

Hacia la Minería de Caminos de Dependencia en Instituciones Electrónicas

Walter M. Grandinetti

Carlos I. Chesñevar

email: {wmg, cic}@cs.uns.edu.ar

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA)
Departamento de Ciencias e Ingeniería de la Computación (DCIC)
Universidad Nacional del Sur (UNS) – Bahía Blanca, Argentina

Resumen

Las instituciones son un entorno ampliamente utilizado por las sociedades para maximizar el beneficio de sus miembros. Se ha detectado además que el comportamiento de los miembros de las instituciones está fuertemente ligado a la historia misma de la institución. Sin embargo, este tipo de análisis es sumamente complejo por la cantidad de variables involucradas y la capacidad limitada para procesar la información existente. Actualmente, el concepto de institución está siendo aplicado a entornos virtuales donde coexisten agentes humanos y electrónicos (conocidas como instituciones electrónicas). Esto brinda la posibilidad de aplicar técnicas automáticas para el análisis del comportamiento de los agentes y reducir así la cantidad de información que necesita ser analizada por los expertos.

1. Introducción

Generalmente, se utiliza el concepto de institución para referirse a un grupo social con hábitos, usos, costumbres o normas que determinan la forma en que se pueden relacionar. Cada institución establece un orden social y su objetivo principal es maximizar el beneficio social. El uso de las instituciones provee las siguientes ventajas [10, 11], *a*) reducir la incertidumbre de la miembros de la sociedad al informar sobre las previsibles decisiones y acciones que tomarán los demás miembros, *b*) reducir el costo de obtener información y adoptar decisiones al reducir las posibilidades de elección del individuo o sugerir una vía de actuación, *c*) determinar el costo de transacción y transformación; y así precisar la viabilidad y potencialidad de una actividad económica. Las instituciones exitosas tienen la propiedad de evolucionar y llegar a un cierto grado de estabilidad y robustez [9]. El análisis de las instituciones ayudan a entender la forma en que evoluciona una sociedad, su estructura, mecanismos y normas.

La componente más importante de una institución es el conjunto de normas o reglas que auto limitan el ámbito de actuación de los individuos. Estas normas se dividen en formales e informales. Las primeras son leyes explícitas y las segundas, originadas en usos y costumbres, son usualmente tácitas y son importantes para explicar comportamientos sociales.

Los miembros de toda institución con su comportamiento son los productores de historia, la cual puede ser utilizada para encontrar *secuencias de comportamientos* sociales, políticos y económicos, y cambios producidos a través del tiempo. La historia afecta (y en algunos casos determina) el comportamiento futuro de los miembros. El concepto básico utilizado para la búsqueda de secuencias de comportamiento es conocido como *camino de dependencia* (en inglés, *path dependence*) o también como *causalidad histórica*.

Los caminos de dependencia buscan comprender los comportamientos y decisiones actuales trazando la evolución de las decisiones y comportamientos previos a través del tiempo. [1, 8, 10].

En [7] se definen las características de los análisis de caminos de dependencia, estableciendo 1) la fuerte relación de los *procesos causales* a los eventos más antiguos de la secuencia, 2) que la secuencia resultante es producto de *eventos contingentes*, 3) la “inercia” de los eventos, los cuales se mantienen en movimiento produciendo el mismo resultado.

Como resultado del análisis de caminos de dependencia se obtendrán secuencias de eventos determinadas causalmente, es decir, secuencias del estilo Z no podría ocurrir si no hubiesen ocurrido X , Y y W , que pueden clasificarse de acuerdo a como se relacionan los eventos entre sí en *autoreforzadas* (self-reinforcing) o *reactivas* (reactive). Otra clasificación, considera si los factores que determinan la secuencia de eventos son externos (exógenos) o internos. Las *secuencias auto-reforzadas* son cadenas de eventos ordenados temporalmente, caracterizadas por el hecho que los eventos iniciales inducirán a eventos posteriores similares de forma que con el tiempo se torna difícil o imposible efectuar eventos antagonistas. Esta inducción se efectúa mediante un refuerzo o recompensa creciente que fomenta que se repita el evento. Por otro lado, las *secuencias reactivas* son cadenas eventos ordenados temporalmente, los cuales se conectan entre sí de forma causal [7], es decir, que cada evento es una reacción de los eventos antecesores y causa de eventos subsecuentes.

