαν δύο προτασιακοί εκπαιδευόμενοι (propositional learners): απλοϊκός Bayes και Bayesian δίκτυα με δομή και παραμέτρους που μαθεύτηκαν χρησιμοποιώντας τον αλγόριθμο VFBN2, έχοντας το πολύ τέσσερις γονείς για κάθε κόμβο.

Οι ιδιότητες δευτέρας τάξεως βοήθησαν τον απλοϊκό Bayes κατηγοριοποιητή αλλά έβλαψαν την απόδοση του κατηγοριοποιητή Bayesian δικτύου, οπότε στα αποτελέσματα που θα αναφερθούν παρακάτω να ξεκαθαριστεί ότι χρησιμοποιήθηκαν ιδιότητες πρώτης τάξεως αλλά και ιδιότητες δευτέρας τάξεως για τον naive Bayes ενώ για τα Bayesian δίκτυα χρησιμοποιήθηκαν μόνο ιδιότητες πρώτης τάξεως.

### 3.9.1.3 Επαγωγικός λογικός προγραμματισμός

Η αρχική KB αποκτήθηκε από εθελοντές. Θα μας ενδιέφερε όμως το αν θα μπορούσε αυτή η KB να προκύψει αυτόματα με χρήση μεθόδων ILP.

Όπως αναφέρθηκε νωρίτερα, χρησιμοποιήθηκε CLAUDIEN για να επαχθεί μια KB από τα δεδομένα.

Το CLAUDIEN "τρέχτηκε" με: τοπική εμβέλεια, ελάχιστη ακρίβεια ίση με 0.1, ελάχιστη κάλυψη ίση με 1, μέγιστη πολυπλοκότητα ίση με 10 και αναζήτηση BFS (Breadth-First Search).

Ο χώρος αναζήτησης του CLAUDIEN ορίζεται από την πόλωση γλώσσας αυτού.

Κατασκευάστηκε μια πόλωση γλώσσας η οποία επέτρεψε: ένα μέγιστο τριών μεταβλητών μέσα σε μια πρόταση, απεριόριστα κατηγορήματα μέσα σε μία πρόταση, μέχρι δύο μη αρνητικές και δύο αρνητικές εμφανίσεις ενός κατηγορήματος μέσα σε μια πρόταση καθώς και χρήση της γνώσεως των τύπων των ορισμάτων των κατηγορημάτων.

Προκειμένου να ελαχιστοποιηθεί η αναζήτηση, δεν χρησιμοποιήθηκαν στον CLAUDIEN κατηγορήματα ισότητας. Αυτό βελτίωσε τα αποτελέσματα του.

Εκτός από την επαγωγή προτάσεων από τα εκπαιδευτικά δεδομένα, είχε επίσης ενδιαφέρον το να χρησιμοποιηθούν δεδομένα για την αυτόματη διύλιση της KB που παρήχθη από τους εθελοντές.

Ο CLAUDIEN δεν υποστηρίζει άμεσα αυτή την ιδιότητα αλλά μπορεί να την μιμηθεί εάν κατασκευαστεί η κατάλληλη πόλωση γλώσσας.

Αυτό επετεύχθη επιτρέποντας στον CLAUDIEN (για κάθε πρόταση της KB) :

- (1) να αφαιρέσει οποιονδήποτε αριθμό λεκτημάτων
- (2) να προσθέσει μέχρι και  $n$  νέες μεταβλητές
- (3) να προσθέσει μέχρι και 1 νέο λέκτημα



Ο CLAUDIEN " έτρεξε" για 24 ώρες σε έναν Sun-Blade 1000 για κάθε  $(v, l)$  μέσα στο σύνολο  $\{(1, 2), (2, 3), (3, 4)\}$ .

Προέκυψαν σχεδόν τα ίδια αποτελέσματα και για τις τρεις περιπτώσεις.

Παρακάτω αναφέρονται τα αποτελέσματα για  $v = 3$  και  $l = 4$ .

#### 3.9.1.4 Λογική Markov

Τα αποτελέσματα συγκρίνουν τα παραπάνω συστήματα με την Μαρκοβιανή λογική. Τα MLNs εκπαιδεύτηκαν χρησιμοποιώντας μπροστά Γκαουσιανό βάρος με μηδενική μέση τιμή και μοναδιαία διακύμανση. Τα βάρη αρχικοποιήθηκαν στον τρόπο της προηγούμενης (prior).

Για την βελτιστοποίηση χρησιμοποιήθηκε υλοποίηση FORTRAN του αλγορίθμου L-BFGS που προτάθηκε από τον Zhu και την ομάδα αυτού καθώς και του Byrd και των συνεργατών του , αφήνοντας όλες τις παραμέτρους στις αρχικές τους τιμές και με κριτήριο σύγκλισης 10<sup>-5</sup>.

Ο συμπερασμός, έγινε με χρήση δειγματοληψίας Gibbs όπως περιγράφηκε στην παράγραφο 3.7, με δέκα αλυσίδες Markov που η κάθε μια αρχικοποιήθηκε σε έναν τρόπο της κατανομής χρησιμοποιώντας τον αλγόριθμο MaxWalkSat.

Ο αριθμός των βημάτων Gibbs προσδιορίστηκε χρησιμοποιώντας το κριτήριο των DeGroot και Schervish.

Η δειγματοληψία συνεχίστηκε μέχρι που επετεύχθη ένα διάστημα εμπιστοσύνης 95% όπου η εκτίμηση ήταν με ανοχή 1% της πραγματικής τιμής για ποσοστό τουλάχιστον 95% των κόμβων (αγνοώντας τους κόμβους που είναι μόνιμα αληθείς ή ψευδείς).

