

Minería de Datos: Aplicaciones en el sector de las telecomunicaciones

Julián Álvarez Menéndez
Universidad Carlos III
NIA: 100084256

100084256@alumnos.uc3m.es

RESUMEN

A lo largo de este trabajo detallaremos cuál es la utilidad de la minería de datos en la industria de las telecomunicaciones actualmente. Del mismo modo, se incorporarán casos prácticos que nos permitirán un conocimiento más profundo y cercano de la aplicación de estas herramientas.

Términos generales

Algoritmos, Documentación, Economía, Fiabilidad, Seguridad.

Palabras clave

Minería, datos, telecomunicaciones, segmentación, predicción,

1. INTRODUCCIÓN

Uno de los rasgos que caracterizan al sector de las telecomunicaciones es la cantidad de información que generan y almacenan sus empresas. Esta enorme fuente de datos puede resultar imposible de abordar desde un punto de vista manual, por lo que es necesaria la utilización de algún tipo de análisis automatizado para obtener algún rendimiento.

Podemos clasificar la información almacenada en:

- Registros de llamadas.

Se generan cada vez que se produce una llamada en una red de telecomunicaciones, almacenándose en una base de datos asociada. Cada uno de estos registros debe guardar al menos el origen, destino, fecha, hora y duración de la comunicación.

- Información de la red.

Las redes actuales alcanzan un nivel de complejidad altísimo y la aparición de errores es detectada en casi todos sus componentes. Estos mensajes deben ser almacenados en una base de datos, indicándose el hardware o software del que proviene cada mensaje.

- Datos de los clientes.

Como en cualquier otra industria, resulta indispensable mantener un registro con la información de sus clientes y realizar el cobro de los servicios prestados.

Dadas las características de esta información y especialmente el gran tamaño de las bases de datos, la minería de datos resulta una

técnica muy adecuada para alcanzar conocimientos que mejoren el negocio desde diversos puntos de vista. Por ello, su entrada en la industria ha sido muy exitosa y su uso se ha extendido a casi todas las firmas involucradas en el negocio.

2. OBJETIVOS

Como en cualquier otra industria, el objetivo principal de las compañías del sector de las telecomunicaciones es la mejora de sus resultados empresariales.

Sin embargo, esta mejora no tiene por qué significar un aumento de clientes o de tarifas, sino que hay que entender cada negocio como un gran sistema donde múltiples factores interactúan. Así, cambios que impliquen un mejor trato hacia los clientes o un aprovechamiento más eficiente de los recursos pueden resultar muy provechosos pese a no ser evidente la forma de conseguirlo.

Ante un entorno tan complejo y con una información disponible de tal magnitud, la minería de datos constituye una herramienta indispensable para conseguir un conocimiento más profundo del sistema y descubrir esas claves ocultas que pueden marcar la diferencia.

Podemos dividir el uso de la minería de datos en las empresas de telecomunicaciones en tres ámbitos fundamentales: marketing, detección de fraudes y control de calidad.

2.1 Marketing y perfil del cliente.

Gracias a los registros de llamadas, las compañías poseen mucha información de sus clientes. En ellos se encuentran numerosos datos que describen su comportamiento (duración de las llamadas, destino de las mismas, momento en el que las realiza...) y pueden ser de gran utilidad para la empresa.

Sin embargo, el tamaño de estas bases de datos y todos los parámetros a analizar hacen que sea necesaria una segmentación de los mismos para poder obtener una conclusión. De esta tarea se encargará la minería de datos, siguiendo los métodos más adecuados para cada situación.

Este conocimiento se puede aplicar para numerosas finalidades. Entre ellas está la correcta planificación de las campañas de marketing emprendidas por la compañía. Hoy en día, éstas se realizan de una forma muy personalizada dirigidas hacia un sector concreto del mercado. Si conseguimos segmentar a nuestra clientela en grupos con características similares, será mucho más sencillo emprender una campaña de promoción al conocer cómo es éste segmento y qué es lo que busca.

Del mismo modo, el conocimiento exhaustivo de la clientela ayuda a detectar nuevas oportunidades de negocio, al conocer mejor los hábitos e intereses de los consumidores.

