

RELACIONES CAUSALES EN REGLAS DE ASOCIACION

M. Amparo Vila F.
Universidad de Granada
DECSAI-España
vila@decsai.ugr.es

Daniel Sánchez F.
Universidad de Granada
DECSAI-España
daniel@decsai.ug.es

Luis Escobar R.
U. Tecnológica Metropolitana
Depto de Informática Chile
laescoba@utem.cl

Resumen

A pesar de que existen técnicas desarrolladas para descubrir relaciones causales entre datos controlados (o experimentales), no está resuelto como puede hacerse en datos solamente observados (capturados y almacenados, y sobre los cuales no es posible experimentar). Para encontrar relaciones causales en datos observados, se han usado métodos basados en grafos, los que resultan excesivamente complejos cuando son aplicados en un típico gran conjunto de datos.

La teoría de la inferencia causal en Minería de Datos consta de dos partes: la inferencia causal en datos estadísticos (la que contempla la matemática, y la filosofía causal) y la parte de Minería de Datos que tiene que ver con técnicas que permitan seleccionar un conjunto de datos observados posibles de relacionarse causalmente. Todo esto en grandes volúmenes de datos almacenados en Base de Datos.

Las técnicas existentes de minería causal se basan en técnicas más complejas de descubrimiento causal en datos estadísticos y en datos experimentales. De esta manera a partir de la filosofía causal y la matemática es posible también inferir relaciones causales en grandes Bases de Datos. Es necesario destacar que la búsqueda de causalidad en Bases de Datos es compleja, en el sentido que no es posible partir de una hipótesis, sin embargo, restringiéndose a un subconjunto de datos a través de la utilización de la técnica de minería de datos denominada “reglas de asociación” se podría obtener información de presunta causalidad entre ellos.

Palabras Clave: Minería de Datos, Causalidad, Reglas de Asociación, KDD

1. INTRODUCCIÓN

1.1. El problema

El interés en las reglas de asociación es que ellas ofrecen la promesa (o ilusión) de causalidad, o al menos entregan relaciones predictivas. Sin embargo, las reglas de

asociación sólo calculan la frecuencia de ocurrencias de uniones entre los atributos; Ellas no expresan una relación causal. Sería de mucha utilidad poder descubrir las relaciones causales a partir de reglas de asociación en el contexto de la minería de datos.

1.2. Objetivos y motivación de la investigación

El objetivo de este trabajo de investigación es realizar una revisión de los conceptos relativos a la causalidad desde diversos puntos de vista, filosófico, matemático, etc., para ello se han recopilado diferentes modelos matemáticos y técnicas utilizados para representar relaciones causales y para juzgar cuando una relación causal es válida o no lo es. Para lo cual se desea proponer dentro del ámbito de la Minería de Datos, una “Técnica de Minería de Relaciones Causales” a través del post-procesamiento de los resultados entregados por la búsqueda de reglas de asociación.

En cuanto a la motivación, se puede precisar como sigue; “En muchas ocasiones cuando un usuario de un sistema de minería de datos se encuentra buscando y obteniendo reglas de asociación, lo que este usuario busca y/o como interpreta dichas reglas de asociación son las relaciones causales, pero en la práctica una regla de asociación no representa en general una relación causal.

2. INTRODUCCION A LA MINERIA DE DATOS Y EXTRACCION DE CONOCIMIENTO EN BASES DE DATOS

Se conoce como Descubrimiento de Conocimiento en Bases de Datos (KDD: Knowledge Discovery in Data Bases) al proceso de búsqueda de nuevo conocimiento a partir de datos almacenados en Bases de Datos. El proceso incluye el análisis inteligente de los datos y los pasos que permiten entregar conocimiento.

La minería de datos se refiere a las técnicas que permiten extraer conocimientos de un gran volumen de datos almacenados en Bases de Datos. En general el concepto Minería de Datos lo utilizan analistas de sistemas, estadísticos y especialistas en gestión informática, mientras KDD es usado por especialistas de aprendizaje en computadores e inteligencia artificial.