2. Visión Generalizada del Proceso de Minería de Datos

En los años recientes, la capacidad para recolectar datos ha superado la capacidad de cualquier humano para procesarlos, siendo imposible el análisis y la comprensión de conjuntos masivos de datos por métodos manuales. Técnicas automáticas que puedan manejar dicha cantidad de datos son necesarias de forma que puedan ayudar a los expertos a tomar decisiones estratégicas.

La minería de datos se define como, *el proceso de extracción de información (patrones) no trivial, implícita, previamente desconocida y potencialmente útil de grandes repositorios de datos tales como bases de datos relacionales, data warehouses, etc* [2, 5]. El proceso de minería de datos consiste de tres elementos:

- Un *repositorio o fuente de datos* (en inglés, *data source*): Es el resultado de recolectar datos de diferentes orígenes, almacenarlos bajo un esquema unificado y realizando una limpieza de potenciales inconsistencias, ruidos y datos superfluos.
- Una *base de conocimiento* (en inglés, *knowledge base*): Consiste del conocimiento específico sobre el dominio de los datos. Es utilizada para guiar la búsqueda y determinar si la información recolectada es potencialmente útil (interesante).
- Un *modelo*: Se refiere al formato en el cual serán presentados los resultados del proceso de minería. Entre los tipos de patrones o modelos más utilizados son, *a) los árboles de decisión, b) los patrones frecuentes, c) las reglas de asociación, d) los patrones y reglas de asociación espaciales, e) los patrones y reglas de asociación temporales, f) los patrones y reglas de asociación secuenciales*, entre otros. Dependiendo de las clases de patrones que se estén buscando, diferentes técnicas podrán ser utilizadas [5]: *a) Descripción de Conceptos y Clases, b) Análisis de Asociaciones, c) Clasificación y Predicción, d) Clustering, e) Análisis Evolutivo y de Desviación.*

Dada una tupla $DM = \langle DS, KB, T, M \rangle$ donde DS representa un repositorio de datos, KB representa la base de conocimiento, T representa la técnica seleccionada, y M representa una clase de patrón o modelo; un *proceso de minería de datos* consiste de aplicar la técnica T sobre el conjunto DS guiado por KB , produciendo como resultado un conjunto de patrones de acuerdo a la representación M .

3. Instituciones Electrónicas como Repositorios de Datos

Como mencionamos en la sección 1, el uso de instituciones provee muchas ventajas para la sociedad. Es interesante notar que la noción de institución ha sido trasladada del contexto humano a entornos electrónicos para organizar sociedades de agentes independientes, heterogéneos y con intereses o metas propias. De esta forma se puede mantener la naturaleza abierta del sistema multi-agente pero estableciendo un ambiente normativo, permitiendo al igual que las instituciones humanas, reducir la incertidumbre acerca de los otros participantes, simplificando el proceso de toma de decisiones debido a la reducción de la cantidad de acciones permitidas en un determinado instante para cada agente [6].

Las instituciones electrónicas se especifican (diseñan) mediante los siguientes componentes, 1) un *conjunto de roles* y una *relación de orden parcial* entre ellos, 2) un *marco dialéctico* (dialogical framework), 3) un conjunto de *escenas* (scenes), 4) una *red de escenas* (performative structure) y 5) un conjunto de *normas*.

La característica más interesante de las instituciones electrónicas es que pueden ser ejecutadas y así *a)* crear instituciones en contextos virtuales con actores independientes y autónomos, o *b)* simular instituciones humanas en un contexto controlado, pero no por ello necesariamente menos complejo, con agentes que pueden estar guiados por humanos o con una mezcla de agentes independientes y delegados. De esta forma es posible recolectar más información de la que podría obtenerse en una institución humana, ya que las instituciones electrónicas podrían registrar todos los eventos, decisiones y caminos realizados por cada agente, como también brindar información contextual. Esto permite que pueda analizarse el desempeño de cada individuo, el comportamiento de la sociedad y potencialmente detectar caminos de dependencia.

Cada escena define un patrón de conversación entre múltiples roles. La ejecución de una escena comienza cuando un grupo de agentes intenta iniciar una conversación siguiendo las reglas especificadas por la escena. Por cada escena que se encuentre en ejecución se registra su estado actual y se dispone del contexto que especifica las acciones que pueden ser realizadas por los agentes para realizar un cambio de estado en la escena. El estado actual es mantenido en Σ_s y estará definido por, *a)* un conjunto de agentes *b)* el rol de cada agente en la escena *c)* las ligaduras (bindings) de variables *d)* nodo (estado) actual del grafo de estados, entre otros.

