

Tecnología Inteligente de Minería de Datos en Telecomunicaciones

Edwin J. Berrios Negrón

Master of Engineering in Computer Engineering

Advisor: Dr. Nelliud Torres

Electrical & Computer Engineering and Computer Science Department

Polytechnic University of Puerto Rico

Abstracto - Las compañías de telecomunicaciones generan una enorme cantidad de datos. Estos incluyen; los datos de detalles de llamadas, que describe las llamadas que atraviesan las redes de telecomunicaciones, aunado a la red de datos que describe el estado de los componentes de hardware y software en la red, y por último los datos del cliente. Este describe sistemas expertos y tecnologías de minería de datos. Al igual que varias aplicaciones de telecomunicaciones y tecnologías inteligentes. Las siguientes tecnologías proporcionan descripciones detalladas de las aplicaciones de telecomunicaciones modernas representativas. El mismo describe el sistema experto *Operation Support System (OSS)* que monitorea y mantiene todos los elementos de conmutación en la red de AT & T. El sistema es notable por el hecho de que combina tecnologías basadas en objetos y reglas, para obtener ventajas de cada una.

Key Terms - *Artificial Intelligence, Data Mining in Telecommunication, TASA in Telecommunication.*

TECNOLOGÍA INTELIGENTE

La construcción de este tipo de aplicaciones, la adquisición de conocimientos valiosos de telecomunicaciones, el uso de expertos humanos y la incorporación de un sistema experto o de minería de datos, se utiliza ahora en las aplicaciones industriales mediante la sustitución de procesos para la adquisición de conocimiento manual, utilizando conocimiento automatizado. Para poder obtener las fallas de las telecomunicaciones, se va a utilizar la aplicación llamada *Operation Support System (OSS)*.

Las redes de telecomunicaciones, que de manera rutinaria generan enormes cantidades de

datos, son candidatos ideales para la minería de datos [1].

Sistemas Expertos

Los sistemas expertos son programas que representan y aplican el conocimiento actual de áreas específicas de experiencia para resolver problemas. Los mismos se han aplicado ampliamente en la industria de las telecomunicaciones, pero no sin problemas. Los primeros sistemas expertos requerían que un ingeniero en el área codificara este conocimiento en un sistema experto basado en reglas [2].

En este orden de ideas, estas reglas eran "ad-hoc", a medida que aumentaba el número de reglas, el sistema experto se volvió más engorroso de entender y modificar. Un sistema experto de primera generación es el 4ESS-ES, descrito más adelante en este artículo, Los de segunda generación intentaron resolver estos problemas utilizando métodos más sólidos, como el razonamiento funcional fundamentados en modelos. El razonamiento basado en modelos es de interés para la industria de las telecomunicaciones ya que los enfoques establecidos en modelos pueden representar la estructura y el comportamiento de los componentes de la red de telecomunicaciones de forma declarativa y luego razonar desde los primeros principios.

El sistema de soporte operativo OSS es el responsable de los servicios de comunicaciones a monitorear, controlar, analizar y administrar un teléfono o una red informática en la red de larga distancia AT&T.

La Base de Conocimiento y Minería de Datos

Este es un campo que ha surgido de diversas disciplinas, incluida la inteligencia artificial, el aprendizaje automático, las estadísticas y las bases de datos. Dicho proceso de base de conocimiento

implica la identificación válida, potencialmente patrones útiles y en última instancia comprensibles en datos. En este sentido, la minería de datos es el tema más investigado en este proceso, que consiste en encontrar patrones interesantes en los datos a través de análisis de datos y algoritmos de descubrimiento.

Sobre el proceso de descubrimiento de conocimiento, se ha hecho referencia en diversas investigaciones recientes, se compone de los siguientes pasos:

- Preparación de datos: seleccionar, limpiar y preprocesar los 'RET' Remote Electrical Tilt (por ejemplo, completar los valores que faltan) y transformarlos para que sean adecuados para la minería de datos.
- Minería de datos: la búsqueda de patrones y diferentes seriales de las antenas Quintel.
- Interpretación y evaluación: interpretar y evaluar los patrones producidos por la minería de datos.
- Una motivación clave para la base de conocimiento es que puede reemplazar o reducir al mínimo la necesidad de que el proceso que consume tiempo de adquirir manualmente el conocimiento de un técnico. La base de conocimiento es especialmente atractiva para la industria de las telecomunicaciones.
- Las redes de telecomunicaciones son típicamente demasiado complejas para construir modelos de simulación completas. Enormes cantidades de datos están disponibles rutinariamente.

