

Aprendizaje de Estructuras de Independencia de Modelos Probabilísticos Gráficos

Facundo Bromberg
fbromberg@frm.utn.edu.ar

Federico Schlüter
fschluter@frm.utn.edu.ar

Laboratorio de Inteligencia Artificial
Departamento de Sistemas de Información,
Facultad Regional Mendoza, Universidad Tecnológica Nacional

RESUMEN

Nuestra investigación se enmarca en el problema del aprendizaje, a partir de datos, de estructuras de independencia de modelos probabilísticos gráficos. Es de especial interés el aprendizaje automatizado de estos modelos a partir de datos, debido principalmente a la presencia cada vez más ubicua de datos digitales. El campo del aprendizaje de máquinas en general, y en particular los miembros de nuestro laboratorio, se han concentrado en el aprendizaje del grafo que representa la estructura de independencias de estos modelos. Durante su tesis doctoral el Dr Bromberg (Bromberg 2007) se ha concentrado en el diseño de algoritmos de aprendizaje de estructuras que utilizan un enfoque basado en independencias (Spirtes et. al. 2000), en contraste con los algoritmos basados en puntaje (Lam and Bacchus 1994, Heckerman 1995). Estos últimos recurren a técnicas para aprendizaje de modelos mas establecidas en la estadística como ser por ejemplo la maximización de la verosimilitud (probabilidad del los datos dado el modelo). El enfoque basado en independencias, en cambio, utiliza un enfoque mas directo para aprender la estructura de independencias del modelo, realizando tests estadísticos de independencia entre las variables aleatorias del sistema. Durante su estadía en Iowa State University, y durante el pasado año ya en UTN-FRM, el Dr. Bromberg ha contribuido con varios algoritmos para el aprendizaje de estructuras de modelos Markovianos con el objetivo de reducir la cantidad de tests estadísticos necesarios durante su ejecución. Recientemente el laboratorio se ha enfocado en un problema más exigente y más importante, el diseño de algoritmos que ante la misma entrada de datos, produzcan modelos de mejor calidad. Estos algoritmos son aplicables tanto a redes Markovianas como Bayesianas.

Palabras clave: *Machine Learning, Probabilistic Graphical Models, Independence-based Structure Learning, Probabilistic Reasoning.*

1. INTRODUCCION

La principal línea de nuestra investigación estudia mecanismos de aprendizaje de estructuras de modelos probabilísticos gráficos, con el objetivo de proponer nuevas perspectivas o enfoques al problema, aportando propuestas que mejoren la calidad y la eficiencia de los algoritmos propuestos hasta la actualidad. El aprendizaje de estructuras se efectúa mediante mecanismos que infieren, a partir de los datos, modelos probabilísticos gráficos (Pearl 1988, Lauritzen 1996). Entre éstos, redes Bayesianas y redes Markovianas son importantes modelos que han contribuido con aumentos considerables en la eficiencia del proceso de razonamiento probabilístico, ofreciendo mecanismos de inferencia estadística prácticos para estimación de valores en diversas ramas de la ciencia.

Las redes Bayesianas han encontrado mayor aplicabilidad en la representación de modelos causales (Pearl, 2000), principalmente debido a la direccionalidad del grafo (Friedman et al., 2000). Las redes Markovianas, en cambio, han sido utilizadas principalmente para representar relaciones espaciales simétricas, por lo que históricamente se las ha llamado *Markov Random Fields* (Kindermann and Snell 1980, Geman and Geman 1984, Besag et. al. 1991). En estas aplicaciones ninguna variable puede identificarse como la causa de las otras. Resaltamos los siguientes ejemplos de campos de aplicación de las redes Markovianas: en visión computacional (Besag et. al. 1991, and Anguelov et. al. 2005) para restaurar imágenes ruidosas, clasificar texturas y segmentar imágenes. Recientemente, la comunidad de Data Mining se ha sumado al uso de estos modelos demostrando gran interés en su uso para mineo espacial de datos, con aplicaciones en geografía, transporte, agricultura, climatología, ecología y otros (Shekhar et al., 2004).