Otra de las utilidades de este profundo estudio de la clientela es su fidelización. Las empresas del sector están especialmente sensibilizadas con la pérdida de clientes que escogen una compañía de la competencia, puesto que es un proceso fácil y está demostrado que el coste de conseguir un cliente nuevo es sustancialmente más costoso que mantener al antiguo.

Ante esto, la minería de datos puede predecir qué clientes son más susceptibles de irse a la competencia basándose en las bases de datos anteriores. De esta forma, la compañía podrá anticiparse y ofrecer a estos clientes descontentos alguna opción para mejorar su satisfacción. La predicción supone sin duda una de las armas más importantes de la minería de datos, puesto que da a las empresas una ventaja competitiva sobre sus rivales.

2.2 Detección de fraudes

El fraude es un problema importante para las compañías de telecomunicaciones, puesto que les hace prestar un servicio sin obtener contraprestación económica alguna. Este fraude puede producirse por diferentes motivos.

Por un lado tenemos el fraude suscriptivo, donde el cliente contrata el uso de una línea pero no paga sus facturas al día. Su detección es bastante simple consultando la base de datos de los clientes. Sin embargo, el adecuado uso de la minería de datos puede elaborar modelos que predigan qué clientes son más susceptibles de dejar de pagar sus facturas.

Por otro lado, tenemos el fraude superimpuesto, donde una persona no registrada accede a la línea de un cliente para hacer un uso fraudulento de ella. Su detección es bastante más complicada, pero posible utilizando la minería de datos para analizar cambios en los comportamientos de los clientes. El tiempo que tardemos en detectarlo será un aspecto clave muy a considerar a la hora de implantar nuestro sistema.

2.3 Control de calidad

Dada la complejidad de las redes desplegadas actualmente, a la compañía le llegan continuamente informes de alarma y de estado de sus componentes. Para poder dar a sus clientes un servicio adecuado, el operador necesita ser capaz de solucionar cualquier error en el menos tiempo posible.

Ante esto, la minería de datos puede ayudar a interpretar esta información permitiendo al operador cuándo y dónde se localiza un problema. Del mismo modo, se deben implementar mecanismos que den respuesta a estos problemas en tiempo real de forma automatizada.

Utilizando los datos recopilados, también se podrían realizar modelos que predigan la aparición de estos problemas, para que la respuesta sea más eficiente y rápida.

3. APLICACIÓN

Una vez comentados los diferentes ámbitos en los que la minería de datos resulta de gran utilidad para las empresas de telecomunicaciones, el paso siguiente será estudiar cómo se llevan a cabo esas aplicaciones.

El esquema general será:

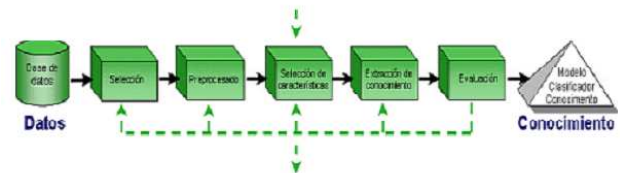


Figura 1. Fases del proyecto de minería de datos.

El primer paso será obtener la información necesaria de las bases de datos de la compañía e introducirla en un almacén de datos donde se pueda trabajar de forma operativa con esos datos provenientes de fuentes muy diversas. La implantación de este almacén debe ser correcta para que se pueda acceder a los datos antes de su procesado.

Seguidamente pasaremos a transformar estos datos mediante diversas técnicas adaptables a cada caso, tal que se puedan solventar problemas como el ruido, valores incompletos...

A continuación deberíamos aplicar un método de minería de datos adecuado, así como los algoritmos a utilizar, adaptando previamente la información a un formato que el método seleccionado pueda tratar. Tras esto debemos especificar un criterio (mediante su función de clasificación, forma de representación, estrategia de búsqueda...) para obtener un determinado modelo de conocimiento.

Tras esto deberemos interpretar y evaluar los resultados obtenidos, volviendo a realizar algunos de los pasos anteriores si es que fuese necesario.

Una vez conocida la base teórica, analizaremos tres casos prácticos donde la aplicación de herramientas de minería de datos significó una notable ventaja competitiva para las compañías.

3.1 Caso práctico: perfil de clientes

3.1.1 Planteamiento inicial

En este primer caso práctico estudiaremos el análisis de clientela realizado por Vodafone en los Países Bajos.