Το ελάχιστο των δειγμάτων που χρησιμοποιήθηκαν ήταν 1000 και το μέγιστο ήταν 500,000 και με ένα δείγμα για κάθε πλήρες πέρασμα Gibbs μέσα από τις μεταβλητές.

Τυπικά, ο συμπερασμός συνέκλινε μεταξύ 5000 και 100,000 περασμάτων.

Τα αποτελέσματα ήταν ανεπηρέαστα σε διακυμάνσεις των κατωφλίων σύγκλισης.

### 3.9.2 Αποτελέσματα

#### 3.9.2.1 Εκπαίδευση με MC-MLE

Το αρχικό σύστημα χρησιμοποίησε MC-MLE για την εκπαίδευση των MLNs, με δέκα αλυσίδες Gibbs και κάθε άτομο γείωσης να αρχικοποιείται ως αληθές με την αντίστοιχη πιθανότητα το κατηγορήματα πρώτης τάξεως να είναι αληθές στα δεδομένα.

Τα βήματα Gibbs μπορούν να γίνουν πολύ γρήγορα βλέποντας ότι λίγες καταμετρήσεις ικανοποιημένων προτάσεων μπορεί να αλλάζουν σε κάθε βήμα.

Στον τομέα UW-CSE, η υλοποίηση πήρε 4-5 ms για κάθε βήμα.

Προκειμένου να προσδιοριστεί πότε οι αλυσίδες έφταναν στην στάσιμη κατανομή τους χρησιμοποιήθηκε το μέγιστο του κριτηρίου Gelman R για όλα τα κατηγορήματα. Για την επιτάχυνση της σύγκλισης, ο δειγματολήπτης Gibbs λάμβανε κατά προτίμηση ως δείγματα άτομα τα οποία ήταν αληθή είτε στα δεδομένα είτε στην αρχική κατάσταση της αλυσίδας.

Αυτό έγινε επειδή τα περισσότερα άτομα είναι πάντα ψευδή και η κατά επανάληψη λήψη δείγματος από αυτά δεν είναι αποτελεσματική.

Αυτό βελτίωσε την σύγκλιση κατά περίπου μια τάξη πλάτους για ομοιόμορφη επιλογή ατόμων.

Παρά τις βελτιστοποιήσεις όμως, ο δειγματολήπτης Gibbs έλαβε έναν απαγορευτικά μεγάλο χρόνο για να φτάσει σε ένα λογικό κατώφλι σύγκλισης (π.χ.  $R = 1.01$ ).

Μετά από "τρέξιμο" 24 ωρών (περίπου 2 εκατομμύρια βήματα Gibbs ανά αλυσίδα), η μέση τιμή του R επί των εκπαιδευτικών συνόλων έγινε 3.04, με κανένα εκπαιδευτικό σύνολο να μην έχει φτάσει τιμή για το R μικρότερη από 2 (στα πρώτα στάδια της διαδικασίας υπήρξε για μικρό χρόνο βύθιση ως το 1.5).

Δεδομένου ότι αυτό πρέπει να γίνει επαναληπτικά καθώς ο L-BFGS ψάχνει για το ελάχιστο, εκτιμήθηκε ότι θα χρειαζόντουσαν ακόμα 20 με 400 ημέρες για να ολοκληρωθεί η εκπαίδευση, ακόμα και για ένα ασθενές κατώφλι σύγκλισης όπως το  $R = 2.0$ .

Τα πειράματα επιβεβαίωσαν την πτωχή ποιότητα των μοντέλων που προέκυψαν αν αγνοηθεί το κατώφλι σύγκλισης και περιοριστεί η εκπαιδευτική διαδικασία σε λιγότερο από δέκα ώρες.

Με καλύτερη επιλογή της αρχικής κατάστασης, προσεγγιστική καταμέτρηση και βελτιωμένες MCMC τεχνικές όπως ο αλγόριθμος Swendsen- Wang, η MC-MLE ενδεχομένως γίνει πιο πρακτική αλλά δεν μπορεί να θεωρηθεί βιώσιμη επιλογή για εκπαίδευση στην τρέχουσα έκδοση της.

### **3.9.2.2 Εκπαίδευση με ψευδο-πιθανοφάνεια**

Σε αντίθεση με το MC-MLE, η εκπαίδευση με ψευδο-πιθανοφάνεια ήταν αρκετά γρήγορη.

Όπως συζητήθηκε στην παράγραφο 3.8, κάθε επανάληψη της εκπαίδευσης μπορεί να γίνει πολύ γρήγορα εφόσον έχουν ολοκληρωθεί πρώτα η αρχική πρόταση και οι καταμετρήσεις ικανοποίησης των ατόμων γείωσης.

Από εκεί και πέρα, η εκπαίδευση χρειάστηκε κατά μέσο όρο 255 επαναλήψεις του L-BFGS για συνολικό χρόνο 16 λεπτών.

### 3.9.2.3 Συμπερασμός

Ο συμπερασμός ήταν επίσης πολύ γρήγορος.

Ο συμπερασμός της πιθανότητας όλων των ατόμων Συμβουλευεται Από (x, y) για την περίπτωση του All Info, πήρε 3.3 λεπτά για το σύνολο των AI test (4624 άτομα), 24.4 για την γραφική (3721), 1.8 για τις γλώσσες προγραμματισμού (784), 10.4 για τα συστήματα (5476) και 1.6 για την θεωρία (2704).

Ο αριθμός των περασμάτων Gibbs κυμάνθηκε από 4270 ως 500,000 και ο μέσος όρος ήταν 124,000.

