Inferencia en Redes Bayesianas Híbridas

Andrés Herranz González
Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid, Spain

Resumen

Las Redes Bayesianas híbridas han recibido una gran atención en los últimos años. La diferencia con respecto a las Redes Bayesianas estándar es que pueden contener variables discretas y continuas simultáneamente, lo cual se acerca más a casos reales. La inferencia es la acción de calcular la probabilidad de cada estado de un nodo en una Red Bayesiana cuando se conocen los valores que toman otras variables de la red. En el caso de las Redes Bayesianas híbridas, se trata de un proceso aún más complejo debido a las variables continuas que lleva, en la mayoría de los casos, a la necesidad de utilizar métodos aproximados.

Keywords: Probabilidad; Inferencia; Híbrido; Redes Bayesianas; CLG; Discretizar

1. Introducción

Las Redes Bayesianas o RB (Pearl, 1988) son un tipo concreto de grafos probabilísticos. Se trata de grafos dirigidos acíclicos (DAG) los cuales nos permiten, entre otras cosas, inferir probabilidades condicionadas. Las Redes Bayesianas son una herramienta con gran cantidad de aplicaciones en situaciones difíciles de estudiar (Pourret Naim & Marcot, 2008). De entre ellas, la inferencia es una de las más comunes y consiste en calcular la distribución marginal a posteriori de una variable, dado el valor de otro conjunto de variables del modelo.

En la práctica, al modelizar este tipo de grafos, nos encontramos simultáneamente con variables tanto discretas como continuas. A este tipo de redes se denominan Redes Bayesianas híbridas y su estudio y modelado implican una mayor complejidad (Murphy, 1998; Koller & Friedman, 2009).

2. Estado del arte

Como ya se ha adelantado, la inferencia consiste en calcular el valor de una variable dado un conjunto de observaciones. De esta manera, dado un conjunto de observaciones $\mathbf{X}_E \subset \mathbf{X}$ y el conjunto de variables de interés $\mathbf{X}_I \subseteq \mathbf{X} \setminus \mathbf{X}_E$ la inferencia se calcula cómo $p(x_i|\mathbf{x}_E)$ para cada $i \in I$.

A lo largo de los años se han utilizado gran cantidad de métodos para la inferencia en Redes Bayesianas, que han servido a su vez como base para el desarrollo de métodos para de inferencia en Redes Bayesianas híbridas. Estos métodos se dividen en: inferencia basada en el esquema de paso de mensajes y métodos basados en Monte Carlo (Salmerón *et al.*, 2018).

Los algoritmos basados en el paso de mensajes se dividen en: por un lado, los que organizan los cálculos en estructuras auxiliares más pequeñas llamadas *join tree* o *junction tree* (Shenoy & Shafer, 1990). El algoritmo transforma la RB original en un árbol equivalente formado por cliques, y ejecuta el paso de mensajes en esa estructura. Por otro lado, los algoritmos basados en *loopy belief propagation* (Pearl, 1988; Peot & Shachter, 1991) que envían los mensajes por los bordes de la RB, de manera que el coste de inferencia es exponencial en el número de nodos de *familia* más larga. Por último, la inferencia variacional (Jordan *et al.*, 1999; Attias, 2000) busca iterativamente aproximar la distribución de interés a posteriori.

Los métodos basados en Monte Carlo calculan de forma aproximada la probabilidad $p(x_i|\mathbf{x}_E)$ a partir de las muestras generadas aleatoriamente. Por un lado, el método importance sampling (Hammersley

& Handscomb, 1964) consiste en muestrear cada variable utilizando una distribución condicionada a las observaciones $\mathbf{x}_E$. Por otro, los métodos basados en Cadenas de Markov Monte Carlo, generan las Cadenas de Markov de tal manera que $p(\mathbf{x}|\mathbf{x}_E)$ sea una distribución estacionaria. Este último método tiene dos aproximaciones bien conocidas: el Gibbs sampler (Hrycej, 1990) y el algoritmo de Metropolis-Hastings (Hastings, 1970).

Sin embargo, al trabajar en dominios híbridos nos encontramos con el problema de integrar las variables continuas que no pertenecen a $\mathbf{X}_E$. Para solventar este problema, se van a revisar los algoritmos y métodos propuestos en la literatura.