Muchas de las aplicaciones en este artículo utilizan base de conocimiento. El paso en el proceso de base de conocimiento que normalmente requiere más trabajo para aplicaciones de telecomunicaciones es la etapa de transformación, que implica la identificación de características útiles para representar los datos. Este paso se complica por el hecho de que las redes de telecomunicaciones producen secuencias de alarmas, donde no es las alarmas individuales que son de importancia, pero el comportamiento en el tiempo de la red.

Minería de Datos

La minería de datos se encuentra interesantes patrones en los datos. Es un ejemplo de aprendizaje inductivo, ya que se basa en la generalización de los casos anteriores. Una aplicación típica de minería de datos de la industria de las telecomunicaciones es predecir el fallo de un componente de red basado en el historial de alarmas pasado. La minería de datos se puede utilizar para resolver muchas tareas, incluyendo las siguientes:

- Clasificación: aprender a trazar un ejemplo en una de varias clases.
- Clustering: ejemplos de partición en categorías que no están predefinidas.
- Modelado de dependencia: la búsqueda de un modelo que explica las dependencias entre las variables.
- Secuencial y detección de patrón temporal: el descubrimiento de patrones secuenciales o temporales entre los ejemplos.

Estas tareas pueden estar asociados con los problemas reales de telecomunicaciones. Todas las aplicaciones descritas en este artículo se puede considerar las tareas de clasificación (por ejemplo, es un elemento de red defectuoso o no). Los sistemas Locator Trouble y APRI, que se describen en este artículo, respectivamente, construir modelos de dependencia. La aplicación TASA se describe más adelante en la Localización de Problemas.

Hay muchos métodos de minería de datos para resolver las diversas tareas de minería de datos. Estos métodos varían de varias maneras, incluyendo: el tiempo que requieren para aprender, su tolerancia al ruido, el formato esperado de los datos y los conceptos que son capaces de expresar.

LOCALIZADOR DE PROBLEMAS: LOCALIZACIÓN DE PROBLEMAS DE LA RED DE CABLE

Debido a la congestión de las redes, existe la necesidad de mantener sistemas automatizados de localización de problemas en la red e identificar la causa real de los mismos. Una vez identificados,

poder tomar las acciones para eliminar o mitigar las fallas. Pacific Bell tiene un sistema inteligente que determina la localización de problemas en una red de cable telefónico local. Este sistema utiliza datos generados por una prueba de todas las noches automatizado para ayudar a reducir cables potenciales o equipo de red que puede ser defectuosa; Sin embargo, los resultados de las pruebas no son suficientes para determinar la causa exacta. El localizador problema utiliza una red bayesiana y la inferencia bayesiana para resolver este problema se alcanza, momento en el que unas listas clasificadas de los componentes defectuosos se pueden generar. Este sistema es utilizado por los analistas de mantenimiento preventivo como un sistema de soporte NOC (Network Operation System).

TASA: Encontrar Episodios de Alarma que Ocurren con Frecuencia

La alarma Secuencia Red de Telecomunicaciones Analyzer (TASA) [3] es un sistema para la base de conocimiento sobre el comportamiento de la red desde una base de datos de alarmas de la red de telecomunicaciones. El objetivo de este sistema es localizar regularidades en las secuencias de alarma con el fin de filtrar las alarmas redundantes, localizar problemas en la red y predecir los fallos futuros.

TASA opera en dos fases. En la primera fase, los algoritmos especializados se utilizan para encontrar reglas que describen que ocurren con frecuencia episodios de alarma partir de los datos de alarma secuenciales. Un episodio describe un conjunto de secuencias de alarma durante un período de tiempo determinado y este conjunto puede incluir secuencias de alarma en la que el orden específico de las alarmas no importa.

En la segunda fase, presentación de patrones colecciones de alarmas se forma interactiva manipulados por el usuario.

TASA ha sido probado con datos de alarma de telecomunicaciones y reglas interesantes se han encontrado y se incorporan en los sistemas de gestión de alarmas existentes.