Αυτό αντιστοιχεί σε 18 ms ανά πέρασμα Gibbs και περίπου 200,000–500,000 βήματα Gibbs ανά δευτερόλεπτο.

Ο μέσος χρόνος για την πραγματοποίηση συμπερασμού για την περίπτωση του Partial Info ήταν 14.8 λεπτά (συγκριτικά με τα 8.3 λεπτά για την περίπτωση του All Info).

### 3.9.2.4 Σύγκριση συστημάτων

Συγκρίθηκαν 12 συστήματα:

- Η αρχική KB (KB)
- CLAUDIEN (CL)
- CLAUDIEN με την αρχική KB ως πόλωση γλώσσας (CLB)
- Η ένωση της αρχικής KB και των αποτελεσμάτων CLAUDIEN και στις δύο περιπτώσεις (KB+CL και KB+CLB)
- Ένα MLN με κάθε μια από τις παραπάνω KBs (MLN(KB), MLN(CL), MLN(KB+CL) και MLN(KB+CLB))
- Naive Bayes (NB)
- Ένας Bayesian network learner (BN)

Σε όλες τις περιπτώσεις χρησιμοποιήθηκε Add-one εξομάλυνση των πιθανοτήτων.

Στον πίνακα 3.4 συνοψίζονται τα αποτελέσματα και στο σχήμα 3.2 φαίνονται οι καμπύλες ακρίβειας-ανάκλησης για όλες τις περιοχές (δηλαδή η μέση τιμή για όλα τα κατηγορήματα Συμβουλευεται Από (x, y)).

Τα MLNs είναι προφανώς πιο ακριβή σε σχέση με όλες τις υπόλοιπες εναλλακτικές, αναδεικνύοντας τις προοπτικές αυτής της προσέγγισης.

Οι καθαρά λογικές ή καθαρά πιθανοτικές μέθοδοι συνήθως αντιμετωπίζουν προβλήματα όταν πρέπει να συμπεραστούν ενδιάμεσα κατηγορήματα ενώ τα MLNs παραμένουν ουσιαστικά ανεπηρέαστα.

Η Naive Bayes έχει καλή απόδοση ως προς AUC για μερικά σύνολα test αλλά κακή απόδοση σε άλλες περιπτώσεις.

Ο CLAUDIEN έχει κακή απόδοση από μόνος του και δεν παρέχει κάποιου είδους βελτίωσης όταν προστίθεται στην KB του MLN.

Η χρήση CLAUDIEN για την δύλιση της KB έχει κακή απόδοση AUC αλλά καλύτερη CLL από ότι αν χρησιμοποιηθεί μόνη της από την αρχή.

Συνολικά, η λογική μέθοδος με την καλύτερη απόδοση είναι η KB+CLB αλλά τα αποτελέσματα της δεν μπορούν να συγκριθούν με αυτά των MLNs.

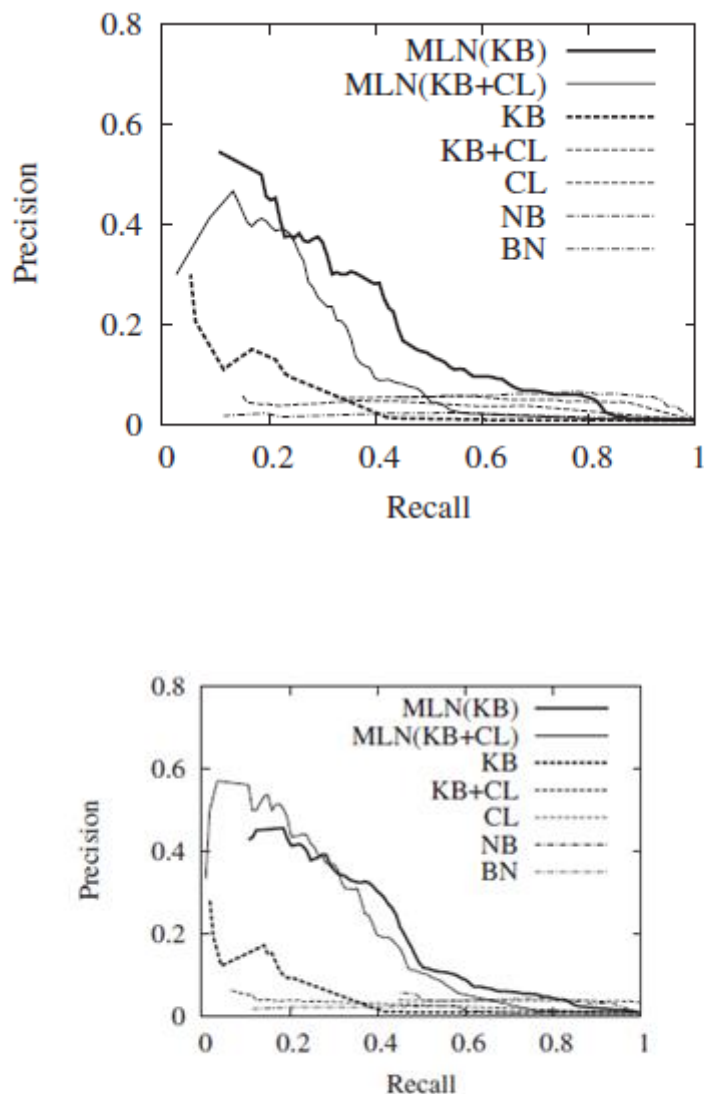
Η γενική πτώση της ακρίβειας περίπου στο 50% της ανάκλησης οφείλεται στο γεγονός ότι η βάση δεδομένων είναι ιδιαίτερος ατελής και επιτρέπει την ταυτοποίηση του μικρότερου μόνο μέρους των σχέσεων Συμβουλευεται Από.