Fundamentalmente Data Mining utiliza técnicas para realizar predicción o descripción de eventos, los cuales

puedan obtenerse en forma automática a partir del gran volumen de datos almacenados en las Bases de Datos. La predicción involucra la utilización de atributos de una Base de Datos para predecir el comportamiento futuro de otros valores (conocer a priori como se presentarán datos desconocidos) y que son de interés para los usuarios, y la descripción se relaciona con encontrar patrones que permitan a los usuarios interpretar los valores de los atributos almacenados (descripción de datos y patrones).

Los objetivos inherentes a la predicción y descripción es posible lograrlos a través del Data Mining, para ello se aplican diversas técnicas, tales como reglas de asociación, árboles de decisión, clustering (agrupamiento), red neuronal, inteligencia artificial, etc.

3. ASPECTOS FILOSOFICOS DE LA CAUSALIDAD

Las generalizaciones habituales nos dicen cómo son las cosas: Las montañas son irregulares. Forman juicios descriptivos de la realidad. A diferencia de ellas, las generalizaciones causales nos explican el porqué de las cosas: ¿Por qué llueve? ¿Por qué hay sol? ¿Por qué hay guerras?

En los años de la antigua Grecia se comenzó a hablar y pensar filosóficamente respecto al concepto de causa. Aristóteles distinguió cuatro tipos de causas:

- 1) la causa material (la materia de la pintura: bastidores, tela, tinta, etc.)
- 2) la causa formal (es la idea representada por el artista en el cuadro)
- 3) la causa eficiente (es la acción del artista sobre la tela)
- 4) la causa final (es el fin para el cual se ha hecho la pintura).

Aristóteles [1] creó la conocida y popular doctrina aristotélica de las causas, que persistió en la cultura oficial de occidente hasta el Renacimiento.

En la ciencia moderna, las causas formales y finales fueron dejadas de lado por considerárselas fuera del alcance del proceso de experimentación científica. Lo cual significó llegar a la conclusión que en la moderna concepción del mundo, la materia es esencialmente el sujeto del cambio, no aquello de que una cosa está hecha y que persiste. "Por lo tanto, de las 4 causas de Aristóteles, sólo la causa eficiente es merecedora de investigación científica" [2].

Galileo Galilei encuentra una doble idea de causa, como sucesión temporal y como necesidad racional, para Hobbes (3), la causa será considerada como la suma de los antecedentes del efecto. Para Galileo la causa eficiente es la condición necesaria y suficiente para la aparición de

algo, aquella y no otra debe llamarse causa, a cuya presencia siempre sigue el efecto y a cuya eliminación el efecto desaparece. El problema de la definición de Galileo es que inutiliza el concepto de causa, pues entonces los análisis causales serían imposibles, debido a la infinidad de factores presuntamente integrantes de la causa y la prueba empírica de la hipótesis causal sería igualmente imposible, pues la supresión de cualquiera de los infinitos factores introducirá una diferencia, y por lo tanto, sería necesario llevar cuenta de una infinidad de parámetros.

El siglo XVII, recogiendo un proceso de crítica intelectual a las causas finales y materiales aristotélicas se pone fin a esta concepción. Para Descartes, la causa final deberá reservarse para el conocimiento de Dios. Respecto a la material, sostendrá que no podrá entenderse como causa de algo su propia materia [4].

Hume señaló en su "Tratado de la Naturaleza Humana" y en la "Investigación sobre el entendimiento humano", que un examen empírico de la causalidad muestra que no hay una conexión necesaria entre causa y efecto, sino una mera secuencia en el tiempo.

4. LA MATEMÁTICA, LA FILOSOFIA CAUSAL Y MINERIA CAUSAL

4.1. Causalidad y la Matemática

Entrando en materia podemos decir que a teoría de la inferencia causal une tres partes, dos partes de matemática y una parte de filosofía. Las dos partes matemáticas son grafos acíclicos dirigidos (DAGs) y la teoría de la probabilidad (enfocada en la independencia condicional de variables), y la parte filosofía involucra la causalidad entre variables.

Uno de los grandes problemas que presenta la búsqueda de relaciones causales en un universo de datos, guarda relación con la selección del subconjunto de datos que prometa corresponder a relaciones causales.