2.1. Algoritmos que asumen Gaussianidad

Una de las ideas más tempranas se basó en el modelo condicional de Gauss (CG) (Lauritzen, 1992), que asume que la distribución condicional de las variables continuas dadas las discretas, es una función Gaussiana multivariante. Un caso particular, es el modelo Gaussiano lineal condicional (CLG) (Lauritzen & Wermuth, 1989), donde la media de la distribución condicional de cada variable continua es una función lineal de sus variables primarias continuas en la red. Este modelo incorpora unas restricciones tales que: una variable discreta no puede tener como padres variables continuas, y cada variable continua debe seguir una distribución Gaussiana lineal condicional.

Uno de los primeros métodos exactos para la inferencia en RB híbridas fue propuesto por Lauritzen (1992) y posteriormente mejorado por Lauritzen & Jensen (2001) el cual era capaz de inferir sobre una RB siempre y cuando la distribución conjunta fuese CLG. Posteriormente Madsen (2008) propuso otro algoritmo para inferencia exacta en redes CLG que factoriza los potenciales tanto en los *cliques* como en los separadores de los *join tree*. Finalmente se ha propuesto una mejora del algoritmo que aprovecha el conocimiento semántico inducido por la estructura del grafo y las evidencias (Zhu *et al.*, 2012).

Para superar las restricciones estructurales indicadas más arriba, Lerner *et al.* (2001) introdujeron las denominadas redes CLG aumentadas. En ellas, la función de densidad condicionada de las variables discretas con padres continuos se modela utilizando la función *soft-max*. El inconveniente de este método es que obtiene una inferencia aproximada, no exacta. Heskes & Zoeter (2003) propusieron utilizar general *belief propagation* en redes CLG aumentadas. El método se basaba en utilizar *loopy belief propagation* guardando únicamente los dos primeros momentos al marginalizar, en vez de almacenar la función de densidad completa.

Para mejorar la escalabilidad, Sun *et al.* (2010) propusieron un algoritmo que obtiene resultados exactos en redes CLG que tengan una estructura de *poly-tree* y que podía extenderse a casos generales de manera aproximada utilizando mixturas de Gaussianas. Recientemente Salmerón *et al.* (2015) propusieron un algoritmo basado en *importance sampling* y la adaptación del algoritmo de ponderación de evidencias (Fung & Chang, 1990) a las redes CLG.

2.2. Algoritmos que asumen familia exponencial

En este apartado se presentan modelos de la familia exponencial conjugada, en los cuales, las distribuciones condicionadas de las variables continuas pertenecen a la familia exponencial. La inferencia en RB híbridas se puede llevar a cabo de una manera eficiente cuando las distribuciones son exponenciales conjugadas (Beal, 2003).

Para este tipo de esquemas, Winn & Bishop (2005) propusieron un algoritmo de paso de mensajes variable (VMP) que, al igual que *belief propagation*, propaga los mensajes a través de los nodos de la red actualizando las convicciones a posteriori. El modelo es generalizable a cualquier tipo de red con modelos conjugados exponenciales. Basado en esta aproximación de inferencia variacional VMP, Masegosa *et al.* (2017) propusieron un nuevo algoritmo más escalable, siguiendo el diseño típico de los procesos big data. Este método propone utilizar VMP como un algoritmo de ascenso de gradiente natural proyectado, que proporciona una base teórica y práctica para la paralelización de métodos variacionales sobre modelos generales de la familia exponencial conjugada.

2.3. Familias de distribución general

La idea generalizada para inferir en RB con familias de distribuciones generales es convertir el modelo en otro similar sobre el cual sea más sencillo y eficiente la inferencia exacta o aproximada.

Uno de los modelos más comunes de transformación es la discretización de las variables continuas. En el caso de las variables continuas que no tienen padres, la discretización se realiza aproximando la función de densidad con una función a trozos. En el caso de las variables continuas X_I con padres continuos $\mathbf{pa}(X_i) \subseteq \mathbf{X}_C$, el dominio de los padres $\Omega_{pa(X_I)}$ se divide en un conjunto de regiones disjuntas $P_1 ... P_{r_i}$. A continuación, para cada región P_ℓ se aproxima la probabilidad condicionada mediante la función q_ℓ, de tal manera que $p(x_i|\mathbf{pa}(x_i)) \approx q_\ell$. Finalmente, se discretizan $q_\ell(x_i)$ como si fuera distribuciones incondicionales.

Este método fue propuesto por Kozlov & Koller (1997). En su propuesta original, la discretización se llevaba a cabo minimizando la divergencia de Kullback-Leibler entre la distribución real y la aproximada. Posteriormente, Neil Tailor & Marquez (2007) propusieron un algoritmo más eficiente que discretizaba las densidades individuales a posteriori en lugar de discretizar dinámicamente las densidades del árbol.