Al tratarse de una gran empresa, acumulaba grandes cantidades de información en su base de datos que era incapaz de manejar y que deseaba utilizar para tomar decisiones operacionales y estratégicas.

Para ello, deseaba hacer una segmentación y catalogación de su clientela. La segmentación consiste en dividir a sus clientes por sus atributos comunes (hábitos, gustos...), mientras que la catalogación los divide por sus atributos (edad, ingresos...). Siendo capaces de obtener la relación entre ambos aspectos,

facilitaría enormemente la labor de la compañía a la hora de fijar sus objetivos y estrategias.

Para realizar la segmentación, se aplicarán varias técnicas de clustering. Para encontrar la relación entre segmentación y catalogación se utilizará una técnica de minería de datos denominada Máquinas de Vectores de Soporte.

3.1.2 Desarrollo

El primer paso será comprender la información que almacenamos en la base de datos.

Para ello, será interesante comprender la estructura a la que están sujetos los clientes en Vodafone.

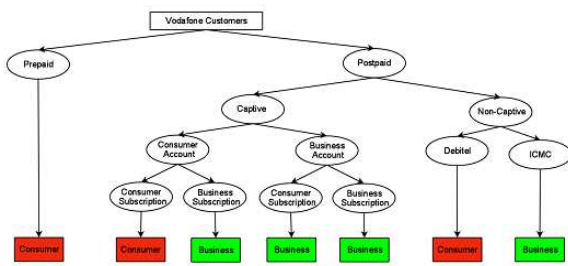


Figura 2: Estructura de la clientela de Vodafone

Cada llamada que se produce es almacenada en un registro de llamadas, con más de 20 millones de nuevas entradas cada día.

Dado que el origen, destino, fecha y duración de la llamada queda registrada, podremos analizar muchos aspectos de cada cliente a partir de este registro. Entre todos ellos, nos quedaremos con estos 12:

1. Duración media de la llamada.
2. Media de llamadas recibidas por día
3. Media de llamadas originadas por día.
4. % de llamadas diurnas
5. % de llamadas entre semana.
6. % de llamadas a móviles.
7. Media de mensajes recibidos por día
8. Media de mensajes originados por día.
9. % de llamadas internacionales.
10. % de llamadas al mismo operador.
11. Nº llamadas a códigos de área única.
12. Nº llamadas a diferentes números.

Estas características pueden usarse para construir los segmentos.

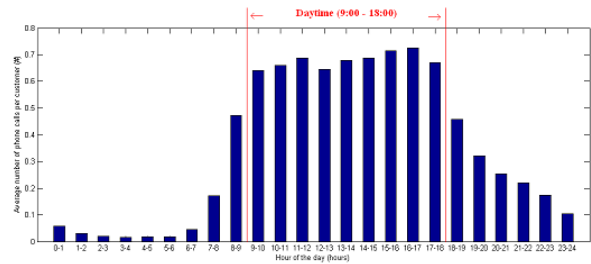


Figura 3: Estudio de las llamadas por hora.

A la hora de catalogar los registros, utilizaremos la información extraída de los registros de Vodafone:

- Grupo de edad (<25, 25-40, 40-55, >55)
- Género.
- Tipo de teléfono.
- Tipo de suscripción.
- Tamaño de la compañía.
- Área donde vive.

La agrupación de algunas variables (como la edad) responde a la necesidad de hacer el problema más manejable. Es importante que el tamaño de los grupos sea lo más parecido posible.

A continuación deberemos preparar los datos para que puedan usarse en el proceso de minería de datos. Para ello debemos chequear valores anormales, cambiar campos vacíos por valores por defecto, mapear valores continuos en rangos, normalizar variables y convertir formatos.

Seguidamente pasaremos a realizar el clustering, un método de aprendizaje no supervisado que organiza los objetos en grupos.

Tras realizar múltiples pruebas con los diferentes algoritmos de clustering posibles, se comprueba que Gath Geva con 4 grupos y Gustafson-Kessel con 6 grupos son los más adecuados y carecemos de medios para decantarnos por uno de los dos.