Los algoritmos actuales que son usados para el descubrimiento en datos observados frecuentemente usan correlación e independencia probabilísticas para encontrar posibles relaciones causales. Por ejemplo, si dos variables son estadísticamente independientes, se puede afirmar que entre ellas no hay relación causal.

Existen dos enfoques para el descubrimiento de causalidad en datos observados, estas son:

- Basadas en reglas (constrain-Based) y
- Enfoque basado en redes Bayesianas.

El descubrimiento causal en redes Bayesianas, consiste en encontrar la estructura causal más probable y los

parámetros correspondientes a la estructura. El primer punto de complejidad encontrado es el número de modelos posibles. Para solo tres variables existen veinticinco modelos posibles. Claramente, no sería viable enumerar todos los modelos posibles incluso para un conjunto pequeño de datos. Este problema, prácticamente hace inviable aplicar el algoritmo Bayesiano a datos observados.

En general, Los algoritmos que se aplican en la búsqueda de causalidad en datos observados son los algoritmos Basados en reglas (constrain-Based).

En este contexto es posible separar la teoría en tres partes: grafos dirigidos, la probabilidad, y la causalidad. La teoría conecta la estructura causal a la probabilidad. Para inferir desde un conjunto de datos estadísticos una estructura causal es necesario hacer una serie de supuestos.

Cabe hacer notar que existen muchas otras teorías que permiten extraer conclusiones causales para datos estadísticos, se ha escogido la Markoviana ya que el objetivo es explicar en algún sentido la relación entre la matemática, la filosofía y la minería causal.

Markov plantea un método para determinar causalidad en datos estadísticos, el cual se fundamenta en la determinación de dependencia probabilística. Explicando este método, se mostrara la relación entre la matemática la filosofía causal.

Un Dag es un objeto matemático totalmente abstracto. En la teoría de inferencia causal, los DAGs tienen dos funciones distintas. En la primera representan conjuntos de las distribuciones de probabilidad y en la segunda representan las estructuras causales. La manera en que representan las distribuciones de probabilidad es dada por la condición de Markov, en la cuál (el DAGs) se convierte en una relación gráfica más útil: d-separation (Pearl 1988) es una relación entre tres conjuntos disjuntos de vértices en un grafo dirigido. La idea básica de la d-separation supone verificar si un conjunto Z de vértices bloquea todas las conexiones (de cierto tipo) entre X e Y en un grafo G . Si es así, entonces X e Y son d-separados por Z en G , es decir, X e Y tienen cierto grado de independencia.

Por ejemplo, se tiene representado en un DAG tres vértices: $\{X_1, X_2, X_3\}$, y un conjunto de dos arcos entre estos vértices: $\{X_1 \rightarrow X_2, X_2 \rightarrow X_3\}$. Casi siempre se representan los DAGs con una figura, o diagrama de ruta, por ejemplo, este Dag se representa: $X_1 \rightarrow X_2 \rightarrow X_3$. En este ejemplo, al aplicar d-separación al DAG se tiene que: X_1 y X_3 son d-separados por X_2 . Luego, se puede suponer que en todas distribuciones que este DAG puede representar, X_1 es independiente de X_3 condicionado en X_2 . Usando la notación para la independencia presentada

por Phil Dawid (1979); $X_1 \perp\!\!\!\perp X_3 \mid X_2$ quiere decir: X_1 y X_3 son independientes condicionadas a X_2 .

Es importante destacar que mientras no se le dé una interpretación a los DAGs, estos son sólo objetos matemáticos que se pueden conectar a las distribuciones de probabilidad de cualquier manera que se desee. Lo importante es que de cualquier forma los DAGs permiten identificar conjuntos de independencias (conjunto de variables independientes). Cuando se da una interpretación causal (filosofía causal) a un DAG, entonces tenemos que la d-separación es la conexión entre un DAG causal y las distribuciones de probabilidad.

Normalmente hay muchos DAGs distintos que representan exactamente el mismo conjunto de relaciones de independencia, y por lo tanto, el mismo conjunto de distribuciones.