Otros métodos para realizar la traslación de un modelo a otro son: la mixtura de exponenciales truncadas (MTE) y la mixtura de polinómicas (MoPs). En el caso de la mixtura de exponenciales truncadas (Moral Rumí & Salmerón, 2001), la función de densidad se aproxima como la suma de funciones exponenciales truncadas en lugar de una constante. En el caso de la mixtura de polinómicas (Shenoy & West, 2011) la función de densidad se calcula como la suma de funciones polinómicas, de manera que es más fácil integrar para calcular las marginales. Posteriormente, combinando ambos métodos, Langseth Nielsen Rumí & Salmerón (1997) propusieron la mixtura de funciones base truncadas (MoTBFs). En este caso, q_ℓ se aproxima como la suma de funciones base, lo cual asegura que la inferencia se puede calcular de manera eficiente usando la arquitectura de Shenoy & Shafer (1990). Esta técnica fue recientemente aplicada por Salmerón & Reche (2018).

En otro método reciente, se introduce la representación de las distribuciones de probabilidad en un árbol de decisión estructurado para llevar a cabo la inferencia (Mori & Mahalec, 2016). Esto permite utilizar el algoritmo *belief propagation* para casos sencillos y *loopy belief propagation* para el resto. En este algoritmo, los mensajes parten de los nodos de las evidencias y así se omite el cálculo de estos mensajes.

Posteriormente, (Cortijo & Gonzales, 2017) propusieron un algoritmo que se basaba en discretizar las variables continuas y generar otras variables continuas artificiales para cada una de ellas que eran modeladas usando funciones de densidad truncadas condicionadas. Así, estas variables artificiales no tienen hijos y la inferencia se puede llevar a cabo de manera eficiente.

Para poder inferir en RB híbridas con distribuciones condicionales determinísticas no lineales, Cobb & Shenoy (2005) propusieron un método basado en aproximar estas distribuciones utilizando para ello funciones lineales a trozos. Posteriormente, Cobb & Shenoy (2017) propusieron una mejorar para estimar las funciones lineales a trozos basado en el uso de la heurística *error medio cuadrático* (EMC).

3. Conclusiones y trabajo futuro

La inferencia en Redes Bayesianas híbridas ha sido ampliamente estudiada. En general, la inferencia exacta es posible con modelos CLG. Cuando estos modelos no son aplicables, es necesario utilizar inferencia aproximada. Para estos casos, lo más común es utilizar funciones de traslación como la discretización o la mixtura de funciones base truncadas (MoTBFs).

Las líneas de trabajo más importante son: por un lado, la escalabilidad, dado que a la mayoría de los algoritmos propuestos les cuesta inferir de forma eficiente con RB grandes y, por otro, una mayor generalización en la inferencia exacta, dado que las asunciones de CLG son inviables en gran cantidad de modelos reales.

Referencias

Attias, H. (2000). A variational Bayesian framework for graphical models. *Advances in neural information processing systems*, 13, 209–215.

Beal, M. J. (2003). Variational algorithms for approximate Bayesian inference. Ph.D. thesis, *Gatsby Computational Neuroscience Unit*, University College London.

Cobb, B.R., Shenoy, P.P. (2012). Nonlinear deterministic relationships in Bayesian networks. *Lecture Notes in Computer Science*, 3571 LNAI, pp. 27-38.

Cobb, B.R., Shenoy, P.P. (2017). Inference in Hybrid Bayesian Networks with Nonlinear Deterministic Conditionals. *International Journal of Intelligent Systems*, 32, 1217-1246.

Cortijo, S., & Gonzales, C. (2017). On conditional truncated densities Bayesian networks. *International Journal of Approximate Reasoning*, In Press.

Fung, R., & Chang, K. (1990). Weighting and integrating evidence for stochastic simulation in Bayesian networks. In Henrion, M., Shachter, R., Kanal, L., & Lemmer, J. (Eds.), *Uncertainty in Artificial Intelligence*, Vol. 5, pp. 209–220. North-Holland (Amsterdam).

Hammersley, J., & Handscomb, D. (1964). Monte Carlo Methods. *Chapman & Hall*.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 97–109.

Heskes, T., & Zoeter, O. (2003). Generalized belief propagation for approximate inference in hybrid Bayesian networks. In *Proceedings of the 9th Workshop on A.I. and Statistics*.

