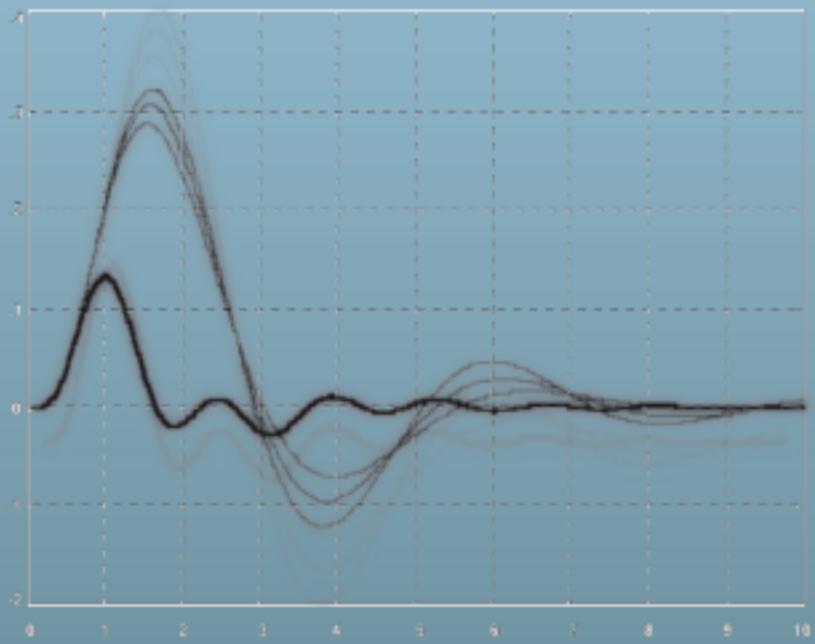
ISSN 1409-2441
www.ucr.ac.cr

Ingeniería

Revista de la Universidad de Costa Rica
ENERO/DICIEMBRE 2002 - VOLUMEN 12 - N°1 y 2



REVISTA INGENIERÍA Volumen 12 Enero / Diciembre 2002 Números 1 y 2

Ingeniería 12 (1,2): 37-45, 2 002 San José, Costa Rica

UN MODELO DE DETECCIÓN DE FRAUDES EN DATOS DE SEGUROS

José Ronald Argüello Venegas
Elzbieta Malinowski Gajda

Resumen

La detección de fraudes en diferentes sectores del quehacer humano cada vez más se convierte en una tarea donde se puede utilizar sofisticadas herramientas computacionales para facilitar y agilizar el proceso investigativo. En el presente artículo, se describen las principales características que envuelve el problema de detección de fraudes en datos de seguros de automóviles, específicamente reclamos de automóviles. Además, se discute como organizar un sistema para descubrir conocimiento en esta área, los diferentes procesos que deben ser realizados, las peculiaridades que éstos deben tener y la interrelación que existe entre ellos. El modelo propuesto presenta también como la utilización de estos procesos permite eventualmente llevar a un modelo de descubrir reglas iterativamente. Este proceso iterativo sirve para afinar los criterios de detección de fraudes hasta que de manera ideal el sistema solo prácticamente dictaminaría si un caso o no clasifica como fraude. Finalmente, se discuten las estrategias y posibilidades actuales para implementar un modelo como el presentado y las ventajas o desventajas de cada una de las opciones de su implementación.

Palabras clave: fraude, seguros de automóviles, procesos por fraude.

Abstract

Fraud detección in different sectors of human behaviour is more often a task where different computational tools can be used to facilitate and speed up the fraud search process. This paper describes the main features around the fraud detection problem in auto-insurance data, particularly in claim data. We will discuss how to organize a Knowledge Discover system in this subject, the group of processes that must be done, features that those processes must have and the interrelationships needed to eventually obtain a model able to discover rules interactively. This iterative process will be useful to refine the criteria to detect fraud until the ideal situation where the system determines if a given case can be classified as fraud. We also discuss the different strategy and options to implement a model like this or similar and the advantages - disadvantages of such approaches.