En este contexto se han desarrollado distintos algoritmos que permiten calcular d-separación para cualquier grafo, y que además son capaces de generar todos los DAGs posibles dado un conjunto específico (y conocido) de relaciones de independencia (Figura 4.1).

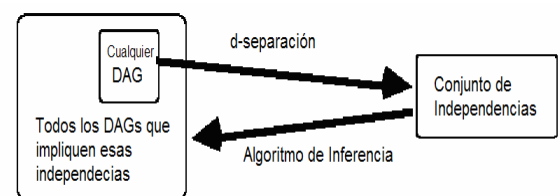


Figura 4.1: Dags generado a partir de un conjunto de relaciones de variables independientes.

Uno de los algoritmos es el denominado “algoritmo PC”. Su entrada es un conjunto de relaciones de independencia sobre un conjunto de variables y su resultado es un conjunto de DAGs sobre estas variables que son su equivalente d-separado, o Markoviano. Al aplicar el algoritmo PC sobre la relación de independencia “ X_1 y X_3 son independientes condicionadas a X_2 ”, se tiene que es posible generar otros dos DAGs que son d-separación equivalentes, esto se puede ver en la figura 4.2. El algoritmo PC genera todos los DAGs que son equivalente en d-separación.

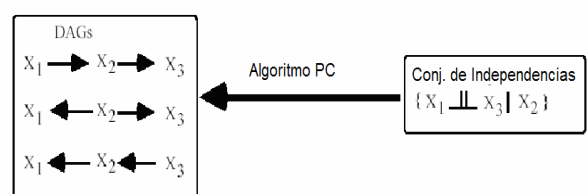


Figura 4.2: DAGs generados por algoritmo PC a partir de la relación de independencias “ X_1 y X_3 son independientes condicionadas a X_2 ”.

Si se considera solamente la d-separación y la condición de Markov como sólo la matemática que conecta a DAGs y las distribuciones de probabilidad y no se involucra la causalidad en nada. Entonces, se cuenta con una teoría matemática para generar representaciones de estructuras de independencia. Sin embargo, si se interpreta un Dag en forma causal (con la filosofía causal) la condición de Markov y la d-separación son de hecho la conexión correcta entre la estructura causal y la independencia probabilística. En este contexto a la interpretación causal de la condición de Markov se le denomina “condición causal de Markov”.

Los DAGs que son interpretados causalmente se denominan grafos causales. Por ejemplo, en un conjunto V de variables se tiene que S es una variable que representa el habito de fumar, la variable Y representa los dedos manchados por nicotina, y C la variable que representa la enfermedad de cáncer al pulmón, entonces el siguiente grafo causal (Figura 4.3.) representa lo que puede ser la estructura causal entre estas variables.

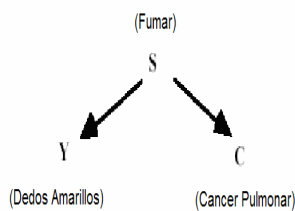


Figura 4.3: Dag causal “habito de fumar y cáncer al pulmón”

Los grafos causales se suponen completos en un sentido y no en otro. Son incompletos en el sentido de que no incluyen todas las causas de cada una de las variables en el sistema. Por ejemplo en el DAG anterior se han omitido: 1) muchas causas del cáncer al pulmón, por ejemplo, polución ambiental, inhalación de asbesto, factores genéticos, etc. 2) muchas variables que pueden estar entre una causa especificada y su efecto, por ejemplo, un problema bronquial podría estar entre el "Verdadero" camino causal de fumar a tener cáncer al pulmón. Sin embargo, estas omisiones no tienen relación con la característica de que el grafo anterior sea completo o incompleto.

Se dice que un grafo causal es completo en el sentido que cumple las siguientes dos características: Primero, deben estar presentes todas las causas comunes de las variables especificadas y segundo, deben estar representadas todas las relaciones causales entre las variables especificadas.

