

Un mecanismo de reducción de alarma lógico por medio de Redes Bayesianas

Luis Orellana Altamirano.

Universidad de Santiago de Chile, Chile
Avenida Ecuador #3659. Estación Central, Santiago de Chile., Chile
luis.orellana.a@usach.cl
fernando.cabrera.ga@usach.cl

Abstracto.

En la medicina moderna un aspecto que se vuelve relevante es la atención efectiva del paciente. Para ello se hace necesario que la toma de decisiones médicas sea precisa e inmediata. La medición continua de los parámetros del paciente como por ejemplo: la frecuencia cardíaca, la frecuencia respiratoria, la presión arterial, la saturación de sangre y oxígeno se realiza a través de monitores electrónicos que son usados para recopilar y mostrar datos fisiológicos. La pregunta principal que busca responder este estudio es ¿Qué tan efectivas son relaciones causales y probabilísticas establecidas por el método de redes bayesianas en comparación a la literatura médica?

Palabras Clave: Monitoreo de pacientes, redes bayesianas, blearn.

1 Introducción

El campo de la medicina ha sido una de las áreas de aplicación favoritas para herramientas de investigación de minería de datos. Una de las razones de este enfoque puede ser el hecho de que la toma de decisiones en medicina es difícil. Esto no solo se debe a los innumerables síntomas, pruebas, enfermedades y opciones de tratamiento que conforman el cuerpo de conocimiento médico, sino también a la complejidad de la fisiología humana.

Dada la falta de práctica de reglas lógicamente exhaustivas y una cierta medida de conocimiento incompleto (teórico o práctico) es posible utilizar redes bayesianas para modelar relaciones causales (o de diagnóstico) usando un grado de creencia, lo que permite razonar bajo incertidumbre.

1.1 Dataset “Alarm”

El dataset Alarm del paquete “blearn” en R está compuesto principalmente por datos provenientes de conocimiento médico. Consta de 37 variables (8 variables de diagnóstico, 13 variables intermedias y 16 recomendaciones) y 20000 instancias. La naturaleza de estas variables son de tipo categóricas ordinales. Los posibles valores que pueden tomar las variables son:

- Variables de diagnóstico: verdadero y falso.
- Variables intermedias: escala de valores (“zero”, “low”, “normal”, “high”).
- Variables de recomendación: escala de valores (“zero”, “low”, “normal”, “high”).

1.2 Método de Minería de Datos

Con la finalidad de obtener conocimiento a través de las técnicas de minería de datos, se utilizará el método de redes bayesianas. Las redes bayesianas (BN) son modelos gráficos capaces de representar y manipular eficientemente las distribuciones de probabilidad n-dimensional [1]. La base de conocimiento que codifica un BN puede ser vista como un modelo de doble representación dividido en un cualitativo y cuantitativo.

Los algoritmos de redes bayesianas utilizadas en este trabajo provienen de la librería “blearn” del software R. Ella proporciona un aprendizaje estructural de redes bayesianas, a través de algoritmos basados en restricciones, puntaje y modelos híbridos. Son evaluados 3 algoritmos de redes bayesianas descritos a continuación:

El método Hill Climbing(HC) posee un buen equilibrio entre los requisitos de la CPU, la precisión del modelo obtenido y la facilidad de implementación. Garantiza obtener un I-map mínimo, además este enfoque es realmente apropiado para lidiar con dominios de alta dimensión [2].

El algoritmo Max-Min Hill-Climbing (MMHC) combina ideas de aprendizaje local, basado en restricciones y técnicas de búsqueda y puntuación de una manera eficaz. Reconstruye por primera vez el esqueleto de una red bayesiana y luego realiza una búsqueda greedy de Hill Climbing bayesiana para orientar los bordes [3].

El maximum minimum parents and child algorithm(MMPC) es un algoritmo de aprendizaje de estructura para determinar el conjunto padre-hijo de una variable de destino T basada en un conjunto de datos. Es un algoritmo de descubrimiento local [4].

1.3 Procedimiento de Ajuste del Modelo

Con la finalidad de obtener un modelo de red bayesiana que permita realizar consultas sobre el de manera eficiente, se empleó una serie de métodos provenientes de la librería “blearn” de R. En la **Tabla 1**, es posible visualizar los distintos métodos empleados

para generar redes bayesianas, los que son comparados con la finalidad de encontrar el mejor posible.