1. INTRODUCCIÓN

Es fácil encontrar en la literatura modelos generales para descubrir conocimiento como los presentados por Piatesky , que si bien sugieren métodos y técnicas para resolver problemas, son poco o no basados del todo en situaciones reales. En el mejor caso, se mencionan solamente los algoritmos de minería de datos, por ejemplo, la utilización de análisis de asociaciones por el Gobierno Federal de los Estados Unidos para resolver el caso de bombardeo de Oklahoma City o la ayuda que encontró el Departamento del Tesoro de Estados Unidos para encontrar los patrones fraudulentos entre la transferencia de fondos a nivel internacional.

También, las casas comerciales relacionadas al análisis de datos proponen, en forma general, tener una fuerte componente encargada de la minería de datos.

La utilización de las herramientas de minería de datos en detección de fraudes se complica porque no solo es difícil construir un conjunto de datos de entrenamiento, necesario por una buena parte de los algoritmos de minería de datos, sino que las acciones fraudulentos pueden basarse en diferentes esquemas y cambiar éstos en el transcurso del tiempo.

Es importante, no solo contar con esta componente del sistema, sino también analizar la situación

más ampliamente y reconocer la situación real y las necesidades de incorporar otros módulos para facilitar la tarea de detección de fraudes.

En este artículo, se presenta un modelo para descubrir conocimiento basado en una situación real: el descubrir irregularidades o situaciones anómalas que son indicios de fraudes en datos de reclamos de seguros.

Un modelo similar fue derivado de un estudio de un año, de las necesidades de información en el Instituto Nacional de Seguros de Costa Rica.

Primeramente, se describe el problema de detección de fraudes para posteriormente exponer la solucción de detección de fraudes en forma de diferentes módulos, como los son extracción, limpieza y carga de datos, filtro dinámico de datos, análisis de similitudes, análisis histórico, análisis de asociaciones y detección de nuevas reglas. Después, se presenta el modelo integrado y se describe la interacción que existe entre sus componentes. Finalmente, se analiza la posibilidad de su implementación incluyendo el uso de los recursos propios de la institución.

2. EL PROBLEMA

En general, el problema de detección de fraudes basado en datos va a depender de la aplicación correspondiente y de la naturaleza misma de la situación a tratar. En general, podemos decir que la detección de fraudes es la detección de patrones irregulares presentes en los registros de datos que no correspondan a las restricciones que el negocio mismo establece.

Un primer acercamiento al estudio de las situaciones fraudulentas es la presencia de montos grandes en las transacciones que se realizan. Por ejemplo, muchos de los fraudes en tarjetas de crédito, en facturación, en retiro de cuentas bancarias, en reclamos de seguros; se descubrieron por los montos excesivos fuera de lo acostumbrado o por la alta frecuencia de la incidencia de eventos correspondientes a los servicios ofrecidos.

Sin embargo, si analizamos más de cerca cada problemática, con los expertos correspondientes, podríamos obtener reglas de detección mucho más complejas y certeras, específicas y refinadas de acuerdo a cada caso en lugar de las reglas generalizadas y no siempre efectivas. Cabe la pregunta ¿Es posible crear un conjunto de reglas generales que puedan implementarse y aplicarse indistintamente de la situación real para ayudar en descubrir el fraude?. Probablemente no. El problema es entonces encontrar un sistema que permita descubrir estas reglas y refinarlas hasta tal grado que la utilización de éstas en un caso específico permita casi dilucidar si es una situación de fraude.

3. SOLUCIÓN A LA DETECCIÓN DE FRAUDES

Vamos a mostrar aquí una solución posible al problema de detección de fraudes en datos de seguros de autómoviles, sin embargo, consideramos que la solución propuesta –con las respectivas modificaciones- puede ser aplicada a otras situaciones de fraude o no en donde se involucren objetos y personas.

El modelo está compuesto de seis componentes: extracción, limpieza y carga de datos, filtro dinámico de datos, detección de entidades o análisis de similitudes, análisis histórico, análisis de asociaciones y extracción de reglas o indicadores. Cada uno de los módulos tiene una funcionalidad definida, descrita a continuación.

