Credal Networks

Graphische Modellierung von unsicherem Wissen

Michael Schomaker

10. März 2005

Inhaltsverzeichnis

Kapitel 1

Bayesianische Netzwerke

1.1 Einführung

Abhängigkeiten verschiedener (Zufalls-) Variablen lassen sich über Bayesianische Netzwerke darstellen (Abb.1.1). Hierbei wird jede Variable durch einen "Knoten" dargestellt. Pfeile symbolisieren eine Abhängigkeitsstruktur zwischen den Variablen. Beispielsweise beeinflußt die Variable Z die Variablen

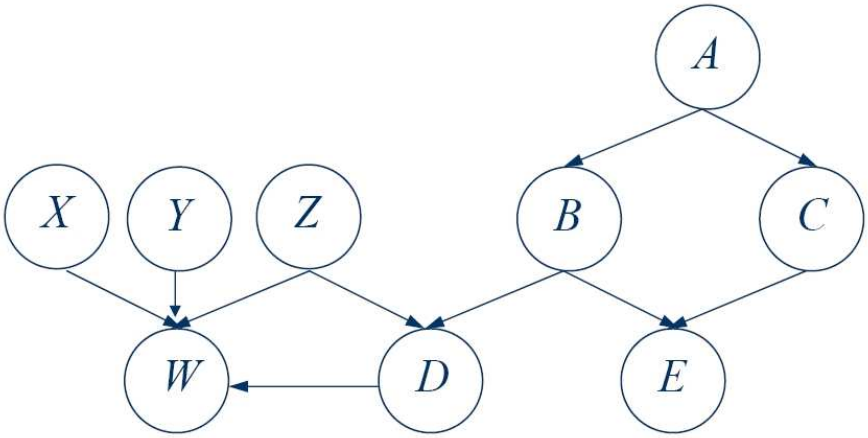

Abbildung 1.1: Beispiel eines bayesiansichen Netzwerkes

W und D, nicht aber umgekehrt.
Weitere wichtige Begriffe sind die der *Eltern* und die der *Nachkommen*. So sind zum Beispiel die Eltern von W die Variablen X,Y,Z. Nachkommen von A sind B,C,E.

1.2 Unabhängigkeitsbeziehungen und d-Seperation

Wie in Abb.(1.2) zu sehen ist, beeinflußt Variable A die Variable B. Variable B beeinflußt Variable C. Sobald B bekannt ist, sind die Variablen A und C unabhängig $\Rightarrow$ *"bedingte Unabhängigkeit"*. Kennen wir den Wert von A in

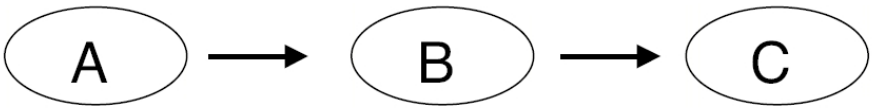

Abbildung 1.2: Einfaches Beispiel eines bayesiansichen Netzwerkes

Abb.(**1.1**), so ist die Verbindung der Variablen B und C blockiert. Das heißt, B und C sind bedingt unabhängig gegeben A.

Das Konzept der *d-Separierung* ist eine graphen-theoretische Eigenschaft von Knoten, die im weiteren helfen wird Inferenzalgorithmen zu vereinfachen.

Definition: Gegeben seien 3 disjunkte Mengen von Knoten X,Y,Z. Der Knoten Z d-separiert X und Y wenn es keinen Pfad gibt von X nach Y für den folgendes gilt:

- jeder Knoten mit konvergierenden Kanten ist in Z (oder hat einen Nachfolger in Z)
- alle anderen Knoten sind außerhalb von Z

Diese Defintion ist ein Formalismus, der hilft bedingte Unabhängigkeitsbeziehungen zu finden.

Ein Beispiel: A "d-separiert" in Abb.(1.1) B und C. E d-separiert *nicht* B und C.