Gustafson-Kessel es una variación del algoritmo Fuzzy c-means, en el que se usa una norma de distancia adaptativa diferente para reconocer las formas de los datos.

$$D_{ikA}^2 = (x_k - v_i)^T \cdot A_i (x_k - v_i), \text{ where } 1 \leq i \leq c \text{ and } 1 \leq k \leq N.$$

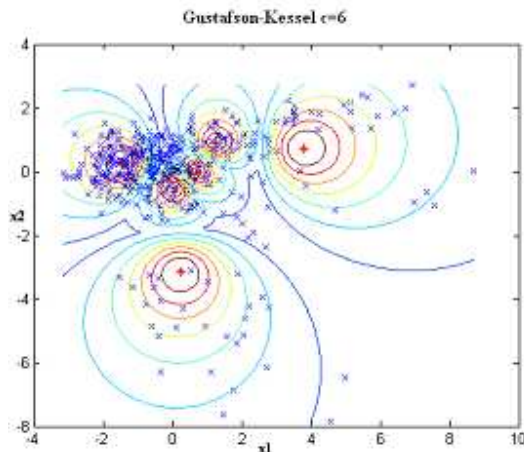


Figura 4: Resultado de Gustafsson-Kessel con mapeado de Sammon para 6 segmentos

Los 6 segmentos serían:

1. Clientes con relativamente pocas llamadas. Duración media menor que la media. Nivel de mensajes alto. No llaman a muchos números diferentes.
2. Clientes con número relativamente alto de contactos. Llamam a muchas áreas diferentes. Tienen más contactos con Vodafone.
3. Clientes con relativamente muchas llamadas. Bajo uso de mensajes. Hacen más llamadas internacionales que otros clientes.
4. Clientes medios. Sin valores altos ni bajos.
5. Clientes que no reciben muchas llamadas. Media de duración baja. Reciben y originan pocos mensajes.
6. Clientes que originan y reciben muchas llamadas. Reciben y mandan muchos mensajes. Duración de las llamadas larga. Porcentaje de llamadas internacionales alto.

Gath Geva tendrá una norma de distancia con un término exponencial, por lo que la norma decrecerá más rápido.

$$D_{ik}(x_k, v_i) = \left(\frac{\sqrt{\det(F_{wi})}}{\alpha_i} \right) \left(\frac{1}{2} (x_k - v_i^{(i)})^T F_{wi}^{-1} (x_k - v_i^{(i)}) \right)$$

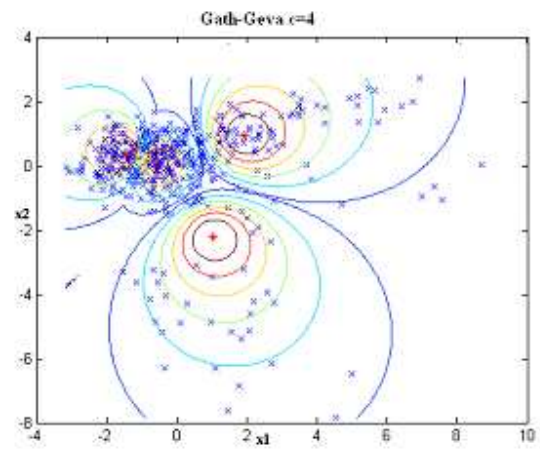


Figura 5: Resultado de Gath Geva con mapeado de Sammon para 4 segmentos

Los 4 segmentos serán.

1. Clientes con un número bajo de llamadas. Llamam más en la noche y a líneas fijas que otros. Su uso de mensajes es mayor que el medio. Su número de llamadas internacionales es bajo.
2. Clientes con un número medio de llamadas. A menudo llaman a móviles durante el día. No mandan ni reciben muchos mensajes.
3. Clientes que hacen relativamente muchas llamadas. Llamam a muchos números diferentes y tienen muchos contactos clientes de Vodafone.
4. Clientes con muchas llamadas. Mandan y reciben muchos mensajes. A menudo llaman durante el día y llaman más que la media a números internacionales. A duración de las llamadas es larga. Tienen un número bajo de contactos a los que llaman.

Seguidamente aplicaremos el algoritmo de Máquinas de Vectores de Soporte, que aprende por ejemplificación para asignar etiquetas a objetos. Lo utilizaremos para reconocer el segmento de un cliente examinando miles de consumidores de cada segmento.