3.1 Extracción, transformación, limpieza y carga de datos.

Por lo general, los datos operacionales de aseguramiento no son todos necesarios para un análisis de fraude potencial. Por ejemplo, el número de recibo de la prima del automóvil, el número de caso del reclamo del reporte de accidente son los tipos de datos necesarios para su manipulación dentro del sistema operacional, sin embargo, no presentan mucha utilidad para el investigador del fraude. Por otra parte, puede presentarse la necesidad de contar con los datos provenientes de diferentes

fuentes, tanto internas como externas a la institución, los cuales pueden tener diferentes formatos o estar incorrectos. También, puede presentarse la situación que muchos de los datos muy importantes para los investigadores de fraudes no se encuentran en los sistemas operacionales disponibles actualmente en la institución.

Aquí podemos encontrarnos con tres escenarios presentes en la institución. Primero, se cuenta con un Almacén de Datos donde mucho del problema de limpieza y transformación de datos ya está resuelto y lo mejor sería alimentarse de este sistema. Sin embargo, dado que la investigación de fraudes, como cualquier aplicación de alto nivel de toma de decisiones, es frecuentemente ignorada en los sistemas operacionales, este acercamiento puede ser desventajoso cuando el Almacén de Datos de la institución fue construido sin considerar las necesidades de información requeridas para este tipo de investigación.

En el segundo escenario, podemos encontrar la situación en la cual la institución está en el proceso de construcción de un Almacén de Datos paralelo a la búsqueda de una solución para mejorar la detección de fraudes. Tomando en cuenta observaciones al primer escenario, podemos concluir que debemos incluir el personal de investigaciones en el grupo de futuros usuarios del Almacén de Datos. La funcionalidad del módulo de Extracción, Transformación, Limpieza y Carga (ETLC) de datos se concentraría principalmente a la carga de datos necesarios para el procesamiento de investigaciones.

El tercer escenario, no ofrece ningún Almacén de Datos ni tampoco planes de su construcción. La solución que proponemos es construir un Mercado de Datos (Data Mart) orientado a la investigación de fraudes. En este caso, el módulo ETLC se convierte a la típica tarea de ETL (Extraction, Transformation and Loading, extracción, transformación y limpieza) presente en el desarrollo de Almacenes de Datos.

Es obvio que el módulo de ETLC es la componente indispensable para el funcionamiento de los demás módulos, como por ejemplo de filtrado de datos.

3.2 Filtro dinámico de datos

A diferencia de la componente anterior de ETCL, la componente de filtrado dinámico de datos tiene la función de focalizar el análisis para las siguientes etapas. Como en cualquier sistema de descubrimiento de conocimiento, la labor de generar conocimiento posterior, es consumidora de recursos . La componente de filtrado garantiza que se analicen únicamente los datos necesarios y que pueden ser objeto de posible fraude. Nuestra posición aquí es que un análisis y procesamiento de todos los datos introducidos a sistemas transaccionales puede generar un volumen de datos demasiado grande y es innecesario, pues confunde, nubla y desvía la investigación que se realiza, llevando a seguir pistas falsas, procesar computacionalmente, y eventualmente manualmente, datos que muy probablemente no son fraudulentos.

Para ajustarse a los diferentes tipos de investigaciones a realizar, a la naturaleza cambiante del fraude, es necesario que este filtro pueda enfocar diferentes datos en diferentes ocasiones. Los criterios o reglas para el filtro de los datos deben ser tan exactos como se puedan, sin embargo, deben cambiarse a criterio del investigador o usuario. Esta componente debe tener la capacidad de establecer la lista de criterios de filtrado (uno o varios) con respectivo peso de cada uno de los criterios de acuerdo a la experiencia del investigador. También, debe permitir utilizar criterios flexiblemente y establecer cuales de los criterios han cumplido los datos filtrados para su subsiguiente análisis, de tal manera que si un grupo de datos cumple los criterios en un 100 %, posiblemente estamos ante una situación de fraude.