La teoría supone un grafo completo en el sentido que deben estar presentes (incluidas en el grafo) todas las causas comunes de las variables especificadas. Por

ejemplo, si se observa el grafo del ejemplo ilustrado en la figura 4.3. es posible detectar que falta la variable “factor genético” que es causa (común) de la conducta de fumar y de la enfermedad del cáncer. Como en el grafo no se ha representado esta causa común a las dos variables se puede afirmar que el grafo causal (figura 4.3) no es una representación exacta de la estructura causal entre estas tres variables, es decir, es un grafo incompleto.

El grafo causal también se supone estar completo en el sentido de que todas las relaciones causales entre las variables especificadas están representadas en el grafo. Por ejemplo, el grafo de la Figura 4.3. no tiene ningún arco dirigido de Y a S, si las manchas de nicotina en los dedos fueran causa de la conducta de fumar esta sería un fundamento para afirmar que el grafo es incompleto o no “exacto”.

4.2. La causalidad, la matemática y la minería de causalidad

En la búsqueda de relaciones causales es muy importante tener la posibilidad de manipular los datos, es decir, poder manejar las variables de modo que sea posible observar el comportamiento de una o más cuando se producen cambios en las otras variables.

La minería causal trata de generar métodos eficientes para descubrir las relaciones causales Bases de Datos observacionales. A la búsqueda de causalidad entre datos de una Base de Datos, se le denomina “Minería Causal o Descubrimiento Causal”.

Por otra parte, es importante la búsqueda de causalidad en datos observados ya que es posible encontrar en ellos causalidades que tal vez sería prácticamente imposible encontrarlas en datos experimentales. Por lo tanto hay un gran potencial para el uso de técnicas de búsqueda de causalidad en grandes Bases de Datos.

La minería causal corresponde a la etapa posterior a la Minería de Datos, ya que a través de esta se interpretan los datos generados por la Minería de Datos.

La teoría de la inferencia causal en Minería de Datos consta de dos partes : la inferencia causal en datos estadísticos (la que contempla la matemática, y la filosofía causal) y la parte de Minería de Datos que tiene que ver con técnicas que permitan seleccionar un conjunto de datos observados posibles de relacionarse causalmente. Todo esto en grandes volúmenes de datos almacenados en Base de Datos.

Las técnicas existentes de minería causal se basan en técnicas más complejas de descubrimiento causal en datos estadísticos y en datos experimentales. De esta manera a partir de la filosofía causal y la matemática es posible también inferir relaciones causales en grandes

Bases de Datos. Es necesario destacar que la búsqueda de causalidad en Bases de Datos es compleja, en el sentido que no es posible partir de una hipótesis, sin embargo, restringiéndose a un subconjunto de datos a través de la utilización de la técnica de minería de datos denominada “reglas de asociación” se puede obtener información de presunta causalidad entre ellos. En realidad, las reglas de asociación entregan información respecto a la frecuencia con que se relacionan dos o más atributos o Datos en la Base. En este contexto, es posible afirmar que los datos identificados por las reglas de asociación podrían corresponder a simples relaciones entre datos, por ejemplo, se puede encontrar a través de aplicación de reglas de asociación que la masa, la aceleración y la fuerza se asocian con alta frecuencia en el conjunto de datos, sin embargo, al analizar este hallazgo, se tiene que la masa y la aceleración no son causa de la fuerza, solo corresponden a una relación en que la fuerza es el resultado del producto masa-aceleración. También, podemos tener casos en que la asociación encontrada corresponde a una correlación y no a causalidad, por ejemplo, se puede tener que la cesantía aumenta y aumenta la delincuencia, recordemos que estadísticamente, se dice, que la correlación no necesariamente implica causalidad.

4.3.- Reglas de asociación

“El Descubrimiento de reglas de asociación, se define como el problema de encontrar todas las reglas de asociaciones existentes con un grado de confianza y soporte mayor que el especificado por el usuario, valores denominados minconf y minsup respectivamente” [5].

5. ALGORITMOS BASADOS EN ILCD

El planteamiento de Silverstein [6] para minar estructuras causales construidas sobre el algoritmo LCD[7] mantienen una complejidad de tiempo polinomial. Estos algoritmos no intentan descubrir la estructura causal completa como los algoritmos IC y PC. El algoritmo LCD encuentra estructuras causales en la forma de cadenas y ramificaciones. Sin añadir complejidad, se encuentra una estructura causal adicional; el autor nombra a esta estructura causalidad CCU. La causalidad CCU esta representada por las estructuras $a \rightarrow b \leftarrow c$, y también es conocida como las v-estructuras.

