

# Métodos y algoritmos determinísticos de inferencia aproximada en Redes Bayesianas

Mariano Ferrero

Universidad Politécnica de Madrid, España

## Abstract

Existen dos grandes formas de tratar con inferencia en una Red Bayesiana: exacta o aproximada. En este artículo se pretende realizar una revisión de técnicas y algoritmos de este último tipo propuestas hasta la actualidad, particularmente aquellas consideradas determinísticas, las cuáles buscan reducir de alguna forma los cálculos requeridos para llevar a cabo esa tarea introduciendo simplificaciones en el modelo original.

KEY WORDS: Redes Bayesianas; Inferencia aproximada; Algoritmos determinísticos.

## 1 Introducción

Una Red Bayesiana (RB) es un modelo gráfico probabilístico representado mediante un grafo dirigido acíclico que describe la factorización de una distribución de probabilidad conjunta determinada, donde la estructura de dicho grafo exhibe las variables junto con las relaciones (dependencias e independencias) existentes entre ellas.

Las RB presentan una gran cantidad de ventajas, dentro de las cuáles se pueden mencionar la posibilidad de representar y trabajar en contextos de incertidumbre. Una vez que el modelo ha sido desarrollado, es posible efectuarle preguntas para obtener, en base a nuevas evidencias, la probabilidad a posteriori de todas o un conjunto de variables. Esto último se conoce como propagación de la evidencia o inferencia, y básicamente se traduce en la realización de una serie de cálculos probabilísticos.

De acuerdo a lo mencionado por Cooper (1990), el proceso de inferencia en una RB es un problema NP-completo en redes medianas o grandes. De acuerdo a esto, se han propuesto diversas alternativas para abordar dicha tarea, las cuáles se enmarcan en dos grandes categorías: exactas o aproximadas. Para esta última categoría, los algoritmos propuestos se pueden clasificar a su vez en determinísticos o estocásticos.

Los algoritmos estocásticos se basan en realizar un número determinado de simulaciones a partir de una distribución de probabilidad existente en la red y estimar la probabilidad de interés a partir de ello. Un ejemplo de esto es el muestreo lógico (Henrion (1988)). Por otra parte, los algoritmos determinísticos buscan simplificar el problema de alguna forma, como por ejemplo mediante la eliminación de determinados arcos presentes en el grafo.

En este trabajo se pretende realizar una revisión de algunas de las técnicas y algoritmos determinísticos desarrollados hasta la actualidad para tratar con inferencia aproximada mediante la introducción de simplificaciones en el modelo original.

---

Se confiere permiso para hacer copias de todo o parte de este trabajo para uso personal o académico sin cargo alguno, siempre y cuando dichas copias no sean efectuadas o distribuidas para obtener un beneficio o ventaja comercial. Dichas copias deberán referenciar siempre el trabajo original. Hacer público este documento sin las consideraciones antepuestas requiere previa autorización.

## 2 Estado del arte

Las diversas formas que se han introducido a lo largo del tiempo para intentar simplificar el problema se pueden clasificar de acuerdo a la estrategia que persiguen: Eliminación de arcos, aproximación de tablas de probabilidad condicionada, representaciones alternativas y eliminación de nodos. A continuación se introducen los trabajos enmarcados dentro de cada una de dichas estrategias.

### 2.1 Eliminación de arcos

Una de las estrategias abordadas en mayor amplitud ha sido la de eliminación de arcos. Una de las ideas elaboradas consiste en eliminar aquellos arcos que codifican relaciones de alguna forma detectadas como "débiles". Uno de los primeros trabajos que introdujo esta idea fue el de Boerlage (1992), el cual se limitaba a variables binarias. Luego, Nicholson y Jitnah (1998) propondrían un enfoque similar utilizando el concepto de información mutua para determinar el peso de cada arco.

Por otra parte, aunque de forma similar, van Engelen (1997) propone la detección de aquellos arcos candidatos a ser eliminados utilizando como medida la divergencia de Kullback-Leibler (1959) entre la distribución original y la aproximada. Este puede verse como una extensión de lo propuesto por Kjaerulff (1994), el cual utilizaba el mismo enfoque sobre grafos moralizados.

Choi, Chan y Darwiche (2005), por su parte, han propuesto la eliminación de aquellos arcos teniendo en cuenta la evidencia proporcionada y utilizando el concepto de entropía. Vale la pena destacar que en este caso, a diferencia de los mencionados anteriormente, la simplificación se aplica específicamente para cada consulta debido a que se utiliza la evidencia proporcionada.

### 2.2 Aproximación de tablas de probabilidad condicionada

La cardinalidad de las variables encontradas en una RB tiene una relación directa con la complejidad demandada en los cálculos de inferencia sobre la misma, estableciendo una relación directamente proporcional entre el número de estados de las variables y el tiempo requerido para realizar una evaluación sobre la red. Debido a esto, una de las formas de simplificar el proceso de inferencia es intentar reducir de alguna forma ese factor. Liu y Wellman (2002) proponen un algoritmo que comienza desde un nivel de abstracción máximo (todos los estados de cada variable se combinan en un único estado) y luego va quitando abstracciones hasta que no sea posible realizar el cálculo sobre la red, o bien, todos los estados originales hayan sido evaluados (lo cuál sería un caso de inferencia exacta).

Un trabajo similar a este es el propuesto por Jensen y Anderson (1990), donde la simplificación se basa en reducir a cero aquellas probabilidades que no superan un umbral determinado.

### 2.3 Representaciones alternativas

Otro de los enfoques adoptados ha sido la utilización de representaciones alternativas para las distribuciones de probabilidad condicionadas presentes en el modelo original. Poole (1998) ha utilizado reglas para codificarlas y realizar simplificaciones a partir de las mismas.