Para ajustar los parámetros aplicaremos el algoritmo con cada una de las distintas funciones, testeando los resultados a través de la matriz de confusión.

Tanto para el caso de 4 segmentos como para el de 6, comprobamos que la función radial es la mejor opción. Sorprendentemente los resultados son muy similares en ambos (80.3% y 78.5% respectivamente).

Si analizamos las matrices de confusión, vemos que hay 2 grupos (clusters) que pueden clasificarse fácilmente con 1 perfil del cliente.

Tabla 1: Matriz de confusión para 4 segmentos

Predicted → Actual ↓	Segment 1	Segment 2	Segment 3	Segment 4
Segment 1	97.1%	0.5%	1.9%	0.5%
Segment 2	3.6%	76.6%	7.8%	12.0%
Segment 3	2.2%	0.8%	96.3%	0.7%
Segment 4	7.1%	13.0%	6.9%	73.0%

Tabla 2: Matriz de confusión para 6 segmentos

Predicted → Actual ↓	Segm. 1	Segm. 2	Segm. 3	Segm. 4	Segm. 5	Segm. 6
Segment 1	74.1%	1.1%	10.1%	8.4%	0.6%	5.7%
Segment 2	0.2%	94.5%	0.6%	1.4%	1.2%	2.1%
Segment 3	5.6%	4.7%	71.2%	9.1%	2.1%	7.3%
Segment 4	12.3%	4.1%	3.9%	68.9%	6.8%	4.0%
Segment 5	2.0%	0.6%	0.7%	1.3%	92.6%	2.8%
Segment 6	12.5%	2.4%	3.7%	10.4%	1.3%	69.7%

Ahora podemos medir la importancia de cada característica. Esto se hará dejando una de ellas fuera del vector y entrenando el algoritmo anterior sin ella.

Así, comprobamos que la edad es una importante característica mientras que el teléfono no lo es.

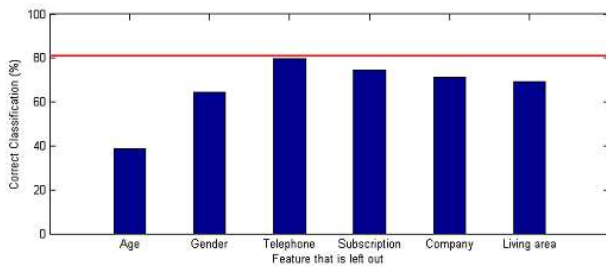


Figura 6: Resultados dejando una característica fuera con 4 segmentos

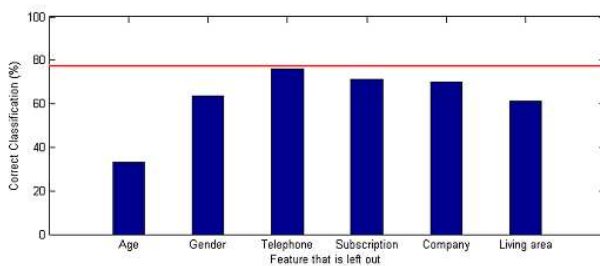


Figura 7: Resultados dejando una característica fuera con 6 segmentos

3.1.3 Resultados

Una vez concluido el estudio, analizaremos los resultados.

A la hora de aplicar el clustering, fue imposible dictaminar cuál de los 2 algoritmos mejores era óptimo. Además, resulta difícil comparar ambos al tener distinto número de segmentos.

Una vez aplicado el algoritmo de Máquinas de Vectores de Soporte, el porcentaje de segmentos correctamente clasificados no es tan alto como esperábamos.

Una posible explicación es que las características del cliente no sean suficientes para realizar con correcto perfil. Probablemente podría mejorarse usando información sobre el estilo de vida, hábitos, ingresos...

También podría ser explicable ya que un teléfono no es siempre usado por la misma persona aunque sólo ella aparezca en los datos del operador.

Finalmente, pudimos comprobar que la edad del cliente es un factor muy a tener en cuenta.

A la vista de estos resultados, podríamos recomendar a la empresa que aumente sus datos almacenados sobre cada cliente y que intente conseguir los datos del usuario en lugar del propietario de cada línea si quiere obtener una mayor precisión en mayores estudios.