Para los seguros de automóviles existe amplia literatura donde se presentan los indicadores de fraudes. Estos indicadores representan una recopilación del conocimiento de expertos en el campo de investigación de fraudes. Esta lista y la experiencia propia de los investigadores pueden ser un conjunto inicial de criterios de filtrado. Por ejemplo, un seguro muy alto pagado en efectivo con un siniestro muy cercano a la fecha de aseguramiento

podría ser un criterio compuesto para analizar el caso. Cabe mencionar, que no necesariamente cada uno de los casos que cumplen las condiciones ejemplificadas anteriormente significa que existe el fraude. El sistema solo reporta los casos de sospecha de fraude y queda como labor del investigador confirmar o rechazar esta sospecha.

Como se mencionó anteriormente, el filtrado de datos permite disminuir la cantidad de datos dejando los casos más sospechosos para su futuro análisis, sin embargo, como los fraudes suelen repetirse entre las mismas personas o usando los mismos objetos, se ocupa una herramienta que permita encontrar objetos similares en el conjunto de datos filtrados.

3.3 Detección de entidades o análisis de similitudes

En una situación de fraude en reclamos de automóviles es factible esperar reclamos duplicados sobre un mismo accidente, personas que cambian ligeramente la escritura del nombre para no ser detectadas, automóviles fungiendo en diferentes reclamos, y otros. Básicamente, la situación común aquí es que un objeto o persona aparece repetidamente en varios sucesos. No obstante, esto no implica un fraude. Es necesario que los protagonistas en accidentes o reclamos sean claramente identificados ya sea como entes diferentes o idénticos. No se trata aquí de ver si una persona o entidad se parece o está emparentada con otra, aunque necesariamente será un producto de esta componente, sino de determinar cuales constituyen una misma para no confundir ni llevar a pistas falsas.

Las funciones de una componente de este tipo incluyen aspectos de detección de similitud por aspectos puramente sintácticos o fonéticos, similitud por transposición de nombres (en caso de que se registren nombres y apellidos en un mismo campo), similitud por sinónimos que provengan de conocimiento semántico del lenguaje, por parte del usuario o por criterios de emparejamiento que el usuario pudiera establecer. Al igual que en el caso anterior, la flexibilidad y presentación gráfica que esta componente es un atributo importante en el sistema.

En algunas situaciones, no es necesario aplicar el análisis de similitudes porque el comportamiento del objeto es fácilmente detectable por medio de análsis histórico.

3.4 Análisis histórico

El análisis histórico corresponde a la asociación de tiempos de ocurrencia de eventos con los objetos o entidades de la aplicación. En este caso, nos referimos a la asociación de tiempos a eventos tales como participación de una persona en accidentes, reclamos, o aseguramientos. No se trata de predecir un comportamiento sino de observar la actuación de personas con hechos importantes en el quehacer del negocio y que pudieran dar pistas a los investigadores sobre posibles hechos delictivos. La presentación visual o gráfica de estos hechos es parte de la funcionalidad requerida de la herramienta.

Es importante notar, que el análisis anterior de similitudes es un requisito para evitar confundir entidades y protagonistas en el análisis histórico.

El análisis histórico descrito puede ser considerado un caso especial de un análisis más amplio llamado análisis de asociaciones.

3.5 Análisis de asociaciones entre entidades

El análisis asociativo es una forma general de análisis en el cual pudiéramos incluir los casos anteriores. En todo el proceso investigativo, es necesario ver las relaciones existentes entre objetos y personas usando para eso las diferentes características presentes en los datos.

La funcionalidad aquí radica en mostrar asociaciones entre personas basada en direcciones, teléfonos, o trabajos similares. En general, se puede decir que el objetivo es buscar atributos y valores comunes entre las diferentes instancias, registros o tuplas de una tabla o varias tablas (de aquí la importancia del análisis de similitudes).

Las reglas de asociación que se obtienen aplicando algoritmos generales de minería de datos, pueden

ser utilizados aquí para encontrar grupos de atributos comunes en accidentes o reclamos, por ejemplo, coincidencia del lugar de accidente y nombre de la persona, es decir, una persona y un lugar ocurriendo con mucha frecuencia en los diferentes casos, situación que es nada usual en accidentes, sin embargo, algo que llama a un potencial fraude en caso de reclamos.