Μια επιθεώρηση δείχνει ότι οι περιστασιακές μικρότερες πτώσεις της ακρίβειας σε πολύ χαμηλές ανακλήσεις οφείλονται σε φοιτητές που αποφοίτησαν ή που άλλαξαν συμβούλους αφότου συνέγραψαν πολλές δημοσιεύσεις μαζί τους.

**Table 3.4** Experimental results for predicting  $\text{AdvisedBy}(x, y)$  when all other predicates are known (All Info) and when  $\text{Student}(x)$  and  $\text{Professor}(x)$  are unknown (Partial Info). CLL is the average conditional log-likelihood, and AUC is the area under the precision-recall curve. The results are averages over all atoms in the five test sets and their standard deviations. (See <http://www.cs.washington.edu/ai/mln> for details on how the standard deviations of the AUCs were computed.)

System	All Info		Partial Info	
	AUC	CLL	AUC	CLL
MLN(KB)	0.215±0.0172	-0.052±0.004	0.224±0.0185	-0.048±0.004
MLN(KB+CL)	0.152±0.0165	-0.058±0.005	0.203±0.0196	-0.045±0.004
MLN(KB+CLB)	0.011±0.0003	-3.905±0.048	0.011±0.0003	-3.958±0.048
MLN(CL)	0.035±0.0008	-2.315±0.030	0.032±0.0009	-2.478±0.030
MLN(CLB)	0.003±0.0000	-0.052±0.005	0.023±0.0003	-0.338±0.002
KB	0.059±0.0081	-0.135±0.005	0.048±0.0058	-0.063±0.004
KB+CL	0.037±0.0012	-0.202±0.008	0.028±0.0012	-0.122±0.006
KB+CLB	0.084±0.0100	-0.056±0.004	0.044±0.0064	-0.051±0.005
CL	0.048±0.0009	-0.434±0.012	0.037±0.0001	-0.836±0.017
CLB	0.003±0.0000	-0.052±0.005	0.010±0.0001	-0.598±0.003
NB	0.054±0.0006	-1.214±0.036	0.044±0.0009	-1.140±0.031
BN	0.015±0.0006	-0.072±0.003	0.015±0.0007	-0.215±0.003

### 3.10 Συμπεράσματα



**Figure 3.2** Precision and recall for all areas: All Info (upper graph) and Partial Info (lower graph).

Η γρήγορη αύξηση στην ποικιλία των SRL προσεγγίσεων και εργασιών οδήγησε στην ανάγκη για ένα ενοποιητικό πλαίσιο εργασίας.

Σε αυτό το κεφάλαιο προτάθηκε η Μαρκοβιανή λογική σαν ένα τέτοιο υποψήφιο ενοποιητικό πλαίσιο.

Η Μαρκοβιανή λογική συνδυάζει λογική πρώτης τάξεως και δίκτυα Markov επιτρέποντας την μορφοποίηση μιας μεγάλης ποικιλίας εργασιών και προσεγγίσεων SRL σε κοινή γλώσσα.

Τα αρχικά πειράματα υλοποίησης Μαρκοβιανής λογικής έδωσαν καλά αποτελέσματα.

Λογισμικό υλοποίησης λογικής Markov, καθώς και αλγόριθμοι μάθησης και συμπερασμού είναι διαθέσιμα στο <http://www.cs.washington.edu/ai/alchemy>.

## 4. Εφαρμογές πιθανοτικών μοντέλων στην διάγνωση και εξέλιξη ασθενειών

Τα Bayesian δίκτυα χρησιμοποιούνται όλο και περισσότερο στην βιοϊατρική και στους διάφορους τομείς της ιατροφαρμακευτικής περίθαλψης (health-care) προκειμένου να βοηθήσουν στην εύρεση διαφόρων τύπων λύσεων για τα προβλήματα που ανακύπτουν.

Εδώ, θα αναφερθούμε σε τέσσερις περιπτώσεις.

### Εφαρμογή 1. Συμπερασμός διάγνωσης (Diagnostic reasoning)

Η θεμελίωση της διάγνωσης για ένα μεμονωμένο ασθενή, ουσιαστικά ισοδυναμεί με την δημιουργία μιας υπόθεσης για την ασθένεια από την οποία πάσχει ο συγκεκριμένος ασθενής, η οποία θα προκύψει από ένα σύνολο έμμεσων παρατηρήσεων από διαγνωστικά tests.

Παρ' όλα αυτά όμως, τα διαγνωστικά tests δεν μπορούν να αποδώσουν την κατάσταση ενός ασθενή με σαφή τρόπο καθότι συνήθως έχουν αληθείς και ψευδείς τιμές που δεν είναι ίσες με 100%.

Προκειμένου να αποφευχθεί μια λάθος διάγνωση θα πρέπει να ληφθεί υπόψη και η αβεβαιότητα που υπεισέρχεται στα αποτελέσματα των tests που γίνονται σε έναν ασθενή κατά την δημιουργία μιας διαγνωστικής υπόθεσης (diagnostic hypothesis).

Τα Bayesian δίκτυα αποτελούν ένα πρόσφορο πλαίσιο εργασίας για αυτού του είδους τον συμπερασμό στον οποίο υπεισέρχεται αβεβαιότητα.

Στο πρόσφατο παρελθόν αναπτύχθηκε ένας σημαντικός αριθμός συστημάτων που βασίζονται σε δίκτυα (network-based systems) και που χρησιμοποιούνται για ιατρικές διαγνώσεις, με τον αριθμό αυτό να αυξάνει συστηματικά στις μέρες μας.