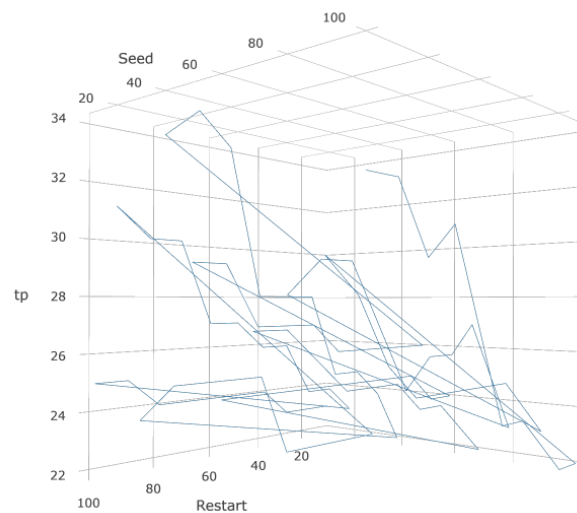
Tabla 1. Resultados de los modelos dados por “bnlearn”.

	TP	BIC
HC	22	-220761.7
MMHC	16	-311189.8
MMPC	0	Error (Parcialmente Direccionado)

Las comparativas de los distintos modelos fueron realizadas por las métricas BIC y TP. Para el caso de las redes bayesianas “TP” hace referencia a la cantidad de arcos entre nodos que coinciden con el modelo planteado originalmente por Beinlich I, Suermondt HJ, Chavez RM, Cooper GF (1989). Como es posible observar en los resultados, el mejor resultado obtenido fue con el algoritmo de búsqueda “HC”.

Los parámetros que fueron empleados para mejorar el modelo “HC” son: “Seed” y “Restart”. El ajuste fue efectuado haciendo variaciones en tales parámetros, los cuales variaron entre 1 y 100 y de 10 en 10, lo cual permitió realizar 100 distintas combinaciones de ellos. Las calidades de los modelos generados fueron evaluadas en base a “TP” y esto se ve reflejado en la **Imagen 1**.

Imagen 1. Ajuste del modelo por medio de “Seed” y “Reset”.

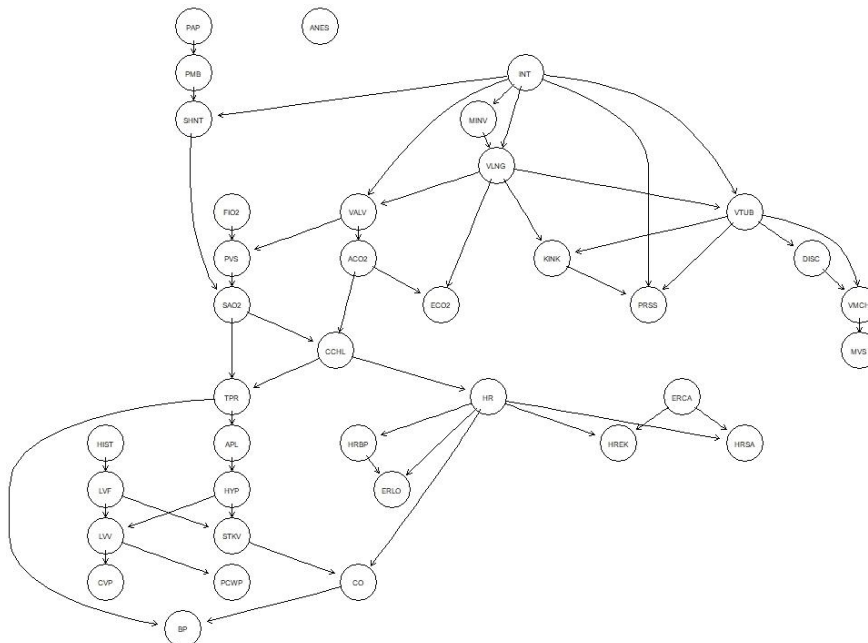


El mejor modelo se obtuvo con “Seed” 40 y “Restart” 90. El “TP” obtenido por medio de tales parámetros es de 37 arcos correctos de un total de 46. El BIC obtenido es -218690.1, el cual mejoró en comparación al obtenido inicialmente y que se muestra en la **Tabla 1**.

1.4 Resultados Obtenidos (discusión)

Por medio de los ajustes fue posible obtener un modelo con 37 arcos correctos. En comparación al grafo inicial, se obtuvo una mejora en la cantidad de arcos correctamente dirigido, recordando que el grafo inicial poseía solamente 22 arcos correctos. Se puede observar el resultado obtenido por el mejor modelo en la **Imagen 2**, el cual es posible de contrastar con el grafo creado gracias a juicios de expertos por Beinlich I, Suermondt HJ, Chavez RM, Cooper GF (1989).

Imagen 2. Mejor grafo generado por medio del proceso de ajuste del modelo.



Se sabe que uno de los tratamientos más recurrentes ante una anafilaxis es la intubación, lo cual es corroborado por Gavalas M, Sadana A, Metcalf S (1998). Ante esto, se realizó la siguiente pregunta con la finalidad de comprobar cuan acertadas pueden ser las respuestas generadas por el modelo:

- **¿Qué tipo de intubación se le suministra a pacientes con anafilaxis?**

Respuesta: La gran mayoría de los pacientes con anafilaxis (92%), se les administra intubación normal.