Los autores utilizan un método llamado soporte (soporte significa que un valor particular debe estar sobre una cierta frecuencia en los datos). Esto restringe la prueba de causalidad a los ítem de mayor interés. Un valor de umbral de confianza es utilizado como un nivel de confianza estadístico para determinar si dos variables son dependientes. La prueba estadística de chi-cuadrado (χ^2) es utilizada para determinar la dependencia de dos variables. El coeficiente β^2 es calculado como: $(O-E)/E$ donde O es el valor observado para la variable y E es el

valor esperado para la variable. Si β^2 es mayor que $\chi^2\alpha$ donde α es el nivel de confianza de la prueba, entonces las dos variables son dependientes. Utilizando estadística en combinación con soporte y confianza, el error en determinar dependencia es reducido.

Silverstein[6] no asume que en el caso $a \rightarrow b$, a es la causa directa. Esto permite que existan posibles variables ocultas. Otras suposiciones que serian deseables de eliminar son tomadas. Se asumen solo datos Booleanos sin datos perdidos. Por ejemplo este método puede ser utilizado en Datos Booleanos extraídos de la canasta de compras.

Al eliminar la representación grafica, este planteamiento es más viable en términos computacionales. Pero, no supera la dificultad de identificar estructuras con cierto grado de confianza. Últimamente estos métodos solo asumen cierta correlación estadística en una orientación particular entre las variables para decidir si una variable causa a otra. Usualmente, el orden temporal de las variables se supone conocido de antemano. Cuando el orden temporal es desconocido, Pearl [8] introduce la noción de tiempo estadístico, a cualquier orden de variables que coincidan con la estructura causal.

6. CONCLUSIONES

5.1. Ideas principales de la causalidad en Descubrimiento de Conocimiento sobre Bases de Datos.

La Minería de Datos cuenta con técnicas que permiten en alguna medida extraer objetos o información en los grandes volúmenes de datos, sin embargo, estos objetos no necesariamente representan conocimiento.

El procesamiento posterior de los resultados entregados por la Minería de Datos es una de las etapas del proceso de Descubrimiento de Conocimiento sobre Bases de Datos, y consiste en el procesamiento de los objetos de conocimiento producidos en etapas previas con el fin de simplificar, depurar, validar y visualizar el conocimiento extraído.

La transformación de estos objetos en conocimiento depende de la interpretación que realice el usuario. En muchos casos la obtención de conocimiento se desprende sin mayor análisis de los objetos, sin embargo, en otros casos requiere de un análisis exhaustivo y complicado, en estos casos es necesario contar con técnicas o herramientas de ayuda en el análisis posterior a la Minería de Datos.

Una de las técnicas utilizadas en Minería de Datos, las reglas de asociación, permite obtener patrones y relaciones entre los datos. El análisis posterior del resultado obtenido al aplicar la técnica de reglas de

asociación, consiste en un análisis semántico realizado en forma manual e íntegramente por el usuario, ya que la Minería de Datos no lo realiza. Esta forma de realizar este proceso de análisis es limitada y puede conducir a errores, ya que aunque puedan utilizarse herramientas estadísticas para algunos de los procesos que involucra dicho análisis, el usuario es siempre el que decide en el proceso de decisión final.

Las reglas de asociación indican la fuerza de la asociación de dos o más atributos de datos. El interés en las reglas de asociación es que ellas entregan la promesa (o ilusión) de causalidad, o al menos de relaciones predictivas. Sin embargo las reglas de asociación solo calculan la frecuencia de ocurrencias de uniones de dos o más atributos de datos; Ellas no expresan una relación causal. Si fuera posible descubrir las relaciones causales, esto sería muy útil para descubrir conocimiento. El objetivo de este trabajo ha sido explorar la causalidad en el contexto de la Minería de Datos.