A medida que avanza la ejecución de cada escena, la información referente al estado es actualizada. En la definición actual de una institución electrónica no se prevee el registro histórico de esta información, sin embargo, como hemos mencionado en la sección 1, la historia del comportamiento social es vital para el análisis de las instituciones.

4. Minería de Instituciones Electrónicas

A fin de utilizar las técnicas de minería de datos como herramientas para el análisis de instituciones utilizaremos la información presente en la institución electrónica en forma de repositorio de datos y como base de conocimiento para realizar la minería, tal como se muestra en la figura 1. En particular, utilizaremos las reglas de asociación, patrones frecuentes, secuencias frecuentes y patrones emergentes, para detectar caminos de dependencia en las instituciones electrónicas.

Necesitamos por lo tanto registrar la información referente a los distintos estados por los que atraviesa cada escena en una bitácora de forma que puede ser utilizada como repositorio de datos para un futuro proceso de minería de datos. Los datos a ser registrados en la bitácora dependerán de que información se esperará encontrar por la minería de datos. En particular, se

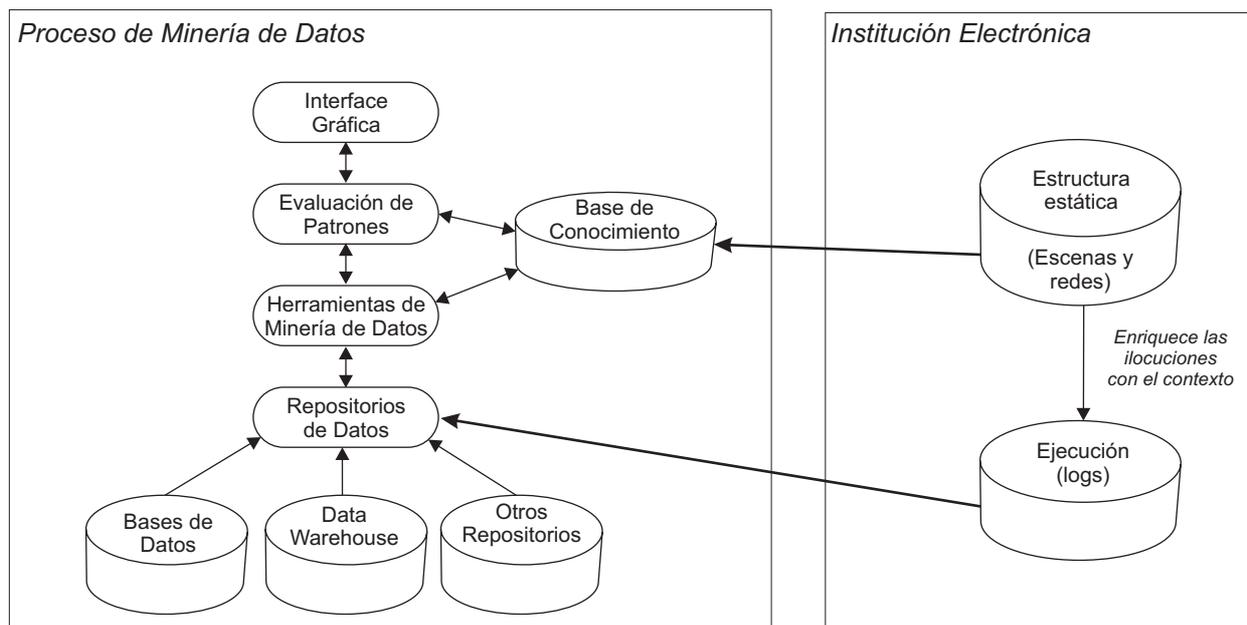


Figura 1: IE como Repositorio de Datos para el Proceso de Minería de Datos.

podría procesar la información presente en cada estado de forma de registrar simplemente que agente produjo el cambio de estado de la escena, mediante que acción y a que estado se arribó.