Sin embargo, los resultados alcanzados aquí, seguro le servirán para poder focalizar sus campañas de marketing y poseer un mayor conocimiento de sus clientes.

3.2 Caso práctico: detección de fraude

3.2.1 Planteamiento inicial

En este ejemplo práctico, estudiaremos el caso de Brasil Telecom, que implementó un sistema para la detección de fraude.

Está basado en 2 modelos de minería de datos. Por un lado un modelo de segmentación basado en el comportamiento del cliente y por otro un modelo de clasificación basado en una red neuronal. El primer modelo analiza cómo los clientes se van haciendo insolventes mientras que el segundo realiza una predicción de los riesgos de insolvencia en tiempo real.

3.2.2 Desarrollo

El primer problema que debemos afrontar es la elección de las variables. De un total de 152 variables disponibles en la base de datos de la compañía, 47 fueron seleccionadas para el estudio.

El siguiente paso será desarrollar un modelo no supervisado para detectar las características más significativas de los clientes insolventes.

La norma escogida para definir un cliente como insolvente será que tenga un retraso en el pago mayor de 3 días.

Tras esto, ya se puede hacer el modelo no supervisado de desarrollo usando los mapas de Kohonen.

Una vez realizado el clustering 5 grupos distintivos fueron identificados, de acuerdo con los diferentes perfiles de insolvencia:

1. Clientes etiquetados como moderados, con facturas medias, insolvencia media, altos periodo de retraso, volumen medio de deudas y trafico usado medio.

2. Clientes etiquetados como malos, con facturas medias, alta insolvencia, retrasos medios, volumen medio de deudas y tráfico usado medio.
3. Clientes etiquetados como buenos, con facturas bajas, insolvencia media, retrasos medios, medio volumen de deudas y bajo uso del tráfico.
4. Clientes etiquetados como muy malos, con facturas altas, alta insolvencia, retrasos medios, medio volumen de deudas y tráfico usado medio.
5. Clientes etiquetados como ofensivos, con facturas altas, insolvencia alta, retrasos altos, alto volumen de deudas y muy alto tráfico usado.

Dado que la empresa también requiere predecir la insolvencia en tiempo real, debemos hacer un modelo de clasificación de estos eventos.

Este modelo de predicción se construye utilizando toda la base de clientes (no sólo los insolventes).

Clasificaremos como malo (B) si su retraso en el pago es superior a 29, considerando al resto como buenos clientes (G).

Seguidamente se dividen los clientes en 10 grupos utilizando los mapas de Kohonen, que usamos para crear los clasificadores basados en redes neuronales.

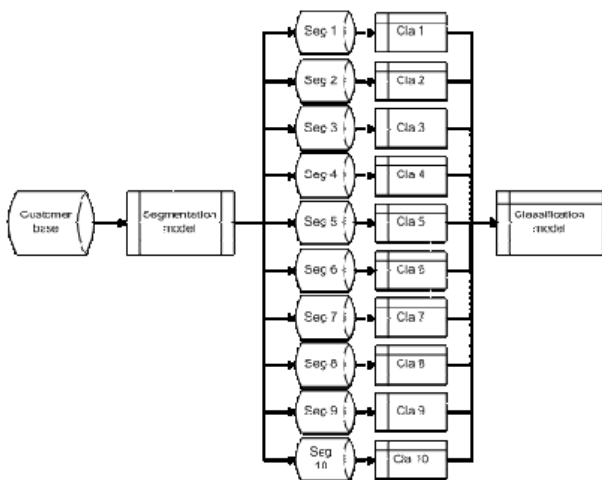


Figura 9: Clasificador basado en redes neuronales.

Tras optimizar los parámetros de cada sub-clasificador, obtenemos los siguientes resultados:

Tabla 3:

Group	# records	Sensitivity (%)	
		Good	Bad
G1	51,173	84.96	96.45
G2	22,051	82.89	96.89
G3	17,116	85.78	92.14
G4	23,276	83.87	89.04
G5	171,789	81.59	88.36
G6	9,902	80.92	88.57
G7	23,827	82.09	89.92
G8	48,068	86.44	97.08
G9	19,194	88.69	91.32
G10	14,868	87.23	93.98
Average		84.45	92.38

Como se ha hecho el modelo clasificador como 10 clasificadores, cada uno usando un segmento, es posible adecuar cada modelo de forma individual, con lo que se consiguen sistemas predictores más fiables.