Scout: Identificación de Fallas de Red a través de la Minería de Datos

El sistema Scout de AT&T identifica proactivamente las fallas transitorias recurrentes. Se opera por la minería de datos históricos de telecomunicaciones utilizando técnicas de aprendizaje automático y correlación. En un enfoque, scout identifica patrones de problemas crónicos directamente de los datos mediante el examen del comportamiento de la red en períodos de días y semanas. Para ello, las características que resumen los datos históricos variables con el tiempo se extraen de los datos para que los algoritmos de aprendizaje de máquina estándar se pueden utilizar (es decir, algoritmos que no son capaces de razonamiento temporal explícita).

Esta factorización se logra mediante el uso de dos ventanas de tiempo consecutivas fijos, W1 y W2. El objetivo es usar las medidas de W1 para predecir problemas en W2. Una forma de resumir estas medidas consiste en contar el número de veces que cada función tiene lugar dentro de la ventana. Scout luego usó Swap-1 para aprender las reglas que predicen recurrentes fallos transitorios.

Otra versión de Scout usa un mapa topológico de la red para mejorar su rendimiento. La adición de este conocimiento topológico permite que Scout aprenda usando intervalos de tiempo de minutos y horas en lugar de días y semanas, algo que es extremadamente importante si se van a predecir las fallas agudas.

Este conocimiento también le permite a Scout identificar problemas de manera efectiva con la misma causa raíz. Scout se ha desplegado en los Centros de servicio de red de AT & T desde 1990 y ha tenido éxito.

4ESS-ES: Gestión de Red para Conmutadores 4ESS

La mayor parte de la capacidad de conmutación para el tráfico de larga distancia nacional en la red de AT&T se proporciona por medio de interruptores 4ESS. Hasta hace muy poco, la 4ESS-ES (Sistema Experto 4ESS) era responsable de la gestión de estos interruptores: para el control de ellos, la realización

de pruebas de diagnóstico y alarmas de filtrado. El 4ESS-ES es una primera generación basado en reglas sistema experto, implementado en 1990 usando C5, una versión C del-5 OPS [4] idioma basado en reglas. Se compone principalmente de reglas ad-hoc poco profundas adquiridas durante un periodo de muchos años de expertos de dominio y sufre los mismos problemas que la mayoría de los sistemas de primera generación que es difícil de modificar y mantener. Este sistema se ha rediseñado y reimplementado en 1996 usando un paradigma basado en reglas híbrido orientado a objetos.

Otras Aplicaciones Inteligentes

Además de las aplicaciones de gestión de red, la industria de las telecomunicaciones Verizon, AT&T, T-Mobile and Sprint ha desarrollado muchos otros tipos de aplicaciones inteligentes, incluyendo: aplicaciones de configuración de productos, aplicaciones de servicio de asistencia y aplicaciones de detección de fraude / seguridad (CyberSecurity). Dado que estas aplicaciones no son exclusivas de la industria de las telecomunicaciones, sólo se describirá una sola aplicación.

OSS OPERATIONAL SUPPORT SYSTEM

Esta describe OSS Operational Support System, el sistema de soporte operativo responsable de servicios de comunicaciones a monitorear, controlar, analizar y administrar un teléfono o una red informática en la red de larga distancia AT&T.

Este sistema reemplaza al 4ESS-ES, el sistema experto de "primera generación" descrito en la parte Sistema Expertos. OSS es notable por el hecho de que utiliza ambas tecnologías y orientado a objetos basados en reglas, mediante el empleo de una extensión basada en reglas de lenguaje de programación orientado a objetos de C ++. La inclusión de la tecnología orientada a objetos facilitó el diseño y la implementación de OSS al proporcionar una forma de principio para modelar el OSS como una colección de dispositivos. A medida que los sistemas telefónicos de voz tradicionales convergen con el tráfico de Internet orientado a

paquetes (incluido VoIP), aplicaciones de banda ancha como teleconferencia y DSL, se necesitan sistemas más sofisticados como OSS, como ordenar y rastrear componentes de red (incluidas direcciones IP), tráfico, facturación e informes.

Descripción Funcional

Los técnicos de vigilancia de los dos centros de control de red de AT&T usan OSS Amdocs para ayudarlos a monitorear y mantener los servicios de comunicaciones, que manejan la mayoría de las llamadas y datos en la red de AT&T. Su desempeño del sistema experto OSS, es minimizar la cantidad de servicios que afectan los incidentes y los valores de los equipos. OSS ayuda a lograr estos problemas al observar miles de mensajes de alarma y notificar a los técnicos de las condiciones que son graves y muy probablemente requerirán la intervención humana. Se define la aplicación de OSS que ayuda a detectar el sistema de alarmas 'RET' Remote Electrical Tilt, esta nos va a optimizar la red. Esto se logra mediante la observación de miles de mensajes y notificaciones a los técnicos de parte del sistema. Este notifica las condiciones que son graves y requieren de intervención humana para que sea posible alertar a los técnicos sobre los problemas que requieren mayor atención. El objetivo de la aplicación OSS es minimizar la cantidad de servicios que afectan los incidentes y el número de llamadas bloqueadas o perdidas.