En particular, creemos que varias bases de datos transaccionales debieran ser generadas a partir de la bitácora. Por ejemplo, las transacciones podrían consistir de la ejecución completa de cada escena con todos sus estados. De esta forma se podrían analizar caminos de dependencia dentro de la escena de acuerdo al comportamiento de todos los agentes presentes. Sino, las transacciones podrían consistir de los mensajes (acciones) de cada individuo (agente) junto su rol. Así podrían encontrarse caminos de dependencia relacionado a cada rol o de la sociedad.

Con respecto a los tipos de patrones a buscar, creemos que dependerá de la premisa que se desee analizar en la institución. Sin embargo, ante la falta de dicha premisa, si simplemente se desea obtener información potencialmente útil, una minería especialmente de patrones secuenciales, reglas de asociación y patrones emergentes podría llegar a revelar datos interesantes. Mediante los patrones secuenciales podremos encontrar secuencias de eventos frecuentes que ayudarían a detectar caminos de dependencia (tanto autoreforzados como reactivos). Las reglas de asociación establecen causalidad y por tal motivo un análisis posterior de las secuencias frecuentes podría establecer si estamos ante caminos de dependencia reactivos o no. Por último, mencionamos los patrones emergentes [3] que detectan características contrastantes entre dos o más conjuntos de datos. Estos se relacionan a los factores internos que hacen que existan agentes con características únicas que generan caminos de dependencia vinculados a dichos factores.

5. Conclusiones

Las instituciones electrónicas poseen el potencial, hasta ahora no aprovechado, para realizar análisis de comportamiento sobre sociedades humanas o electrónicas. Es especialmente importante realizar este análisis en las instituciones para mejorar su eficiencia adaptativa [9]. Las técnicas de minería de datos han demostrado ser eficientes en este tipo de tareas.

Tal como mencionamos, la rápida difusión de la redes sociales electrónicas (ej. myspace.com,

secondlife.com, etc) probablemente ayudará a la instalación de instituciones electrónicas en contextos reales que ayuden a organizar a los agentes y disminuyan su incertidumbre acerca de los demás agentes, y en este contexto es altamente deseable tener herramientas automáticas para la detección de dichos fenómenos.

Además, las instituciones electrónicas pueden ser utilizadas como herramienta de simulación y pudiendo ser programados los agentes para que tengan un comportamiento similar a sus pares reales. Utilizando minería de datos en este contexto de simulación podrán efectuarse pruebas empíricas sobre datos reales para el testeo de hipótesis de trabajo. Por ejemplo, como se menciona en [4, pág.10], para determinar si un cierto evento es un potencial evento inicial de una secuencia reactiva o es necesario seguir analizando eventos previos.

Referencias

- [1] Bonn C. Mantzavinos, Max Planck Project Group, *Learning, change, and economic performance*, Washington University in St.Louis Syed Shariq (2000), 16.
- [2] Ming-Syan Chen, Jiawei Han, and Philip S. Yu, *Data mining: an overview from a database perspective*, Ieee Trans. On Knowledge And Data Engineering **8** (1996), 866–883.
- [3] Guozhu Dong and Jinyan Li, *Efficient mining of emerging patterns: Discovering trends and differences*, Knowledge Discovery and Data Mining, 1999, pp. 43–52.
- [4] Ian Greener, *Theorising path dependence: how does history come to matter in organisations and what can we do about it*, University of York, Department of Management Studies.
- [5] Jiawei Han and Micheline Kamber, *Data mining: Concepts and techniques*, 2 ed., The Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann Publishers, march 2006.
- [6] J.A.Rodriguez, *On design and construction of agent-mediated electronic institutions*, Ph.D. thesis, Instituto de Investigación en Inteligencia Artificial, 2002, Nro. 14.
- [7] James Mahoney, *Path dependency in historical sociology*, Theory and Society **29** (2000), 507–548.
- [8] C. Mantzavinos, *Individuals, institutions, and markets*, Cambridge and New York (2001).
- [9] Douglass North, *Understanding the process of economic change*, Journal of Evolutionary Economics **17** (1992), no. 3, 361–363.
- [10] Douglass C. North, *Institutions, institutional change, and economic performance*, Cambridge University Press, Cambridge, 1990, Pp. 152.
- [11] Thelen S. Steinmo and Longstreth, *Structuring politics: Historical institutionalism in comparative analysis*, Cambridge University Press (1992).
- [12] Qiankun Zhao and Sourav Bhowmick, *Sequential pattern mining: A survey*, Technical report, Nanyang Technological University, Singapore, 2003.