Con esto se corrobora la fiabilidad de la respuesta obtenida del modelo en cuanto a esta pregunta planteada ya que concuerda con la literatura.

Por otra parte, se observaron dos nodos padres importantes los cuales son embolia pulmonar (PMB) y anafilaxis (APL). Estas dos variables tienen dos niveles (verdadero y falso) y constituyen enfermedades de alto riesgo para la vida de pacientes lo cual es corroborado por el trabajo de Pumphrey (2000). Sobre esta premisa, además de la pregunta anteriormente mencionada, se formularon algunas preguntas adicionales:

- **¿Cuál es la probabilidad de que el paciente padezca de embolia pulmonar (PMB) si la catecolamina es alta (CCHL) y el ritmo cardiaco es alto (HR)?**
Respuesta: Entre los pacientes que poseen embolia pulmonar, se presenta una gran cantidad de casos (69,2%) en que tienen un ritmo cardiaco alto y su catecolamina es alta.
- **¿Cuál es la probabilidad de que el paciente padezca de embolia pulmonar (PMB) si la catecolamina es alta (CCHL), si el ritmo cardiaco es alto (HR) y el volumen de sangre bombeado por el corazón (CO) es alto?**
Respuesta: Entre los pacientes que poseen embolia pulmonar, se presentan casos (56%) en que tienen un ritmo cardiaco alto, catecolamina es alta y el volumen de sangre bombeado por el corazón es alto.
- **¿Cuál es la probabilidad de que un paciente con ritmo cardiaco alto, catecolamina alta e irrigación sanguínea pulmonar baja padezca de anafilaxis?**
Respuesta: La mayoría de las personas (86,36%) que presentan estas características presentan anafilaxis.

2 Conclusiones

A partir de los resultados obtenidos por medio de la red bayesiana, fue posible corroborar en primera instancia que el modelo se ajusta a la respuesta esperada para el caso del tratamiento de urgencia recurrente para anafilaxis. Por otro lado, se observó por medio de la bibliografía procedente del dataset “alarm”, que el modelo generado guarda importantes similitudes con el esquema planteado por el juicio de expertos.

Ante esto, es posible mencionar que el gran potencial y utilidad que reviste al método de redes bayesianas dado que posee la capacidad de dar respuestas a interrogantes técnicas complejas, las cuales son de mucha utilidad en el presente y permiten ser un apoyo en la medicina moderna. De esta manera, se pudo descubrir por medio de las preguntas efectuadas, que una enfermedad tan grave como la anafilaxis, está íntimamente relacionada con la alta irrigación sanguínea pulmonar y altos niveles de ritmo cardiaco. Otro hallazgo interesante fue algunos factores que están muy relacionados con la embolia pulmonar (la cual es una enfermedad que puede llegar a ser mortal para el paciente), los cuales son la catecolamina, el ritmo cardiaco y la sangre bombeada por el corazón. Para todos los hallazgos mencionados, hay una significancia estadística la cual indica mayores chances de que los eventos consultados ocurran. Ante esto, es posible afirmar

que este método permite obtener conocimiento no trivial, pertinente y de gran valor para no tan solo labores médicas, sino para una amplia gama de escenarios posibles donde una alta complejidad de conformación de los datos está presente.

3 Referencias

1. Pearl, J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo (1988).
2. Gámez J.A., Mateo J.L., Puerta J.M. (2007) A Fast Hill-Climbing Algorithm for Bayesian Networks Structure Learning. In: Mellouli K. (eds) Symbolic and Quantitative Approaches to Reasoning with Uncertainty. ECSQARU 2007. Lecture Notes in Computer Science, vol 4724. Springer, Berlin, Heidelberg.
3. Tsamardinos, I., Brown, L., Aliferis, C. The max-min hill-climbing Bayesian network structure learning algorithm. Springer, (2006).
4. Furrer, L. (2009). Bayesian Network: Graph discovery. China.
5. Beinlich I, Suermondt HJ, Chavez RM, Cooper GF (1989). "The ALARM Monitoring System: A Case Study with Two Probabilistic Inference Techniques for Belief Networks." In "Proceedings of the 2nd European Conference on Artificial Intelligence in Medicine", pp. 247-256. Springer-Verlag
6. Gavalas M, Sadana A, Metcalf S (1998). Guidelines for the management of anaphylaxis in the emergency department. Emergency Medicine Journal 15:96-98.
7. Pumphrey (2000), Lessons for management of anaphylaxis from a study of fatal reactions. Clinical & Experimental Allergy, 30: 1144–1150. doi:10.1046/j.1365-2222.2000.00864.x