Una vez implementados los 2 modelos, ya podremos obtener resultados que serán de utilidad para la compañía.

3.2.3 Resultados

Tras observar estos resultados, Brasil Telecom cambió alguna de sus políticas.

Como ejemplo, solía mandar un aviso por carta a los insolventes a los 15 días, pero hasta los 30 días no podía cortar el suministro. Tras aplicar el modelo, se observó que al menos el 85% de los clientes pagarían antes de alcanzar los 30 días por lo que todas las cartas destinadas a ellos era un gasto innecesario.

Del mismo modo, la empresa pudo disminuir su presión fiscal. Dado que según la legislación brasileña las compañías deben pagar en impuestos el 25% de la factura (sea esta abonada o no por el cliente), la no aceptación de clientes potencialmente insolventes con este sistema supuso unas ganancias de varios millones de dólares con una tasa de error inferior al 5%.

3.3 Caso práctico: control de calidad

3.3.1 Planteamiento inicial

Actualmente las complejas redes de telecomunicaciones producen grandes cantidades de alarmas a diario. Sin embargo, analizarlas no es fácil ya que la información se encuentra escondida o fragmentada.

Esto es debido a que las alarmas pueden ser remotas implicaciones de fallos y que a menudo sólo se pueden analizar en el contexto de otras alarmas y otro conocimiento. Además, el hecho del continuo crecimiento y evolución de las redes y las alarmas dificulta aún más las cosas.

Para tratar de mitigar este problema se creó TASA, un sistema que descubre regularidades en las alarmas y que puede localizar episodios frecuentes de alarmas presentándolos como normas.

3.3.2 Desarrollo

En los últimos años la búsqueda automatizada de regularidades en grandes bases de datos se ha convertido en un campo de gran interés.

Se aplican técnicas de minería de datos y descubrimiento del conocimiento para encontrar regularidades en las alarmas, por ejemplo cómo las alarmas ocurren juntas.

El proceso puede dividirse en las siguientes tareas:



Figura 10: proceso de descubrimiento del conocimiento

TASA se basa en esta búsqueda para proporcionar a los operadores información nueva sobre el funcionamiento de sus redes. Soporta 2 fases centrales del proceso de descubrimiento de conocimientos.

En primer lugar analizamos la fase del descubrimiento de patrones. En ella se utilizan algoritmos que describen asociaciones entre alarmas y asociaciones entre atributos de alarmas para localizar episodios de alarma frecuentes.

TASA descubre las llamadas reglas de episodios, afirmando que pronto ocurrirá una determinada alarma si una determinada combinación de alarmas ocurre previamente.

Por tanto, se puede decir que durante esta fase TASA busca automáticamente reglas de episodio y reglas de asociación,

Así mismo, el sistema realiza mediciones estadísticas para conseguir una visión general de las alarmas.

Estos conocimientos son fáciles de entender por lo que permite hacerse a una idea de las alarmas y pueden usarse en aplicaciones de vigilancia de redes.

Esta fase de descubrimiento es automática y sólo depende de unos pocos parámetros introducidos por el usuario. Así, el usuario podrá definir qué órdenes especiales son permitidas en los episodios, qué tipo de identificación de alarma se usa, los períodos de tiempo considerados en las reglas y la frecuencia con la que las reglas deben estar en la secuencia.

Para el operador muchos de los fenómenos analizados aquí serán triviales. En la segunda fase sin embargo, conocida como fase de presentación, el papel del experto ya será vital.

En ella la exploración de largos conjuntos de reglas es soportada por simples pero poderosas herramientas de poda (seleccionar o rechazar reglas basándose en sus propiedades), ordenación (clasificar las reglas según varios criterios) y agrupamiento (segmentación de las reglas en grupos de reglas con efectos similares).

Entre ellas es especialmente poderosa la utilización de plantillas para fijar el rechazo de un conjunto de reglas.