Como se mencionó anteriormente, cada una de las componentes presentes ayuda a investigador a encontrar los casos sospechosos de fraude. Sin embargo, el fraude es una situación cambiante en el tiempo y es indispensable que de la misma manera como el investigador desarrolla nuevos indicadores de sospecha de fraude, la herramienta los puede encontrar basándose en los datos que se tiene acumulados en la base de datos.

3.6. Extracción de reglas o indicadores

Cuando un grupo de casos haya sido designado como fraudulentos, es cuando se pueden aplicar algoritmos de minería de datos para extracción de reglas o para clasificación con el propósito de definir y establecer los criterios que pueden caracterizar uno o varios casos de reclamos. Estos criterios una vez extraídos pueden utilizarse en la componente de Filtro Dinámico de Datos para aumentar la calidad y la confiabilidad de los casos filtrados.

Aunque el análisis de asociaciones permite en cierta forma extraer reglas o indicadores, no es hasta que se disponga claramente de un criterio claro sobre la naturaleza fraudulenta del caso cuando se debe aplicar este análisis. Nótese que bien pudiera ser que el caso no se determine como fraudulento pero de interés para el usuario que requiere analizar casos similares en el futuro y por lo tanto es importante filtrar los casos que satisfagan criterios de clasificación similares.

Cada una de las componentes caracteriza las tareas que realiza un investigador real. Aquí es importante mencionar que estas componentes no deben ser elementos separados sino integrarse y formar un sistema global para análisis de fraude.

4. EL MODELO GLOBAL

Habiendo discutido la naturaleza de la principales componentes es posible dilucidar un modelo, como mostrado en la Figura 1.

El modelo propuesto no necesariamente representa la situación física, sino es el modelo lógico para entender la interacción que existe entre diferentes componentes y los datos que produce cada uno de ello.

El proceso de análisis empieza con la extracción de datos de los sistemas operacionales o almacenes de datos si éstos éxisten. La Figura 1. muestra únicamente la situación para el caso operativo. Debido al gran volumen de casos posibles, la componente de Filtrado seleccionará o enfocará el subsiguiente análisis en un subconjunto de los datos de acuerdo a las reglas o indicadores que en una primera instancia se introducirán a este módulo.

Las otras tres componentes de análisis actuarán sobre estos datos ya filtrados modificando o agregando nuevos, de ser necesario, para posterior procesamiento por otros módulos. También, pueden ser obtenidos reportes particulares de cada componente para investigación manual o para control del proceso investigativo.

Seguidamente, es necesario un proceso manual o semiautomático que permite evaluar cada caso y asignar un estado del caso. Este estado puede ser una simple variable binaria indicando la decisión tomada (fraude o no) o mucho más compleja que indique estados alternos como pendiente, irresoluble, vencido u otros.

Es sobre estos casos ya clasificados en los cuales un sistema de extracción de reglas basado en redes neuronales, agrupamiento o árboles de decisión puede trabajar consistiendo así el quinto módulo de Extracción de Reglas.

Estas reglas encontradas tienen dos fines: servir como nuevos criterios de filtro de datos en el módulo de Filtro de datos u como información de toma de decisiones para el negocio, útil para cambiar la estrategia o modus operandi para evitar o