1.3 Inferenz

Bayesianische Netzwerke bestehen aus einem gerichteten azyklischen Graphen (DAG), verschiedenen Variablen, die eine Beziehung zu jedem Knoten haben und einer **bedingten Verteilung** für jede Variable. Die gemeinsame Verteilung für alle Variablen ergibt sich aus:

$$P(X) = \prod_{i=1}^{n} P(X_i|pa(X_i))$$

$pa(X_i)$ bezeichnet hierbei die Eltern (parents) von X_i.
Grundsätzlich taucht das Problem auf, daß Wissen nur über einige Variablen bekannt ist, aber nicht über alle. Man möchte also mit Hilfe der Informationen, die einem über bestimmte Variablen vorliegen die einem unbekannten Parameter schätzen. In der Regel ist dies nichts anderes als die Berechnung der posteriori Wahrscheinlichkeit:

$$P(X|y) = \frac{P(y|X)P(X)}{P(y)}$$

y bezeichen hierbei die bereits beobachtete Variable, X die zu schätzende. Eine genauere Beschreibung dieser Problemstellung wird in Kapitel 2 noch folgen.

Kapitel 2

Credal Networks

2.1 Einführung

Was passiert, wenn ich meinen Variablen nicht mehr einen genauen Wahrscheinlichkeitswert zuordnen kann, sondern nur noch Intervallwahrscheinlichkeiten? Seien beispielsweise die Variablen in Abb. 1.2 binär. Dann könnte folgende Struktur gegeben sein:

$$0.2 \leq P(A) \leq 0.3$$

$$0.1 \leq P(B|A) \leq 0.2 \qquad 0.8 \leq P(B|A^c) \leq 0.9$$

$$0.4 \leq P(C|B) \leq 0.5 \qquad 0.5 \leq P(C|B^c) \leq 0.6$$

Nun stellt sich die Frage, welche Konzepte der ganz normalen bayesianischen Netzwerke man hier übernehmen kann. Wann ist eine Variable von einer anderen unabhängig? Wie konstruiere ich Unabhängigkeitsbeziehungen in einem solchen (credal) Netzwerk (größtes theoretisches Problem)? Wann sind Variablen hier d-separiert? Wie gehe ich hier bei der Inferenz vor? Wie berechne ich bei der Inferenz meine oberen und unteren Intervallgrenzen (größtes praktisches Problem)? Und wie kann ich eventuelle Unsicherheiten bezüglich einer (bedingten) Verteilung modellieren?

2.2 Theoretische Grundlagen und Umgebungsmodelle

2.2.1 Konvexität

Eine Menge $\mathcal{K}$ heißt konvex, wenn für alle Elemente P_i und P_j dieser Menge auch deren "Verbindung" $\alpha P_i + (1-\alpha)P_j$ der Menge angehört (für alle $\alpha \epsilon [0,1]$)

Eine konvexe Menge von Wahrscheinlichkeitsverteilungen nennt man "Credal set". "Credal networks" verbinden "Credal sets" mit einem gerichteten azyklischen Graphen. In Analogie zu den Bayesianischen Netztwerken repräsentiert jeder Knoten eine Variable, und jede Variable ist mit einem *lokalen* "Credal set" $K(X|pa(X))$ verbunden.

2.2.2 Umgebungsmodelle

Im wesentlichen gibt es zwei verschiedene Methoden zur Modellierung von Unsicherheiten bezüglich einer Verteilung. Auf der einen Seite gibt es den Ansatz der (ε, δ)-Umgebungen bei der alle Verteilungen in einer gewissen ε-Umgebung betrachtet werden.
Zum anderen kann man Unsicherheiten mit Hilfe von Verteilungsintervallen modellieren. Hierbei wird eine obere und untere Verteilungsfunktion definiert. Alle Verteilungen bei der jeder Punkt dieser Verteilung in den Intervallgrenzen liegt und den Bedingungen einer Verteilungsfunktion genügt, wird hier als Umgebung betrachtet.
Im folgenden nun zwei Beispiele zur Verdeutlichung des Unterschiedes:

- ε - *Kontaminationsumgebungen (im engeren Sinne):*

  Eine ε - kontaminierte Klasse ist ein "credal set", das von einer Wahrscheinlichkeitsverteilung r(X) und einer reellen Zahl $\varepsilon \epsilon (0,1)$ charakterisiert wird. Diese Klasse beinhaltet alle Wahrscheinlichkeitsverteilungen p(X), mit

  p(X) = (1-ε) r(X) + εq(X) (q(X) beliebige Wahrscheinlichkeit)