Hrycej, T. (1990). Gibbs sampling in Bayesian networks (research note). *Artificial Intelligence*, 46, 351–363.

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., & Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning*, 37, 183–233.

Koller, D., & Friedman, N. (2009). Probabilistic graphical models: Principles and techniques. *MIT Press*

Kozlov, A. V., & Koller, D. (1997). Nonuniform dynamic discretization in hybrid networks. In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, pp. 314–325.

Langseth, H., Nielsen, T. D., Rumí, R., & Salmerón, A. (2012). Mixtures of truncated basis functions. *International Journal of Approximate Reasoning*, 53(2), 212–227.

Lauritzen, S. L. (1992). Propagation of probabilities, means and variances in mixed graphical association models. *Journal of the American Statistical Association*, 87(420), 1098– 1108.

Lauritzen, S. L., & Jensen, F. (2001). Stable local computation with conditional Gaussian distributions. *Statistics and Computing*, 11(2), 191–203.

Lauritzen, S., & Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *The Annals of Statistics*, 17, 31–57.

Lerner, U., Segal, E., & Koller, D. (2001). Exact inference in networks with discrete children of continuous parents. In *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, pp. 319–328.

Madsen, A. (2008). Belief update in CLG Bayesian networks with lazy propagation. *International Journal of Approximate Reasoning*, 49, 503–521.

Masegosa, A. R., Martínez, A. M., Langseth, H., Nielsen, T. D., Salmerón, A., Ramos- López, D., & Madsen, A. L. (2017). Scaling up Bayesian variational inference using distributed computing clusters. *International Journal of Approximate Reasoning*, 88, 435–451.

Moral, S., Rumí, R., & Salmerón, A. (2001). Mixtures of truncated exponentials in hybrid Bayesian networks. In EQSCARU'2001, Vol. 2143 of *Lecture Notes in Artificial Intelligence*, pp. 145–167. Springer, Berlin, Germany.

Mori, J., & Mahalec, V. (2016). Inference in hybrid Bayesian networks with large discrete and continuous domains. *Expert Systems with Applications*, 49, 1–19.

Murphy, K. P. (1998). Inference and learning in hybrid Bayesian networks. Tech. rep. UCB/CSD-98-990, EECS Department, University of California, Berkeley.

Neil, M., Tailor, M., & Marquez, D. (2007). Inference in Bayesian networks using dynamic discretisation. *Statistics and Computing*, 17(3), 219–233.

Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. *Morgan Kaufmann Publishers Inc.*, San Mateo, CA.

Peot, M. A., & Shachter, R. D. (1991). Fusion and propagation with multiple observations in belief networks. *Artificial Intelligence*, 48(3), 299–318.

Pourret, O., Naim, P., & Marcot, B. (2008). Bayesian networks. A practical guide to applications. *Statistics in Practice*. Wiley.

Salmerón, A., Ramos-López, D., Borchani, H., Masegosa, A., Fernández, A., Langseth, H., Madsen, A., & Nielsen, T. (2015). Parallel importance sampling in conditional linear Gaussian networks. CAEPIA'2015. *Lecture Notes in Artificial Intelligence*, 9422, 36– 46.

Salmerón, A., Langseth, H, Nielsen, T., & Madsen, (2018), A. A review of Inference Algorithms for Hybrid Bayesian Networks. *Journal of Artificial Intelligence Research*, 62, 799-828

Salmerón, A. & Reche, F. (2018). Mixtures of Gaussians as a Proxy in Hybrid Bayesian Networks. Studies in Systems, *Decision and Control*. Volume 142, 2018, Pages 367-374

Shenoy, P. P., & Shafer, G. (1990). Axioms for probability and belief functions propagation. In Shachter, R., Levitt, T., Lemmer, J., & Kanal, L. (Eds.), *Uncertainty in Artificial Intelligence*, 4, pp. 169–198. North Holland, Amsterdam.

Shenoy, P., & West, J. (2011). Inference in hybrid Bayesian networks using mixtures of polynomials. *International Journal of Approximate Reasoning*, 52, 641–657.

Sun, W., Chang, K., & Laskey, K. (2010). Scalable inference for hybrid Bayesian networks with full density estimations. In *Proceedings of the 13th Conference on Information Fusion*, pp. 1–8.

Winn, J., & Bishop, C. (2005). Variational message passing. *Journal of Machine Learning Research*, 6, 661–694.

Zhu, M., Liu, S., & Yang, Y. (2012). Propagation in CLG Bayesian networks based on semantic modeling. *Artificial Intelligence Review*, 38, 149–162.