Γνωστά παραδείγματα πρώιμων συστημάτων είναι το Pathfinder και το MUNIN.

Φορμαλιστικά, μια διάγνωση μπορεί να οριστεί ως μια ανάθεση τιμών σε ένα υποσύνολο των τυχαίων μεταβλητών που μας ενδιαφέρουν που είναι τέτοια ώστε

$$\mathcal{D}^* = \underset{\mathcal{D}}{\operatorname{argmax}} \operatorname{Pr}(\mathcal{D}|\mathcal{E})$$

Όπου  $\mathcal{E}$  είναι τα παρατηρούμενα αποδεικτικά τα οποία αποτελούνται από συμπτώματα, σημάδια και αποτελέσματα διαφόρων tests.

Συνεπώς, μια διάγνωση είναι μια μέγιστη a posteriori ανάθεση τιμών (maximum a posteriori assignment - MPA) σε ένα δεδομένο υποσύνολο μεταβλητών.



Η πραγματοποίηση όμως μιας μέγιστης *a posteriori* ανάθεσης τιμών από ένα Bayesian δίκτυο είναι ένα εξαιρετικά δύσκολο έργο από υπολογιστικής πλευράς.

Από την στιγμή επιπλέον που δεν προκύπτουν πολύ συχνά συνδυασμοί ασθενειών ο διαγνωστικός συμπερασμός επικεντρώνεται γενικώς στις μεμονωμένες ασθένειες.

Μια προσέγγιση είναι να υποθεθεί ότι όλες οι ασθένειες είναι αμοιβαία αποκλειόμενες.

Οι διαφορετικές πιθανές ασθένειες τότε λαμβάνονται σαν οι τιμές μιας μεταβλητής ασθένειας (disease variable).

Μια άλλη προσέγγιση είναι να θεωρήσουμε κάθε δυνατή ασθένεια και ως μια ξεχωριστή μεταβλητή.

Ο συμπερασμός τότε ισοδυναμεί με τον υπολογισμό της πιθανοτικής κατανομής για κάθε μεταβλητή ξεχωριστά. Ο συνδυασμός όμως των πιο πιθανών τιμών για αυτές τις ξεχωριστές μεταβλητές ασθενειών δεν χρειάζεται να είναι μια μέγιστη *a posteriori* ανάθεση σε αυτές.

Συχνά, για την βοήθεια των ιατρών στην πολύπλοκη εργασία του διαγνωστικού συμπερασμού, ένα Bayesian δίκτυο είναι εξοπλισμένο με μια μέθοδο επιλογής test (test-selection method) η οποία αποσκοπεί στο να δείξει ποιά tests είναι εκείνα που ελαττώνουν περισσότερο την αβεβαιότητα σχετικά με την ασθένεια που παρουσιάζεται σε έναν συγκεκριμένο ασθενή.

Μια μέθοδος επιλογής test συνήθως χρησιμοποιεί ένα πληροφοριακό-θεωρητικό μέτρο για την εκτίμηση της αβεβαιότητας στην διάγνωση.

Ένα τέτοιο μέτρο ορίζεται επί μιας πιθανοτικής κατανομής για μια μεταβλητή ασθένειας και εκφράζει την αναμενόμενη ποσότητα πληροφορίας που απαιτείται για τον προσδιορισμό της τιμής αυτής της μεταβλητής με βεβαιότητα.

Ένα παράδειγμα μέτρου που χρησιμοποιείται συχνά για αυτόν τον σκοπό είναι η εντροπία Shannon.

Το μέτρο μπορεί να επεκταθεί ώστε να συμπεριλαμβάνει πληροφορίες σχετικά με τα κόστη που εμπλέκονται για την πραγματοποίηση ενός συγκεκριμένου test καθώς και σχετικά με τις παρενέργειες που μπορεί να έχει.

Μιας και είναι υπολογιστικά δύσκολο να δούμε πέρα από το αμέσως επόμενο διαγνωστικό test, η επιλογή test συνήθως γίνεται με έναν ακολουθιακό τρόπο.

Η μέθοδος τότε προτείνει κάποιο test που πρέπει να γίνει και περιμένει από τον χρήστη να εισάγει τα δεδομένα. Αφού λάβει υπόψη της τα αποτελέσματα του test, η μέθοδος προτείνει το επόμενο πείραμα και ούτω καθ' εξής.

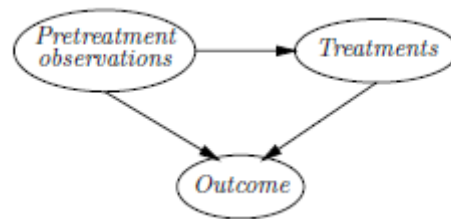
## ***Εφαρμογή 2. Συμπερασμός πρόγνωσης (Prognostic reasoning)***

Ο συμπερασμός πρόγνωσης στην βιοϊατρική και την ιατροφαρμακευτική περίθαλψη συνίσταται στην δημιουργία πρόβλεψης σχετικά με το τι θα συμβεί στο μέλλον.

Καθώς η γνώση του μέλλοντος είναι εγγενώς αβέβαιη, στον προγνωστικό συμπερασμό η αβεβαιότητα είναι ακόμα περισσότερο δεσπόζουσα από ότι είναι στον διαγνωστικό συμπερασμό.

Άλλο ένα διακεκριμένο γνώρισμα του προγνωστικού συμπερασμού σε σύγκριση με τον διαγνωστικό συμπερασμό είναι η εκμετάλλευση της γνώσης σχετικά με την εξέλιξη των διαδικασιών ως προς τον χρόνο.