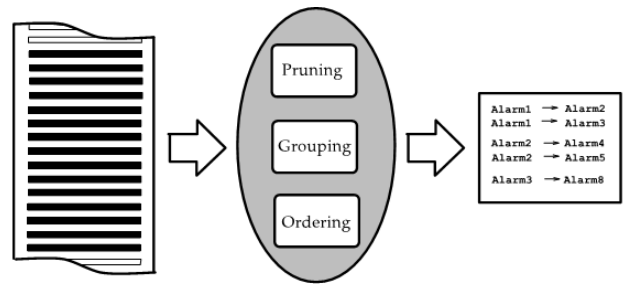


Figura 11 : Herramientas de exploración

3.3.3 Resultados

Como hemos podido observar, TASA resulta una herramienta muy útil para comprender el funcionamiento de la red o indicar configuraciones erróneas.

De esta forma, constituye una herramienta de gran utilidad para la localización de fallos en las redes de telecomunicaciones y su solución en el menor tiempo posible. Del mismo modo, también puede ser utilizada por sistemas expertos para la correlación de alarmas y la predicción de fallos.

4. CONCLUSIÓN

En un mundo en el que las empresas se encuentran con enormes volúmenes de información almacenada, las compañías sienten la necesidad de sacar el mayor partido posible a estos datos.

Ante esto, técnicas como la minería de datos resultan de un valor incuestionable por flexibilidad, potencia y eficiencia.

En un mundo globalizado en el que las diferencias geográficas empiezan a ser inapreciables, herramientas como el conocimiento son las que marcan la diferencia y constituyen una auténtica ventaja competitiva para aquellas empresas capaces de sacarle partido.

El sector de las telecomunicaciones no es ajeno a este fenómeno y el uso de la minería de datos ya tiene una gran penetración en el sistema.

A lo largo de los tres ejemplos prácticos detallados hemos comprobado cómo tres problemas completamente distintos pueden abordarse desde la misma perspectiva.

Por un lado, una correcta segmentación de la clientela permitiría agruparla de acuerdo a unas características concretas y enfocar de una manera completamente distinta cualquier decisión estratégica o de marketing.

Por otro lado, vimos cómo la segmentación unida a un modelo de predicción era capaz de disminuir notablemente la morosidad en una compañía, suponiendo un gran ahorro de costes.

Por último, la gestión de un aspecto tan complejo como los avisos de alarma en una compleja red podía facilitarse notablemente mediante las técnicas de minería de datos.

El hecho de que sea tan complicado conseguir ejemplos prácticos detallados indica que son un bien valioso para las empresas, que no desean compartir con la competencia.

Por tanto, es esperable que este tipo de técnicas adquieran un mayor peso si cabe en la toma de decisiones empresariales y se consolide como uno de los aspectos clave a la hora de mejorar la situación competitiva de un compañía.

5. REFERENCIAS

- [1] Gary M. Weiss. Data Mining in Telecommunications. Fordham University
- [2] DAEDALUS. Minería de Datos. Nov 2002.
- [3] Sofía J. Vallejos. Minería de Datos. 2006. Universidad Nacional del Nordeste.
- [4] S.M.H. Hansen. Customer Segmentation and Customer Profiling for a Mobile Telecommunications Company Based on Usage Behavior, July 2007
- [5] Pinheiro, C.; Evsukoff, A.; Ebecken, N. Revenue Recovering with Insolvency Prevention on a Brazilian Telecom Operator. SIGKDD Explorations. Volume 8 Issue 1
- [6] Hatonen, K.; Klemettinen, M.; Mannila, H.; Ronkainen, P.; Toivonen, H. TASA: Telecommunication Alarm Sequence Analyzer. IEEE/IFIP 1996 Network Operations and Management Symposium. Kyoto, Japan. April 1996.
- [7] Fayyad, U.; Piatetsky, G.; Padhraic, S. From Data Mining to Knowledge Discovery in Databases. AI Magazine. Fall 1996.
- [8] Frawley, W.; Piatetsky, G.; Matheus, C.; Knowledge Discovery in Databases: an Overview. AI Magazine. Fall 1992.
- [9] Apte, C.; Liu, B.; Pednault, E.; Smyth, P. Business Applications of Data Mining. May 2002
- [10] Acosta, M. Minería de Datos y Descubrimiento de Conocimiento. Universidad de La Habana.