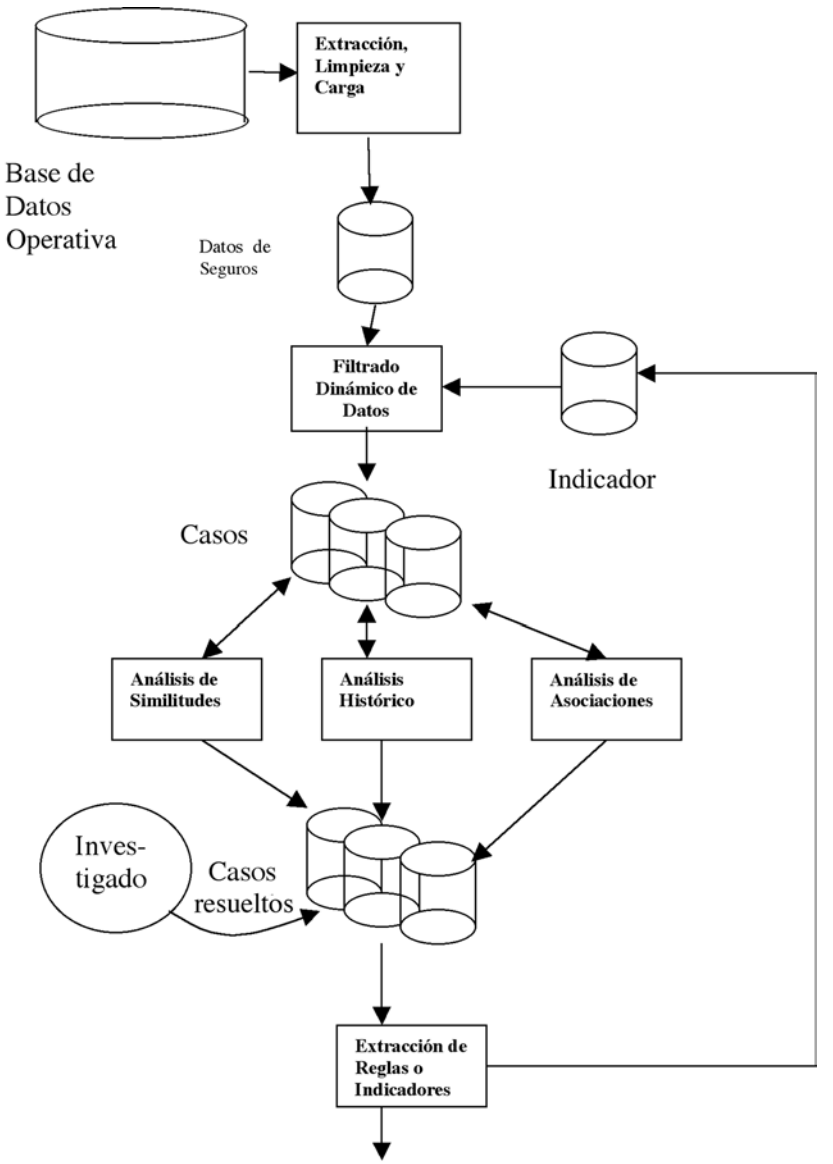


Figura 1. Un modelo para la detección de fraudes en seguros.

disminuir la incidencia de fraude en el negocio. Es importante notar que no todas las reglas extraídas pueden ser incorporadas plenamente en el módulo de filtro debido a las capacidades que pueda tener éste, objeción de los investigadores o aplicación es más adecuada al nivel del sistema operacional.

Habiendo mostrado el modelo, nos interesa ahora el mecanismo por el cual podemos llevarlo a cabo: implantación con herramientas existentes o desarrollo de una solución propia.

5. IMPLANTACIÓN CON HERRAMIENTAS EXISTENTES

Entre los mecanismos de implementación está la aquisición de sistemas de descubrir conocimiento que cumplan con las especificaciones de todos y cada uno de los módulos previamente descritos.

Desgraciadamente, un estudio realizado donde se analizaron herramientas provenientes de compañias como Cognos, SAS, SPSS y programas específicos Apoyo, SQL Server 2000 y otros, muestra que dada la particularidad del modelo, no existen sistemas en el mercado diseñados que se adapten o tengan un modelo similar al mostrado. Consideramos que existen y existirán aún más en el futuro, sistemas que implementen parcialmente soluciones a la problemática planteado con un menor o mayor acercamiento a las necesidades del usuario aquí expresadas.

Desesperanzadamente, muchos sistemas trabajan en problemáticas específicas tales como el DTS (*Data Transformación Services*, Servicio de transformación de Datos) de Microsoft que permite adecuadamente filtrar datos pero que las reglas deben ser programadas una a una a través de una interfase SQL o ActiveX (Visual Basic Script o Java Script) .

Otros sistemas son muy buenos encontrando Reglas de Asociación para canasta básica en supermercados, pero difíciles de adaptar a bases de datos generales. También, algunos permiten minar reglas en formas de Árboles de Decisión pero les falta mucho para automatizar el proceso de exportarlas a otros sistemas para que sean usadas en esos ambientes, por ejemplo, salidas en sintaxis SQL listas para ser incorporadas en sistemas como el DTS.