Ακόμα και αν η προσωρινή γνώση δεν αναπαρίσταται με σαφήνεια, τα προγνωστικά Bayesian δίκτυα εξακολουθούν να έχουν μια καθαρή γενική προσωρινή δομή η οποία αναπαρίσταται στο επόμενο σχήμα.



Το αποτέλεσμα που προβλέπεται για έναν συγκεκριμένο ασθενή, γενικώς επηρεάζεται από την συγκεκριμένη ακολουθία δράσεων θεραπείας που θα πραγματοποιηθούν, που με την σειρά τους μπορεί να εξαρτώνται από τις πληροφορίες που είναι διαθέσιμες για τον ασθενή πριν αρχίσει η αγωγή.

Συχνά, το αποτέλεσμα επηρεάζεται επίσης από την εξέλιξη της ασθένειας καθ' εαυτής.

Φορμαλιστικά, μια πρόγνωση μπορεί να οριστεί σαν μια πιθανοτική κατανομή

$$\Pr(\text{outcome} | \mathcal{E}, \mathcal{F})$$

Όπου  $\mathcal{E}$  είναι και πάλι τα διαθέσιμα για τον ασθενή δεδομένα τα οποία περιλαμβάνουν συμπτώματα, σημάδια και αποτελέσματα από test, ενώ με  $\mathcal{F}$  συμβολίζεται μια επιλεγμένη ακολουθία δράσεων θεραπείας.

Το αποτέλεσμα που μας ενδιαφέρει μπορεί να εκφραστεί με μια μεταβλητή, πχ. μοντελοποίηση του προσδόκιμου ζωής.

Μπορεί όμως, το αποτέλεσμα που μας ενδιαφέρει να είναι πιο πολύπλοκο. Για παράδειγμα η μοντελοποίηση όχι μόνο του προσδόκιμου ζωής αλλά και οι διάφορες άλλες πλευρές που σχετίζονται με την ποιότητα της ζωής.

Για την περίπτωση της μοντελοποίησης ενός τέτοιου σύνθετου αποτελέσματος μπορεί να χρησιμοποιηθεί ένα υποσύνολο μεταβλητών.

Τα προγνωστικά Bayesian δίκτυα είναι μια καινούργια ανάπτυξη στον χώρο της ιατρικής.

Μόνο πρόσφατα ξεκίνησαν οι ερευνητές να αναπτύσσουν τέτοια δίκτυα, για παράδειγμα στον χώρο της ογκολογίας και των μεταδοτικών νοσημάτων.

Υπάρχει μικρή εμπειρία όσον αφορά τον τρόπο ενσωμάτωσης στα Bayesian δίκτυα ιδεών όπως είναι η παραδοσιακή ανάλυση επιβίωσης.



Δεδομένης όμως της σημασίας των προγνώσεων στον τομέα της υγείας είναι σίγουρο ότι στο άμεσο μέλλον θα αναπτυχθούν περισσότερα προγνωστικά δίκτυα.

### **Εφαρμογή 3. Επιλογή θεραπείας**

Ο φορμαλισμός των Bayesian δικτύων προσφέρεται μόνο για τον ορισμό ενός συνόλου τυχαίων μεταβλητών και τον ορισμό μιας από κοινού πιθανοτικής κατανομής επί αυτών.

Κατά συνέπεια, ένα Bayesian δίκτυο επιτρέπει μόνο τον πιθανοτικό συμπερασμό, όπως τον προσδιορισμό μιας διάγνωσης για έναν συγκεκριμένο ασθενή και την πρόβλεψη των επιδράσεων της θεραπείας

Για την λήψη αποφάσεων όμως, όπως για παράδειγμα την επιλογή της πιο κατάλληλης εναλλακτικής θεραπείας για έναν συγκεκριμένο ασθενή, ο φορμαλισμός του δικτύου δεν μπορεί να μας δώσει κάποια απάντηση.

Ο συμπερασμός για εναλλακτικές θεραπείες όμως σχετίζεται με τον συμπερασμό περί των επιδράσεων που αναμένονται από τις διάφορες εναλλακτικές.

Συνεπώς, εμπεριέχει διαγνωστικό συμπερασμό και ακόμα πιο εμφανώς προγνωστικό συμπερασμό.

Προκειμένου να υπάρχει η δυνατότητα επιλογής βέλτιστης θεραπείας, ένα Bayesian δίκτυο και οι σχετιζόμενοι με αυτό αλγόριθμοι συμπερασμού συχνά ενσωματώνονται σε ένα σύστημα υποστήριξης απόφασης (decision-support system) το οποίο παρέχει τις απαραίτητες δομές από την θεωρία λήψης αποφάσεων ώστε να επιλεγεί η βέλτιστη θεραπεία δεδομένων των προβλέψεων.

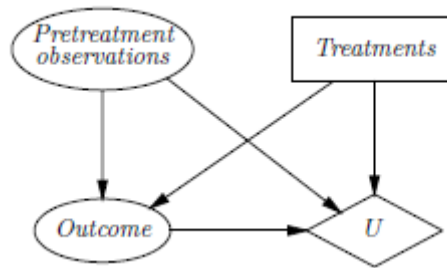
Εναλλακτικά, ο φορμαλισμός του Bayesian δικτύου αναμένουμε να συμπεριλαμβάνει γνώση σχετικά με αποφάσεις και προτιμήσεις.

Ένα παράδειγμα τέτοιου επεκτεταμένου φορμαλισμού είναι ο φορμαλισμός διαγράμματος επιρροής (influence diagram formalism).

Όπως και ένα Bayesian δίκτυο, ένα διάγραμμα επηρεασμού περιλαμβάνει έναν ακυκλικό προσανατολισμένο γράφο.