Un modelo gráfico está conformado por un grafo que representa una estructura de independencias entre las variables del modelo (representadas por los nodos del grafo), y por un conjunto de parámetros numéricos para modelar las probabilidades. Por ello, el aprendizaje automatizado de modelos gráficos presenta dos desafíos: primeramente, aprender la estructura de independencias, para luego estimar los

parámetros numéricos (Pearl 1988). Para redes Bayesianas el grafo está conformado por aristas dirigidas, mientras que para redes Markovianas, está conformado por aristas no dirigidas. El propósito principal de estos grafos es codificar las independencias condicionales existentes entre las variables aleatorias del sistema. La identificación explícita de estas independencias es la principal causa en la importante reducción de los requerimientos computacionales y espaciales de la inferencia probabilística, evitando cálculos innecesarios entre variables independientes.

A pesar de ser interesante por sí mismo y ser un problema de gran dificultad, no consideramos en nuestra investigación el problema de aprendizaje de parámetros, concentrándonos en el problema del aprendizaje de estructuras. Tradicionalmente, el aprendizaje del grafo a partir de los datos se realiza por medio del método de maximización de algún "puntaje" (e.g., maximización de la verosimilitud), el cual consiste en una optimización en el espacio de todas las estructuras posibles, identificando aquella con mayor puntaje (Lam and Bacchus 1994, Heckerman 1995). Desde un principio, el aprendizaje de la estructura de los modelos Markovianos, en contraste con los Bayesianos, ha presentado importantes dificultades computacionales, principalmente por el hecho de que el cálculo del puntaje para ciertos modelos requiere estimar los parámetros del mismo, un problema NP-completo (Barahona 1982). Esta dificultad computacional ha forzado a los usuarios de estos modelos a recurrir al conocimiento de expertos para estimar las estructuras gráficas.

Los resultados obtenidos en investigaciones recientes (2009, Gandhi et al. 2008, Margaritis and Bromberg 2009) han producido una serie de algoritmos capaces de aprender eficientemente la estructura de modelos Markovianos basándose en el enfoque de independencias (Spirtes et. al. 2000), cuya ventaja radica en que se puede prescindir del aprendizaje de los parámetros numéricos para aprender la estructura de independencias, eliminando así una de las principales causas de ineficiencia de los algoritmos basados en puntaje. El enfoque de aprendizaje basado en independencias consiste en la ejecución de tests estadísticos para estimar independencias condicionales entre las variables aleatorias de un sistema (e.g., es la variable aleatoria X independiente de la variable aleatoria Y dado el conocimiento de los valores del conjunto de variables aleatorias Z ?). Desafortunadamente, la enorme ventaja computacional obtenida por los algoritmos basados en independencias se ve opacada por la sensibilidad de la calidad de las estructuras producidas por estos algoritmos cuando los datos disponibles son insuficientes.

La calidad de las estructuras resultantes de los algoritmos basados en independencias depende fuertemente de la calidad de los tests estadísticos de independencias llevados a cabo durante su ejecución. Se sabe que estos tests pueden resultar en decisiones erradas cuando la cantidad de datos es insuficiente. Si bien la disponibilidad de datos crece continuamente con la innovación en materia de instrumentación digital, para mantener su calidad los tests requieren una cantidad de datos exponencial en la cantidad de variables involucradas en el test (Agresti 2002).

2. LINEAS DE INVESTIGACION Y DESARROLLO

Por lo expuesto, nuestra investigación se ha orientado recientemente en la búsqueda de enfoques novedosos para atacar la problemática que subyace a la falta de certeza de los resultados de los tests, lo que disminuye la calidad de los modelos generados. Hasta el momento hemos detectado dos enfoques complementarios para la mejora de la calidad de las estructuras: (i) diseñar algoritmos para mejorar la calidad de los tests, (ii) ante un test dado, diseñar algoritmos para mejorar la calidad de las estructuras.