- Eine (sehr spezielle) Definition von Verteilungsintervallen ist auf den Fall eindimensionaler Verteilungen beschränkt. Die Verteilungsintervalle haben die Form:

$$\mathrm{U}(F_l; F_u) = \{Q \,\epsilon\, \mathfrak{M} : F_l(x) \leq F_Q(x) \leq F_u(x) \;\text{ für alle } x\}$$

$\mathfrak{M}$ sei hierbei die Menge aller Verteilungen P auf dem $(\mathbb{R}_l, \mathfrak{B}_l)$. F_l bezeichnet die untere, F_u die obere Verteilungsfunktion.

2.2.3 Unabhängigkeitskonzepte

Es gibt verschiedene Unabhängigkeitskonzepte die im Zusammenspiel mit den "Credal Networks" von Interesse sein können. Im folgenden sind zwei davon aufgelistet:

- Epistemische Unabhängigkeit:

  $\underline{E}[f(X)|Y] = \underline{E}[f(X)]$ und $\underline{E}[g(Y)|X] = \underline{E}[g(Y)] \qquad \forall f, g$

- starke Unabhängigkeit:

  jeder Eckpunkt $p(X,Y)\ \epsilon\ K(X,Y)$ genügt den Bedingungen

  $$p(X|Y) = p(X) \text{ und } p(Y|X) = p(Y)$$

  für alle möglichen bedingten Werte.

Wie sich noch herauskristallisieren wird ist besonders das Konzept der starken Unabhängigkeit von Interesse.

2.3 Extension eines Netzwerks

Gegeben sei ein "credal network". Dann existieren verschiedene Mengen an Wahrscheinlichkeiten die den Einschränkungen des Netzwerkes entsprechen. Jede dieser Mengen ist eine *extension*.

Die größte gemeinsame Menge eines Credal Network, bei der jede Variable stark unabhängig von allen anderen Variablen gegeben deren Eltern ist, wird strong extension genannt.

$$\mathcal{M}_x = \{P(X) = \prod_{i=1}^{n} P(X_i|pa(X_i))|\ P\epsilon\mathcal{M}\}$$

Folgendes Beispiel soll dies noch einmal illustrieren:

Zehn Maschinen produzieren Tischdeckenhalter. Der Ausschußanteil bei jeder Maschine liegt im Intervall $I = [0.1, 0.2]$. Wie hoch ist die Wahrscheinlichkeit, daß der Ausschuß bei einer elften Maschine, die von den anderen zehn beliefert wird genau 10 Tischdeckenhalter ist?

$$P(X = 10) = \binom{n}{10} p^{10}(1-p)^{n-10} = \binom{n}{10}(p_1 \cdot p_2 \cdot ... \cdot p_{10}) \cdot (1-p_{11}) \cdot ... \cdot (1-p_n)$$

$p_1 = p_2 = ... = p_{10} \, \epsilon \, [0.1, 0.2] \Rightarrow$ strong extension möglich!

Die größte Menge von Verteilungen die (nur) den Bedingungen des Netzwerks genügt, wird natural extension genannt.

$$\mathcal{M}_x = \{P(X) = \prod_{i=1}^{n} P_i(X_i|pa(X_i))| \; P_i \epsilon \mathcal{M}\}$$

Eine natural extension genügt also jeder Einschränkung im Netzwerk, aber keiner anderen. Wäre in obigem Beispiel etwa $p_1, p_2, ..., p_{10} \, \epsilon \, [0.1, 0.2]$, so wäre eine natural extension erforderlich.

Natural extensions können daher mit unsicherem Wissen weit flexibler umgehen, der Aufwand im Zuge der Inferenz ist jedoch bei weitem höher als bei der Arbeit mit strong extensions.

Bei einem gegebenen Credal network entspricht jede graphische d-seperations Beziehung einer bedingten Unabhängigkeitsbeziehung in der strong extension des Netzwerks!