Σε αυτόν τον γράφο, το σύνολο των κόμβων διαμερίζεται σε ένα σύνολο πιθανοτικών κόμβων που αναπαριστούν τις τυχαίες μεταβλητές, ένα σύνολο κόμβων αποφάσεων που αναπαριστούν τις διάφορες εναλλακτικές θεραπείες και έναν κόμβο τιμής (value node) ο οποίος αναπαριστά τις προτιμήσεις που σχετίζονται.

Τα διαγράμματα επιρροής για την επιλογή θεραπείας έχουν για άλλη μια φορά ξεκάθαρη δομή όπως φαίνεται και στο επόμενο σχήμα.



#### **Εφαρμογή 4. Ανακάλυψη λειτουργικών αλληλεπιδράσεων**

Μέχρι στιγμής εστίασαμε στην κατασκευή και χρήση Bayesian δικτύων για την επίλυση προβλημάτων στην βιοϊατρική και στην περίθαλψη υγείας.

Η διορατικότητα όμως που αποκτάτε από την διαδικασία κατασκευής καθ' εαυτή, ειδικά όταν γίνεται αυτόματα χρησιμοποιώντας κάποια από τις μεθόδους εκμάθησης που περιγράφηκαν στα προηγούμενα, μπορεί να χρησιμοποιηθεί για την επίλυση προβλημάτων.

Καθώς η τοπολογία ενός Bayesian δικτύου μπορεί να ερμηνευθεί ως μια αναπαράσταση των τυχαίων αλληλεπιδράσεων μεταξύ μεταβλητών, υπάρχει ένα αναπτυσσόμενο ενδιαφέρον στον τομέα της βιοπληροφορικής για την χρήση των Bayesian δικτύων για την ανάδειξη των μοριακών μηχανισμών σε κυτταρικό επίπεδο.

Για παράδειγμα, η εύρεση αλληλεπιδράσεων μεταξύ γονιδίων οι οποίες βασίζονται σε εκφράσεις δεδομένων που ανακτήθηκαν πειραματικά σε μικροπεριοχές (microarrays) είναι αυτή την στιγμή ένα σημαντικό ζήτημα έρευνας.

Τα βιολογικά δεδομένα συχνά συλλέγονται με την πάροδο του χρόνου οπότε η ανάλυση των προσωρινών προτύπων (temporal patterns) μπορεί να αναδείξει τον τρόπο με τον οποίο αλληλεπιδρούν οι μεταβλητές συναρτήσει του χρόνου.

Αυτή είναι μια τυπική εργασία που πραγματοποιείται στο χώρο της μοριακής βιολογίας.

Τα Bayesian δίκτυα χρησιμοποιούνται επίσης στις μέρες μας για την ανάλυση τέτοιων χρονοσειρών βιολογικών δεδομένων.

## 5. Μελλοντικές κατευθύνσεις και προοπτικές

Παρ' όλο που ο τομέας των Μαρκοβιανών και Μπεϋζιανών δικτύων στον οποίο αναφερόμαστε εξελίσσεται με ταχύτατους ρυθμούς, εξακολουθούν να υπάρχουν ακόμα πολλές κατευθύνσεις στις οποίες έχουμε ανοικτά προβλήματα. Προφανώς δεν θα ήταν εύκολο να παραθέσουμε μια λεπτομερή λίστα η οποία θα περιελάμβανε όλα αυτά τα ανοικτά προβλήματα. Ούτως ή άλλως, ο εντοπισμός και μόνο ενός τέτοιου προβλήματος είναι πολλές φορές το πρώτο βήμα για την δημιουργία ενός καινούργιου ερευνητικού προγράμματος. Μπορούμε όμως να περιγράψουμε μερικές ευρείες κατηγορίες προβλημάτων στις οποίες είναι ξεκάθαρο ότι υπάρχει πρόσφορο έδαφος για εργασία.

Είναι γεγονός ότι τα πιθανοτικά μοντέλα έχουν χρησιμοποιηθεί ως βασικό εργαλείο για την υλοποίηση πολλών εφαρμογών αυξημένου ενδιαφέροντος όπως για παράδειγμα είναι ο αυτοματοποιημένος συμπερασμός και η λήψη αποφάσεων, η ανάλυση δεδομένων, η αναγνώριση προτύπων καθώς και η ανακάλυψη γνώσης. Για κάποιες από αυτές τις εφαρμογές έχουμε ήδη μιλήσει στα πλαίσια αυτής της εργασίας. Υπάρχουν όμως και άλλες πολλές περιπτώσεις εφαρμογών για τις οποίες δεν έχουμε μιλήσει και στις οποίες εφαρμόζεται η εν λόγω τεχνολογία και ενδεχομένως ακόμα περισσότερες κατηγορίες εφαρμογών στις οποίες δύναται να εφαρμοστεί. Απαιτείται πολύ δουλειά ακόμα για την περαιτέρω ανάπτυξη αυτών των μεθόδων, προκειμένου να μπορούν να εφαρμοστούν αποδοτικά σε μια διαρκώς αυξανόμενη γκάμα προβλημάτων του πραγματικού κόσμου (συμπεριλαμβανομένου προφανώς και του τομέα της υγείας).