En términos generales, el primer enfoque detecta tests incorrectos y corrige sus valores. Esto es posible explotando restricciones existentes entre los tests de un conjunto de variables. Estas restricciones, llamadas *axiomas de independencia* (Pearl, 1988, Dawid 1979) son fundamentales y son resultado directo de los axiomas de Probabilidad, es decir, solo tests erróneos pueden violarlos. Por ello, nuestro enfoque propone buscar violaciones de los axiomas de independencia para detectar los tests erróneos. Dada una restricción violada, una dificultad importante es la de detectar cual (o cuales) de los tests involucrados en esta restricción es el incorrecto. En general la decisión requiere información de las otras violaciones en donde participa cada test. Lamentablemente existe un número exponencial de restricciones por test (con respecto al número de variables del dominio). En (Bromberg and Margaritis, 2009), se propone con éxito la utilización del formalismo de Argumentación (Dung, 95) para decidir eficientemente si un test es errado o no en base al conjunto de restricciones satisfechas y violadas por el test. Bajo el punto de vista de la Argumentación, vemos al problema como una base de conocimiento inconsistente, compuesta por proposiciones de independencias (e.g., es X independiente de Y dado el conjunto de variables Z ?), reglas lógicas de Pearl que las relacionan, y potenciales inconsistencias donde ciertas proposiciones violan una o más reglas. La importancia del formalismo radica en que formalismos tradicionales de razonamiento lógico no son sólidos cuando la base de conocimiento es

inconsistente. Para refinar aún más el formalismo de inferencia por Argumentación, es posible cuantificar las reglas con preferencias (Amgoud and Cayrol, 2002), un valor numérico que permite decidir entre dos reglas conflictivas (una demuestra la negación de la otra) cuando no existe otra información. En nuestro caso, se tomó como preferencias a la probabilidad de que un test estadístico es correcto (calculable a partir de valores numéricos arrojados por tests estadísticos, ver referencia para mas detalles). De esta manera, el formalismo de Argumentación con preferencias se convierte en una instancia de un formalismo para el razonamiento bajo incertidumbre. Los resultados positivos obtenidos con este formalismo demuestran la viabilidad del enfoque de mejora de la calidad basado en restricciones y propone una importante línea de investigación de nuevos algoritmos. Recientemente Federico Schlüter se ha sumado al laboratorio con beca doctoral de UTN-PRH para trabajar en esta línea de investigación, planeando investigar en el corto plazo el uso de las *Markov Logic Networks* (Richardson 2006) como formalismo de razonamiento bajo incertidumbre para la inferencia del valor de independencia de los tests, basado en restricciones. Este formalismo es prometedor por razonar con reglas en lógica de primer orden, reduciendo la complejidad presentada por el formalismo de Argumentación y cualquier otro que requiera de la proposicionalización de las reglas de Pearl.

El segundo enfoque para la mejora de la calidad considera el diseño de algoritmos de aprendizaje novedosos que tomen en cuenta el hecho de que las independencias arrojadas por los tests de independencia pueden estar errados. Este enfoque es el utilizado por PFMN (Margaritis and Bromberg 2009) y por GSMNS, un algoritmo que generaliza GSMN que se encuentra actualmente en la etapa de diseño de heurísticas para lidiar con su costo computacional exponencial, pero cuyos resultados preliminares demuestran mejoras en la calidad de los modelos generados altamente promisorias (ver resultados en Sección 3). Pasamos a explicar el algoritmo GSMNS en detalle en la siguiente sección.

2.1. ALGORITMO GSMNS

Dado un conjunto de variables aleatorias V distribuidas de acuerdo a una distribución \Pr (desconocida), y una muestra estadística de asignaciones conjuntas de las variables V distribuidas de acuerdo a \Pr , llamado simplemente conjunto de datos D , el algoritmo **GSMNS** (Grow Shrink Markov Network with Search), arroja como resultado un grafo no-dirigido G que representa la estructura de independencias de una red Markoviana M . Cuando los tests son correctos, lo ocurre cuando la muestra de datos es correctamente sampleada de

\Pr y es infinita, es posible demostrar que la distribución de probabilidad representada por la red aprendida M coincide con la distribución subyacente \Pr . En la práctica esta condición rara vez se cumple, produciendo redes incorrectas.