Además, muchas de estas herramientas aunque aseguran que son amigables, los usuarios poco experimentados los encuentran difíciles de manejar y de entender los resultados.

Consideramos que la solución global vía herramientas existentes cae en sistemas modulares que permiten cierto grado de programación en algún lenguaje de alto nivel y un alto grado de compatibilidad, interoperabilidad y comunicación entre componentes, de manera que una ligera modificación a uno de sus componentes sirve de solución equivalente a cuales quiera de los módulos aquí discutidos. En este respecto, los más aproximados -en un 60 % a 70 %- para el modelo presentado recaen, en nuestra opinión en sistemas como SAS o SPSS.

Sin embargo, como las instituciones que encuentran problemas de detección de fraudes en general cuentan con el personal informático propio, es necesario analizar la posibilidad de desarrollar una solución propia.

6. DESARROLLO DE UNA SOLUCIÓN PROPIA

Una alternativa de poner el marcha lo propuesto en el modelo, es desarrollar programación que sirva para crear el modelo desde sus inicios.

Este enfoque tiene sus ventajas (+) y desventajas (-):

- No permite aprovechar la experiencia de muchísimos años de compañías desarrolladoras ya existentes. (-)

- No existe personal informático disponible para el desarrollo de sistemas propios debido a la alta rotación de personal informático y esto no permite una continuación adecuada en el proceso de desarrollo. (-)

- Existe la situación de encontrar los costos relativamente bajos de proveedores internacionales. (+)

- Se permite mantenimiento adecuado de sistemas y adaptarlo rápidamente y fácilmente al cambio. Al contrario, al adquirir sistemas la realización de cambios es lenta y costosa dado que el proveedor es el único capacitado o autorizado para hacerlo. (+)

- Se requiere un alto grado de experticia que el personal disponible no tiene. (-)

- Se permite fomentar el conocimiento y el desarrollo de sistemas locales y tecnología propia. (+)

Consideramos que debido al alto costo de la adquisición de herramientas especializadas -principalmente en países en vías de desarrollo- una inclinación hacia desarrollo propio es bien justificada siempre y cuando exista el conocimiento y la capacidad técnica adecuadas.

También es importante analizar una solución intermedia donde se adquieren los productos con la suficiente flexibilidad que permiten incluir la programación propia en el caso necesario.

7. CONCLUSIONES

Hemos descrito las principales carácterísticas que envuelve el problema de detección de fraudes en seguros de automóviles, la necesidad de contar con los datos adecuados, la descripción de los diferentes procesos (módulos) que deben ser desarrollados, las peculiaridades que éstos deben tener y la interrelación que existe para contar con un modelo de apoyo en descubrimiento de fraude y eventualmente llevarlo al modelo de descubrir reglas iterativamente para afinar el proceso de detección del fraude. También, se discutió las estrategias y posibilidades actuales para implementar un modelo como el presentado o similar.

El problema de detección de fraudes en reclamos de carros tiene sus particularidades sin embargo, consideramos que puede ser generalizado a otro tipo de fraudes. El modelo global propuesto consistente de los módulos de Filtrado Dinámico, Análsis de Similitudes, Análsis Histórico, Análsis de Asociaciones representan el trabajo de un investigador visto por un analista de sistemas y aplicado a las herramientas computacionales. Más aún, la incorporación del Módulo de Extracción de Reglas amplia la "visión" del sistema sobre los indicadores y mejora el proceso investigativo.

El análisis de posibilidades de implementación basándose en herramientas existentes nos impulsó a divulgar los resultados de esta investigación.

La importancia de modelos específicos, como el aquí ilustrado, radíca en que permite a las casas productoras de programas y sistemas para descubrir conocimiento, conocer los requerimientos funcionales necesarios para la solución de los problemas del modelo y no limitarse a realizar sistemas generales que aunque altamente sofisticados en el tratamiento cientiífico de la información carecen de las funcionales básicas para su adaptación a sistemas de detección de fraudes o sistema particulares en general.