Η ικανότητα μας να εφαρμόζουμε γραφικά μοντέλα για την επίλυση προβλημάτων συνήθως περιορίζεται από το γεγονός ότι τις περισσότερες φορές η προσέγγιση γίνεται περισσότερο υπό ένα πρίσμα τέχνης παρά επιστήμης. Υπάρχουν πάντοτε πολλές δυνατότητες επιλογής για την μέθοδο αναπαράστασης, την διαδικασία μάθησης και τον αλγόριθμο συμπερασμού. Δυστυχώς όμως, δεν υπάρχουν και αντίστοιχες κατευθυντήριες γραμμές που να μπορούν να μας βοηθούν στο πως να κάνουμε τις κατάλληλες επιλογές. Η όλη διαδικασία σχεδίασης γίνεται με την λογική πειραματισμού "trial and error" σε συνδυασμό με την ενδεχόμενη εμπειρία που μπορεί να έχει αποκομίσει κατά καιρούς ο εκάστοτε ερευνητής. Θα ήταν εξαιρετικά ευτυχές γεγονός αν μπορούσε να βρεθεί μια συστηματική μέθοδος προσέγγισης.

Είναι επομένως εύλογο να αναρωτηθεί κανείς αν τα γραφικά πιθανοτικά μοντέλα επαρκούν για την αντιμετώπιση των προβλημάτων που συνεχώς προκύπτουν και θα προκύπτουν στο μέλλον. Τα μέχρι στιγμής στοιχεία που διαθέτουμε δείχνουν ότι πράγματι τα πιθανοτικά σχεσιακά μοντέλα είναι πολλά υποσχόμενα και ότι στο μέλλον θα δώσουν απαντήσεις και λύσεις σε πολλά προβλήματα. Δηλαδή είναι πραγματικά άξια προσοχής και περαιτέρω μελέτης.

## Βιβλιογραφία

- [1] R. Braz, E. Amir, and D. Roth. Lifted first-order probabilistic inference. In Proceedings of the International Joint Conference on Artificial Intelligence, 2005.
- [2] W. Buntine. Operations for learning with graphical models. *Journal of Artificial Intelligence Research*, 3:159–225, 1994.
- [3] S. Chakrabarti, B. Dom, and P. Indyk. Enhanced hypertext categorization using hyperlinks. In Proceedings of ACM International Conference on Management of Data, 1998.
- [4] D. Cohn and T. Hofmann. The missing link—a probabilistic model of document content and hypertext connectivity. In Proceedings of Neural Information Processing Systems, 2001.
- [5] M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam, and S. Slattery. Learning to extract symbolic knowledge from the World Wide Web. In Proceedings of the National Conference on Artificial Intelligence, 1998.
- [6] J. Cussens. Loglinear models for first-order probabilistic reasoning. In Proceedings of the Conference on Uncertainty in Artificial Intelligence, 1999.
- [7] L. Dehaspe, H. Toivonen, and R.D. King. Finding frequent substructures in chemical compounds. In International Conference on Knowledge Discovery and Data Mining, 1998.
- [8] T. Dietterich and R. S. Michalski. Inductive learning of structural descriptions: Evaluation criteria and comparative review of selected methods. *Artificial Intelligence*, 16:257–294, 1986.
- [9] S. Dzeroski and N. Lavrac, editors. *Relational Data Mining*. Kluwer, Berlin, 2001.
- [10] P. Finn, S. Muggleton, D. Page, and A. Srinivasan. Discovery of pharmacophores using the inductive logic programming system Progol. *Machine Learning*, 30(1-2):241–270, 1998.
- [11] D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, Upper Saddle River, NJ, 2002.
- [12] N. Friedman, L. Getoor, D. Koller, and A. Pfeffer. Learning probabilistic relational models. In Proceedings of the International Joint Conference on Artificial Intelligence, 1999.
- [13] L. Getoor. *Learning Statistical Models from Relational Data*. PhD thesis, Stanford University, Stanford, CA, 2001.
- [14] L. Getoor and J. Grant. PRL: A probabilistic relational language. *Machine Learning Journal*, 62(1-2):7–31, 2006.
- [15] L. Getoor, E. Segal, B. Taskar, and D. Koller. Probabilistic models of text and link structure for hypertext classification. In Proceedings of the IJCAI Workshop on Text Learning: Beyond Supervision, 2001.
- [16] G. W. Flake, S. Lawrence, and C. L. Giles. Efficient identification of Web communities. In International Conference on Knowledge Discovery and Data

Mining, 2000.

- [17] N. Friedman, L. Getoor, D. Koller, and A. Pfeffer. Learning probabilistic relational models. In Proceedings of the International Joint Conference on Artificial Intelligence, 1999.
- [18] M. R. Genesereth and N. J. Nilsson. Logical Foundations of Artificial Intelligence. Morgan Kaufmann, San Mateo, CA, 1987.
- [19] C. J. Geyer and E. A. Thompson. Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society, Series B*, 54(3):657–699, 1992.
- [20] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London, 1996.
- [21] J. Halpern. An analysis of first-order logics of probability. *Artificial Intelligence*, 46:311–350, 1990.
- [22] D. Heckerman, D. Geiger, and D. M. Chickering. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20:197–243, 1995.
- [23] D. Heckerman, D. M. Chickering, C. Meek, R. Rounthwaite, and C. Kadie. Dependency networks for inference, collaborative filtering, and data visualization. *Journal of Machine Learning Research*, 1:49–75, 2000.
- [24] D. Heckerman, C. Meek, and D. Koller. Probabilistic entity-relationship models, PRMs, and plate models. In Proceedings of the ICML-2004 Workshop on Statistical Relational Learning and Its Connections to Other Fields, 2004.
- [25] G. Hulten and P. Domingos. Mining complex models from arbitrarily large databases in constant time. In *International Conference on Knowledge Discovery and Data Mining*, 2002.
- [26] M. Jaeger. On the complexity of inference about probabilistic relational models. *Artificial Intelligence*, 117:297–308, 2000.