El algoritmo GSMNS al igual que el algoritmo **GSMN** (Grow Shrink Markov Network) al cual generaliza, aprende la estructura de independencias G infiriendo el *Markov blanket* $MB(X)$ de cada variable X en V . De acuerdo a (Pearl 88), $MB(X)$ consiste en un subconjunto de $V - \{X\}$ de variables aleatorias, vecinas de X en la estructura. En otras palabras, aprender el MB de cada variable en V es suficiente para aprender la estructura. En (Dimitris and Thrun 2000) se presenta el algoritmo **GS** (Grow Shrink) que dado una variable aleatoria X , y el conjunto de datos D y de variables V , aprende $MB(X)$ realizando tests estadísticos de independencia en dos etapas:

Grow Phase:

for all Y in $V - \{X\}$, if $\neg I(X, Y | S)$ then $S \leftarrow S \cup \{Y\}$

Shrink Phase:

for all Y in S , if $I(X, Y | S - \{Y\})$ then $S \leftarrow S - \{Y\}$

donde $I(X, Y | S)$ indica si la variable X es independiente de la variable Y condicionado en el conjunto de variable S , de acuerdo al test de independencia.

El algoritmo GSMN utiliza una versión del algoritmo GS que preestablece el ordenamiento de variables para recorrer los bucles de ambas fases por medio de una heurística sencilla. Este algoritmo confía plenamente en el resultado de los tests (potencialmente de baja calidad o hasta erróneos), dando indicios de un posible error. Esto produce un algoritmo eficiente (cuadrático en el número de variables) pero que ante falta de datos, y por ende potenciales errores en los tests, el ordenamiento utilizado y la confianza ciega en las independencias obtenidas puede resultar en una estructura errónea.

Por esto proponemos una generalización de GS, el algoritmo GSS, que flexibiliza tanto el ordenamiento de variables, como la asignación de valores de independencias de los tests. Para ello proponemos primero una medida de calidad de un *Markov blanket*, calculada a partir de la medida de calidad de los tests utilizados para aprenderlo (ver apartado 2.1.1), y segundo, la realización de una búsqueda sistemática de todos los posibles ordenamientos y todas las posibles asignaciones (ver apartado 2.1.2.), devolviendo como resultado el *blanket* que maximice la medida de calidad. Por el momento la búsqueda se realiza en la etapa de *grow*, esperando diseñar e implementar una búsqueda en la etapa de *shrink* en el mediano plazo.

2.1.1. Medida de calidad de un Markov blanket

Dado un triplete $(X, Y | \mathbf{Z})$, un test estadístico arroja, además de un valor de independencia, un valor numérico que indica la “calidad” del test. En nuestros experimentos utilizamos un test de independencia Bayesiano propuesto por (Margaritis and Bromberg 2009) que calcula la probabilidad de independencia/dependencia. Si denotamos por $\Pr(X, Y | \mathbf{Z})$ la probabilidad de independencia entre X e Y dado \mathbf{Z} , proponemos calcular la medida de calidad del Markov blanket $\mathbf{MB}(X)$, denotada $\Pr(\mathbf{MB}(X))$, utilizando el conjunto de tests $T(\mathbf{MB}(X))$ realizados durante la etapa de grow de la siguiente manera:

$$\Pr(\mathbf{MB}(X)) = \text{Prod}_{t \in T(\mathbf{MB}(X))} \Pr(t),$$

es decir, como la productoria de las probabilidades de independencia de los tests realizados durante la etapa de grow.

2.1.2. Búsqueda de ordenamiento y asignación óptima

Para encontrar el ordenamiento de $V-\{X\}$ y asignación de independencias que maximicen la medida de calidad $\Pr(\mathbf{MB}(X))$ del blanket de X , proponemos realizar una búsqueda en árboles del tipo A^* (Russell and Norvig, 2002), a través de los distintos estados de la etapa de grow, es decir, el estado de \mathbf{S} y las variables en $V-\{X\}$ ya testeadas, que denotamos \mathbf{P} (por predecesores). Los operadores entre estados son los tests estadísticos que pueden realizarse en ese estado, y el costo de cada operador asociado al test t es $-\log(\Pr(t))$. De esta manera maximizar la calidad de un blanket equivale a minimizar la sumatoria de los costos. Para explorar todos los ordenamientos y todas las asignaciones, proponemos como operadores de un estado $[\mathbf{P}, \mathbf{S}]$ aquellos que nos lleven a los siguientes estados sucesores:

$$\{[\mathbf{P} \square \{U\}, \mathbf{S}] \text{ y } [\mathbf{P} \square \{U\}, \mathbf{S} \square \{U\}] \mid U \square \mathbf{P}\},$$

donde el primer caso corresponde a una asignación de independencia (i.e., $I(X, U | \mathbf{S})$) y el segundo a una asignación de dependencia (i.e., $\neg I(X, U | \mathbf{S})$).