Adicionalmente, a lo descrito en el artículo, se puede mencionar que uno de los problemas ilustrados o descubiertos, con el desarrollo del modelo, fue la falta de estándares de intercambio de datos entre diferentes proveedores.

Eventualmente, una solución debe ser adquirida a una sola casa proveedora o se corre el riesgo de "secuestro" de los datos en el sentido de que una vez que estos son incorporados en un sistema no se pueda convertir o trasladar a otro sistema. Los proveedores prefieren los datos en su propio sistema pues así los pueden procesar más eficientemente.

Aunque existen métodos y estandáres tales como OLE DB y ODBC para permitir acceso a datos de diferentes bases de datos, éstos deben ser importados con la consiguiente duplicación de almacenamiento o esfuerzo – almacenamiento que en casos extremos puede ser prohibitivo.

La implementación del modelo en conjunto de los pros y contras de adquirir o implementar totalmente la solución, sugieren un estrategia intermedia: adquirir ciertos módulos con la flexibilidad suficiente para permitir la programación del modelo en base a esos módulos y en algunos casos programar dentro de la empresa los módulos que definitivamente existen soluciones externas muy alejadas.

8. AGRADECIMIENTOS

Queremos agradecer a los investigadores del Departamento de Investigaciones del Instituto Nacional de Seguros, Lic. Omar Soto, Lic. Enrique Mora y Sr. Rafael Blando, así como a los encargados del proyecto en el INS: M.Sc. Isabel Ortega, Lic. Guillermo Artavia y la Lic. Sandra Castro, quienes brindaron la información y realizaron las observaciones correspondientes.

9. BIBLIOGRAFÍA

1. Agrawal, R. T. Imielinsky, amd A. Swami; *"Mining Association Rules between sets of items in large data bases"*. Proceeding of the ACM SIGMOD International Conference on Management of Data. Washington USA. Pp 207-216. 1993.
2. Argüello J.R. , E. Malinowski; *"Especificación de Requerimientos para el INS"*. Reporte Interno. INS. Julio 2001.
3. Argüello J. R. , S. Chakravarthy; *"Extensions to Decision Tree Algorithms for classification and data mining in large databases"* en XIII Brasilian Symposium on Data Bases. 1998.
4. Barquin R.; et al, *"Building, Using, and Managing the Data WareHouse"*. Prentice Hall, 1997
5. Berry M. y Linoff G.; *"Data Mining Techniques"*. John Wiley &Sonsa, 1997
6. Insurance Investigation Specialists. http://www.iisjax.com/fraudind.html-#claims. Enero del 2002.
7. Inmon W. H.; *"Building the Data Warehouse"*, John Wiley & Sons, 1996
8. Kimball R.; *"The Data Warehouse Toolkit"*. John Wiley & Sons, 1996
9. Kimball R., Reeves L, Ross M. y Thornthwaite W. *"The Data Warehouse Lifecycle"*. John Wiley & Sons, 1998.
10. Mathews C. J.., P.K. Chan y G. Piatesky S.; *"Systems for Knowledge Discovery in Data Bases"*. .IEEE Transactions on Knowledge and Data Engineering. Vol. 5(6). Pags 903-913. 1993.
11. Microsoft; *"SQL Server 7.0 Data Warehousing Training Toolkit"*. Microsoft Press, 2000
12. SAS, "Using Data Mining Techniques for Fraud Detection", SAS Institute y Federal Data Corporation, 1999.
13. SPSS, "Using data mining to detect fraud", White Paper – Technical Report. http://www.spss.com 2001

SOBRE EL AUTOR

Dr. José Ronald Argüello Venegas
Profesor Catedrático de la Escuela de Ciencias de la Computación e Informática, Universidad de Costa Rica
Tel: 285-6610 Fax: 207-5527
Correo Electrónico:
jarguell@costarricense.cr

M. Sc. Elzbieta Malinowski Gajda
Profesora Catedrática de la Escuela de Ciencias de la Computación e Informática, Universidad de Costa Rica.
Tel: 207-4020 Fax: 207-5527