Por otra parte, también se han utilizado como representaciones alternativas estructuras basadas en árboles (Sarkar (1993)), tales como árboles de probabilidad (Cano *et al.* (2000), Cano *et al.* (2011)). Continuando sobre esta misma línea, recientemente Pérez-Ariza (2013) ha propuesto como una nueva forma de representación los árboles de probabilidad iterativos, los cuáles se presentan como una generalización de los mencionados previamente.

### 2.4 Eliminación de nodos

Finalmente, otra categoría dentro de la cual se pueden enmarcar las técnicas propuestas trata con la selección de un conjunto de nodos que serán tenidos en cuenta para realizar los cálculos necesarios (descartando de alguna forma los demás). Draper y Hanks (1994) han desarrollado un algoritmo

llamado "Localized Partial Evaluation" (LPE), el cuál incrementa iterativamente el número de nodos a tener en cuenta, descartando aquellos que se encuentran más alejados del nodo de interés. A su vez, el método propuesto por Poole (1998) de representación alternativa puede verse como un caso especial de eliminación de nodos, debido al modo en el que realiza las simplificaciones sobre esas estructuras.

Un caso particular dentro de esta categoría es el método propuesto por Horvitz *et al.* (1989), denominado *Bounded conditioning*, el cuál propone la simplificación de los cálculos requeridos mediante la instanciación de determinados nodos, lo cuál genera nuevas estructuras llamadas *cutsets*. A diferencia de otros métodos exactos que computan cada una de estas nuevas estructuras en particular, en este caso se realizan los cálculos de los *cutsets* de manera progresiva, de acuerdo a los recursos disponibles.

### 3 Conclusiones y trabajo futuro

Como se introdujo anteriormente, las RB son modelos potentes que permiten razonar y tomar decisiones en contextos de incertidumbre. Debido a que en determinadas situaciones realizar cálculos sobre dichos modelos puede volverse inmanejable, las distintas técnicas y algoritmos que buscan simplificar esta tarea han sido exploradas en las últimas décadas por la comunidad científica.

La naturaleza de estos modelos ha permitido abordar esta problemática desde distintos enfoques, lo que ha dado lugar a la clasificación introducida anteriormente. Debido a las particularidades de cada uno de ellos, no es posible especificar un método eficiente que pueda utilizarse en todos los problemas.

Vale la pena destacar que, en base a las fechas de publicación de los artículos obtenidos, pareciera existir una disminución en los últimos años con respecto al interés sobre esta temática en particular. No obstante, es posible imaginar posibles líneas de investigación emergentes en los próximos años relacionadas con la utilización de estructuras alternativas o adaptaciones de los algoritmos utilizados para operar con ellas.

### Referencias

- Boerlage, B. (1992). *Link strength in Bayesian Networks*, Master Thesis, University of British Columbia, 1992, 109 páginas.
- Cano, A., Moral, S. y Salmerón, A. (2000). Penniless propagation in join trees. *International Journal of Intelligent Systems*, 15, pp. 1027-1059.
- Cano, A., Gómez-Olmedo, M. y Moral, S. (2011). Approximate inference in Bayesian networks using binary probability trees. *International Journal of Approximate Reasoning*, 52, pp. 49-62.
- Choi, A., Chan, H. y Darwiche, A. (2005). On Bayesian network approximation by edge deletion. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence (UAI)*, AUAI Press, Arlington, Virginia, pp. 128-135.
- Cooper, G. F. (1990). The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42, 393-405.
- Draper, D. y Hanks, S. (1994). Localized partial evaluation of belief networks. In *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, pp. 170-177.
- Henrion, M. (1988). Propagating uncertainty in Bayesian Networks by probabilistic logic sampling. In *Uncertainty in Artificial Intelligence*, Elsevier Science, pp. 149-163.
- Horvitz, E., Suermondt, H. J. y Cooper, G. F. (1989). Bounded conditioning: Flexible Inference for Decisions Under Scarce Resources. In *Proceedings of the 5th Conference on Uncertainty in Artificial Intelligence*, Windsor, Ontario, USA, pp. 182-193.

- Jensen, F. y Andersen, S. K. (1990). Approximations in Bayesian belief universes for knowledge based systems. In *Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence*, Cambridge, MA, USA, pp. 162-169.
- Kjaerulff, U. (1994). Reduction of Computational Complexity in Bayesian Networks Through Removal of Weak Dependencies. In *Proceedings of the 10th Conference Uncertainty in Artificial Intelligence*, pp. 374-382.
- Kullback, S. (1959). *Information Theory and Statistics*, New York: John Wiley, 1959.
- Liu, C. y Wellman, M. P. (2002). Evaluation of Bayesian networks with flexible state-space abstraction methods. *International Journal of Approximate Reasoning*, 30, 1-39.
- Nicholson, A. E. y Jitnah, N. (1998). Using Mutual Information to determine Relevance in Bayesian Networks. In *Proceedings of the 5th Pacific RIM International Conference on Artificial Intelligence*, Springer, pp. 399-410.
- Pérez-Ariza, C. B. (2013). *New data structures and algorithms for uncertainty treatment with Probabilistic Graphical Models*, PhD Thesis, Universidad de Granada, 2013, 270 páginas.
- Poole, D. (1998). Context-specific approximation in probabilistic inference. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, Madison, WI, USA, pp. 447-454.
- Sarkar, S. (1993). Using Tree-Decomposable Structures to Approximate Belief Networks. In *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, Washington, DC, USA, Morgan Kaufmann, pp. 376-382.
- van Engelen, R. A. (1997). Approximating Bayesian belief networks by arc removal. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19, 916-920.