De esta manera, la optimalidad del costo de camino de la solución garantizada por A^* equivale a una maximización de la productoria de la calidad de los tests. En otras palabras, a una calidad del *blanket* máxima. Para completar la descripción de la búsqueda A^* requerimos proponer una heurística que estime, para cada nodo del árbol de búsqueda, el costo de camino óptimo hasta el objetivo, i.e., la calidad óptima de los tests no realizados aún. En nuestros experimentos consideramos como

heurística la probabilidad de los tests que restan calculada en forma voraz (eligiendo para expandir directamente el mejor de entre los hijos de un nodo). Esto garantiza un cálculo eficiente de la heurística.

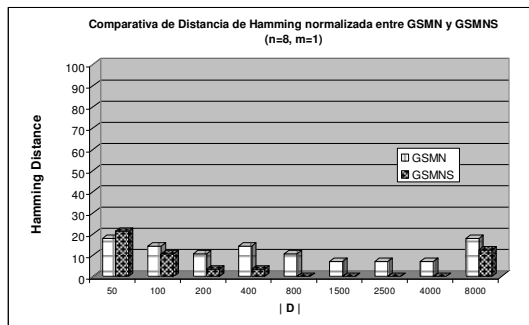
3. RESULTADOS OBTENIDOS

La optimalidad garantizada por las búsquedas en árboles conllevan un costo en tiempo y en memoria exponenciales, salvo para heurísticas muy buenas cuya estimación del costo restante es logarítmicamente cercano al costo real. Esta condición es extremadamente difícil de satisfacer en la práctica y por ello no esperamos, en nuestros experimentos, complejidades espaciales y temporales polinomiales. Nos interesamos, sin embargo, en poder estimar a través de estos experimentos una cota superior (por ser óptimos los algoritmos) en la mejora en la calidad a la cual podemos aspirar. Esperamos en un futuro encontrar heurísticas, o algoritmos de optimización alternativos eficientes, que no reduzcan en demasía las mejoras en la calidad obtenidas con el algoritmo óptimo presentado aquí.

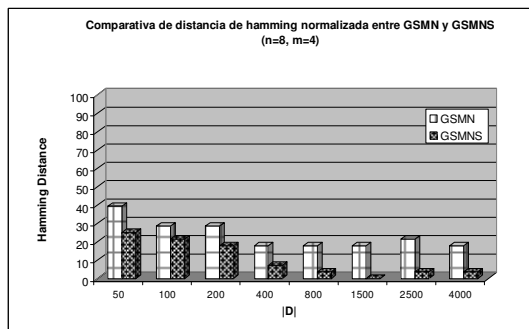
Para evaluar la performance del algoritmo GSMNS comparamos la calidad de las redes que produce con la calidad de las redes producidas por GSMN. Para ello, usamos diversos datasets generados a través de algoritmo de sampleo (Gibbs sampler) a partir de redes Markovianas generadas aleatoriamente. De esta manera conocemos la red subyacente posibilitando una comparación precisa entre el modelo aprendido y el real. Se generaron redes aleatorias de $n=8$ variables, uniendo cada nodo de la red con m variables seleccionadas aleatoria y uniformemente de entre los nodos restantes. Se presentan resultados para $m=1,2,4$. Para estimar la calidad de la red generada medimos la *distancia de Hamming normalizada* entre la red generada y la red real. La distancia de Hamming entre dos redes consiste en la cantidad de aristas existentes en una de las redes e inexistentes en la otra (y viceversa). Se normaliza dividiendo por la cantidad total de pares de variables (i.e., el valor máximo de la distancia de Hamming no normalizada), obteniendo un valor entre 0 y 1. De esta manera, una distancia igual a 0 indica que las dos redes son iguales, mientras que un valor igual a 1 indica que son una el negativo de la otra (siempre que una tiene una arista la otra no).