El descubrimiento de causalidad en datos observados se apoya en técnicas que reúnen elementos matemáticos y filosóficos. Sin embargo, el problema no está resuelto. Los algoritmos para el descubrimiento en datos observacionales por lo general utilizan correlación e independencia probabilística. La aplicación de estos algoritmos solo permite afirmar que dos variables son estadísticamente independientes, y de ello se puede asegurar de que no existe causalidad asociada. Sin embargo, lo inverso no es necesariamente verdadero. Dada esta situación, el usuario de realizar un trabajo de análisis manual muy exhaustivo para poder determinar si existe causalidad entre las reglas de asociación generadas por la Minería de Datos. De otra forma, cuando se encuentran dos variables que están correlacionadas, ¿cómo se puede decidir que una causa a la otra?

5.2. Trabajo Futuro

En el desarrollo de este trabajo se han comentado ciertas materias que, si bien en algunos casos han quedado resueltas, pueden mejorarse, dejando abiertas varias líneas de investigación de cara al desarrollo de trabajos futuros. A continuación se explica brevemente una propuesta de investigación.

Adaptar y aplicar (probar) en una Base de Datos específica el algoritmo planteado por Silverstein (1998) "Scaleable Techniques For Mining Causal Structures" para encontrar estructuras causales en datos observados.

El algoritmo planteado por Silverstein (1998) para encontrar estructuras causales, utiliza la prueba estadística de chi-cuadrado (χ^2) para determinar la independencia de dos variables en el universo de datos observados. De esta forma obtiene el subconjunto de datos observados sobre los cuales aplica el algoritmo de búsqueda de causalidad entre los datos.

La implementación de esta adaptación se podría realizar aplicando la técnica de reglas de asociación sobre una Base de Datos determinada, seleccionando las reglas que supere un cierto umbral y con ellas montar un dígrafo para discutir cuáles de las reglas obtenidas podrían corresponder a relaciones causales y cuáles no a partir del algoritmo.

En este trabajo se debiera de hacer uso de factores de certeza en la búsqueda de reglas de asociación. Para ello aplicar (probar) el factor de certeza propuesto por el Dr. Daniel Sánchez Fernández [5]. El factor propuesto es capaz de detectar el grado de dependencia o independencia entre antecedente y consecuente de la regla. Es importante probar el factor de certeza propuesto por el Dr. Sánchez ya que normalmente los otros métodos no son capaces de detectar el grado de dependencia o independencia entre antecedente y consecuente de la regla.

Este trabajo permitiría evaluar la utilidad que podría obtener una organización en la búsqueda de causalidad en sus Bases de Datos, aplicando la técnica de reglas de asociación en la selección de datos a analizar con el algoritmo propuesto por Silverstein (1998) [6].

Referencias

- [1] Aristóteles, "La Metafísica".
- [2] Mario Bunge. Causalidad. Editorial Universitaria, Buenos Aires, 1978.
- [3] Hobbes, Thomas. 1996 (1651). Leviatán (México, Fondo de Cultura Económica).
- [4] Jean Wahl. Introducción a la filosofía. Fondo Cultura Económica, México, 1954.
- [5] Fernando Berzal, Ignacio Blanco, Daniel Sánchez. and María Amparo Vila, Measuring the accuracy and interest of association rules: A new framework, Department of Computer Science and Artificial Intelligence, University of Granada, E.T.S.I.I, March 2002
- [6] C. Silverstein, S. Brin, et al. [1998] "Scaleable Techniques For Mining Causal Structures," Proceedings. 1998 International Conference Very Large Data Bases, NY, 594-605
- [7] G. Cooper [1997] "A Simple Constraint-Based Algorithm for Efficiently Mining Observational For Causal Relationships" in Data Mining and Knowledge Discovery, v 1, n 2, 203-224
- [8] J. Pearl, J. [2000] Causality: Models, Reasoning, And Inference, Cambridge University Press, NY.
- [9] W. Frawley and G. Piatetsky-Shapiro and C. Matheus, Knowledge Discovery in DataBases: An Overview. AI Magazine, Fall 1992, pgs 213-228.