Dies bedeutet natürlich implizit, daß weite Teile der Bayesianischen Netzwerktheorie hier übernommen werden können. Dies minimiert den Rechenaufwand enorm. Daher existieren auch schon einige Algorithmen zur Lösung von Problemen in Credal Networks, wenn eine strong extension angenommen werden kann. Algorithmen zur Berechnung von Intervallgrenzen bei natural extensions stecken noch in der Entwicklung.

2.4 Inferenz

2.4.1 Prinzip der Inferenz

Inferenz bei Credal Networks ist ein Optimierungsproblem. Ziel ist, eine Verteilung $P(X_i|pa(X_i))$ in $K(X_i|pa(X_i))$ für jede Variable zu finden, die den

Wahrscheinlichkeitswert für eine gesuchte Variable X_Q maximiert:

$$\bar{P}(X_Q = x_Q|E) = \max\left(\frac{P(X_Q = x_Q), E}{P(E)}\right)$$

Die Maximierung unterliegt linearen Einschränkungen wie sie uns durch die Annahmen des Netzwerkes gegeben sind. Doch wie löst man dieses Problem unter welchen Voraussetzungen? Hierzu gibt es verschiedene Ansätze. Die bisher besten Algorithmen die im Zuge von strong extensions entwickelt wurden beruhen auf dem "'Branch and Bound"-Prinzip.

2.4.2 Branch and Bound

Branching: Dieser Schritt dient dazu, das Problem in mehrere Teilprobleme aufzuteilen, die eine Vereinfachung des ursprünglichen Problems darstellen. Rekursives Ausführen läßt eine Baumstruktur entstehen. Es existieren verschiedene Auswahlverfahren für die Wahl des als nächsten zu bearbeitenden Knotens:

- Tiefensuche: Von den noch nicht bearbeiteten Teilproblemen wird das gewählt, welches als letztes eingefügt wurde. Mit dieser Auswahlregel arbeitet sich das verfahren im Baum möglichst schnell in die Tiefe, so daß sehr schnell eine zulässige Lösung erlangt wird, über deren Qualität jedoch nichts ausgesagt wird

- Breitensuche: Von den noch nicht bearbeiteten Teilproblemen wird das gewählt, welches als erstes in den Baum eingefügt wurde. Bei Verwendung dieser Auswahlregel werden die Knoten im Baum pro Ebene abgearbeitet, bevor tiefer gelegene Knoten betrachtet werden. Eine zulässige Lösung wird in der Regel erst realtiv spät erhalten, hat aber tendenziell einen guten Lösungswert.

- Bestensuche: Von den noch nicht bearbeiteten Teilproblemen wird dasjenige gewählt, welches die beste untere Schranke vorweist. Durch diese qualitativ arbeitende Auswahlregel wird versucht, direkt einen sehr guten Lösungswert als erste Lösung zu erhalten

Bounding: Dieser Schritt hat die Aufgabe, bestimmte Zweige des Baumes abzuschneiden, d.h. in der weiteren Berechnung nicht mehr zu betrachten, um so den Rechenaufwand zu begrenzen. Der Weg von der Wurzel des Baumes zu einem Teilproblem bildet die *untere Schranke* dieses Teilproblems Als *obere Schranke* dient eine vorerst optimale zulässige Lösung. Diese erhält man durch Berechnung eines kompletten Zweiges. Übersteigt die untere Schranke

in einem Teilproblem die obere Schranke, so wird dieser "Teilbaum" abgeschnitten. Wird eine Lösung gefunden, die besser als die bisherige obere Schranke ist, so ist eine neue optimale Lösung erreicht.

2.5 JavaBayes

Mit Hilfe des Programms JavaBayes finden die "Credal Networks" ihre Anwendung. Es können eigene Netzwerke mit all ihren Eigenschaften kreiert werden. Umgebungsmodelle werden hier ebenso berücksichtigt, wie die Eigenschaften eines Netzwerks im Hinblick auf Intervallwahrscheinlichkeiten. Unter `http://www-2.cs.cmu.edu/~javabayes/Home/node2.html` kann JavaBayes getestet werden.