Dado que la calidad de los tests dependen de la cantidad de datos en el data set, corrimos el algoritmo para subconjuntos de D con un número $|D|$ creciente de renglones, también sampleados aleatoriamente del data set original. Las figuras 1, 2 y 3 comparan la distancia de Hamming normalizada entre la red obtenida utilizando GSMN y GSMNS, y la red real, para los tres valores de m . Puede observarse que en todos los casos la distancia de

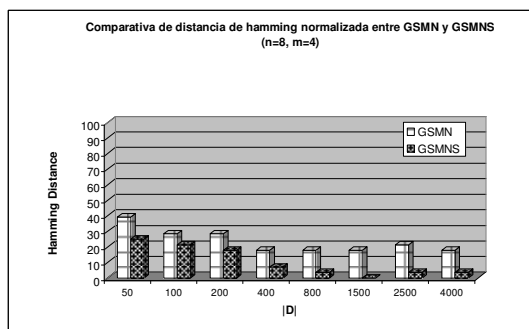
Hamming disminuye para $|D|$ crecientes tanto para GSMN como para GSMNS, lo cual es esperado ya que la calidad de los tests crece con $|D|$. Puede notarse sin embargo que la distancia de Hamming de GSMNS disminuye más rápido que la de su contraparte GSMN, demostrando una mejora en la calidad del primero vs. el segundo. Actualmente los resultados obtenidos para el mismo son prometedores, ya que registran mejores estructuras de independencia, pero aún debemos continuar trabajando sobre la performance, ya que el procesamiento de la solución requiere un tiempo exponencial respecto de la cantidad de variables de la estructura.



-Figura 1: Comparación de la distancia de Hamming normalizada de las redes obtenidas por GSMN y GSMNS para $n=8$, $m=1$, para tamaños crecientes del dataset de entrada.



-Figura 2: Comparación de la distancia de Hamming normalizada de las redes obtenidas por GSMN y GSMNS para $n=8$, $m=2$, para tamaños crecientes del dataset de entrada.



-Figura 3: Comparación de la distancia de Hamming normalizada de las redes obtenidas por GSMN y GSMNS para $n=8$, $m=4$, para tamaños crecientes del dataset de entrada.

4. FORMACION DE RECURSOS HUMANOS Y LINEAS DE INVESTIGACIÓN EMERGENTES

El laboratorio de Inteligencia Artificial en la FRM-UTN ha sido creado a comienzos del 2008 por el Dr. Bromberg tras su incorporación a la Universidad como docente-investigador. Proyectos en curso son:

Aprendizaje automatizado de estructuras de independencias de modelos probabilísticos: En esta investigación colaboran Dr. Bromberg (beca de Fundación UTN y beca post doctoral de reinscripción de CONICET), el Dr. D. Margaritis, profesor asociado en la Iowa State University y ex-director de tesis del Dr. Bromberg, Ing. Federico Schlüter, doctorando bajo la dirección del Dr. Bromberg, financiado con beca UTN, bajo el marco del PRH-2007, y por último Ing. Franco Farinelli (ad-honorem).

GEARS: (<http://ai.frm.utn.edu.ar/gears>) Producto del interés de un grupo de alumnos el laboratorio esta explorando líneas de investigación en el área de la Web Semántica, para lo cual creo el grupo GEARs (Grupo de Estudio de Aplicaciones de Redes Semánticas) en el que participan alumnos de quinto año de la carrera de sistemas A. Edera, M. Garsiolo y M. Pasquier. Este grupo se encuentra desarrollando la aplicación KHIPU para autenticación semántica en el marco de su proyecto final de carrera.

ByOS: (<http://ai.frm.utn.edu.ar/byos>) Búsqueda y Optimización Subjetiva. F. Bromberg, junto con los estudiantes de quinto año de Ing. en Sistemas M. Spertino y S. Perez están investigando técnicas de optimización interactivas basadas en algoritmos genéticos.

De los mencionados arriba, A. Edera, M. Spertino y S. Perez han demostrado interés en continuar con sus estudios realizando un doctorado bajo la dirección del Dr. Bromberg, con intenciones de presentarse a la próxima convocatoria de becas doctorales de CONICET

5. BIBLIOGRAFIA

Agresti, A. (2002). *Categorical Data Analysis*. Wiley, 2nd edition.
 Amgoud and Cayrol (2002). *A reasoning model based on the production of acceptable arguments*. *Annals of Mathematics and Artificial Intelligence*, 34:197-215.
 Angelov, D., Taskar, B., Chatalbashev, V., Koller, D., Gupta, D., Heitz, G., and Ng, A. (2005). *Discriminative learning of Markov random*

- fields for segmentation of 3D range data. Proc. of CVPR.
- Barahona, F.** (1982). *On the computational complexity of ising spin glass models*. Journal of Physics A: Mathematical and General, **15**(19): 3241-3253.
- Besag, J., York, J., and Mollie, A.** (1991). *Bayesian image restoration with two applications in spatial statistics*. Annals of the Institute of Statistical Mathematics, 43:1-59.
- Bromberg, F., Margaritis, D. and Honavar, V.** (2006). *Efficient Markov Network Structure Discovery using Independence Tests*. Proceedings of SIAM Data Mining 2006. Bethesda, Maryland.
- Bromberg, F.** (2007). *Markov network structure discovery using independence tests*. PhD Thesis, Iowa State University.
- Bromberg, F. and Margaritis, D.** (2009). *Improving the Reliability of Causal Discovery from Small Data Sets using the Argumentation Framework*. Journal of Machine Learning Research, **10**(Feb):301—340.
- Dawid, A. P.** (1979). *Conditional independence in statistical theory*. Jnal. Royal Stat. Society, **41**:1-31.
- Dung, P. M.** (1995). *On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games*. Artificial Intelligence, **77**:321-357.
- Friedman, N., Linial, M., Nachman, I., and Peér, D.** (2000). *Using Bayesian networks to analyze expression data*. Computational Biology, **7**:601-620.
- Gandhi P., Bromberg F. and Margaritis D.** (2008) *Dynamic inference based learning Markov Network Structure*. Proc. of SIAM Data Mining 2008, Atlanta Georgia.
- Heckerman, D., Geiger, D., and Chickering, D. M.** (1995). *Learning Bayesian networks: The combination of knowledge and statistical data*. Machine Learning. **20**:197-243, 1995.
- Kindermann, R. and Snell, J. L.** (1980). *Markov Random Fields and Their Applications*. AMS.
- Lam, W. and Bacchus, F.** (1994). *Learning Bayesian belief networks: an approach based on the MDL principle*. Comp. Intelligence, **10**:269-293.
- Lauritzen, S. L.** (1996). *Graphical Models*. Oxford Statistical Science Series, vol 17. Oxford U. Press.
- Bromberg F. and Maragaritis D.** (2009) *Efficient Markov Network Discovery Using Particle Filters*. This is a preprint of an Article accepted for publication in *Computational Intelligence*.
- Margaritis, D. and Thrun, S.** (2000). *Bayesian network induction via local neighborhoods*. In Solla, S., Leen, T., and Müller, K.R., editors, Advances in Neural information Proccessing Systems 12, pp. 505-511. MIT Press.
- Pearl, J.** (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*.
- Pearl, J.** (2000). *Causality*. Cambridge Univ. Press.
- Russel, S. and Norvig, P.** (2002). *Artificial Intelligence. A Modern Approach*. Prentice Hall 2nd Edition.
- S. Shekhar, P. Zhang, Y. Huang, and R. Vatsavai.** (2004) Trends in spatial data mining. In H. Kargupta, A. et. al. editors, Data Mining: Next Generation Challenges and Future Directions, **19**, 357–379. AAAI Press / The MIT Press, 2004.
- Richardson R., Domingos, P.** (2006) *Marklov Logic Networks*. *Machine Learning*, **62** (2006), pp 107-136.
- Spirtes, P., Glymour, C., and Scheines, R.** (2000). *Causation, Prediction, and Search*. Adaptive Computation and Machine Learning Series. MIT Press, 2nd edition.