OSS es un sistema completo de soporte para operaciones de telecomunicaciones e incluye muchos componentes. El componente central y más importante es la base de conocimiento, que es responsable de todo el comportamiento inteligente de OSS. En la Figura 1, se muestra la funcionalidad básica de entrada/salida de OSS. Las alarmas se pueden bloquear o se pueden perder ya que, en la triangulación del sistema de células, estas no están alineadas de acuerdo con su diseño, desde la perspectiva de su base de conocimiento.

Los mensajes de alarma proporcionan información sobre condiciones que ocurren en un dispositivo específico dentro de OSS, mientras que los mensajes que no son de alarma proporcionan

información de estado o los resultados de diagnósticos previamente solicitados por la base de conocimiento. Los resultados clave de la base de conocimiento son los comandos del OSS (por ejemplo, para solicitar que se ejecuten los diagnósticos) y las notificaciones a los técnicos de problemas que requieren mayor atención. Estas notificaciones vienen en forma de alertas y tickets.

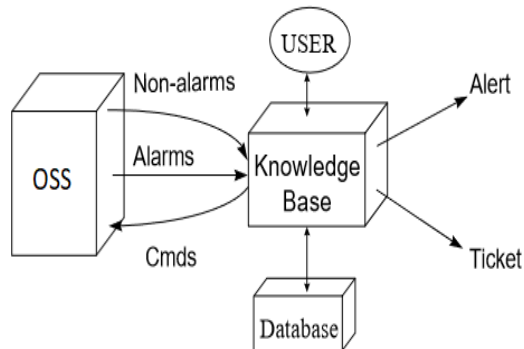


Figura 1
Vista Funcional de OSS

Las alertas se envían a los técnicos de vigilancia en los centros de control de red NOC (Network Operation Center), que luego deciden qué medidas tomar. Pueden ignorar la alerta, diferir el procesamiento de la alerta en espera de información adicional, o crear manualmente un ticket de alarmas de "orden de trabajo" y enviarlo a la fuerza de trabajo en el sitio. Los miembros de la fuerza de trabajo en el sitio están ubicados en cada conmutador OSS y pueden realizar reparaciones físicas del equipo antenas y radios RRU, como la sustitución de un TMA, SPF, fibras, controladoras etc. En los casos donde la necesidad de la intervención física humana es clara, la base de conocimiento puede crear el boleto de alarmas de forma autónoma y enviarlo directamente a la fuerza de trabajo en el sitio, evitando a los técnicos de vigilancia centralizados.

La tarea principal de la base de conocimiento es decidir cuándo generar una alerta o un boleto de alarma. Estas notificaciones no se generan para cada alarma recibida por la base de conocimiento, y, por lo tanto, una de las funciones clave de la base de conocimiento es realizar un filtrado de alarma. Por ejemplo, en muchos casos, una alerta o boleto de alarma solo se genera si se excede un número de

umbral de alarmas dentro de un período de tiempo específico. La base de conocimiento es solo una parte de OSS. Además de las entradas y salidas que acabamos de describir, la base de conocimiento también está conectada a una base de datos y a los usuarios, a través de una interfaz gráfica de usuario. La base de datos es utilizada por OSS para proporcionar un almacenamiento persistente de alarmas, alertas y otra información, parte de la cual está ubicada allí por la base de conocimientos y parte de la cual es colocada allí por otras partes del sistema. Los usuarios también pueden interactuar con la base de conocimiento a través de la interfaz gráfica de usuario para recuperar información y establecer varias opciones para personalizar el comportamiento de la base de conocimiento.

PREDICCIÓN DE FALLAS EN LOS EQUIPOS DE TELECOMUNICACIONES A PARTIR DE LOS DATOS DE SERIES TEMPORALES

Las técnicas de minería de datos se han aplicado con éxito a las tareas de clasificación. Muchas técnicas de clasificación han sido investigadas en los últimos veinte años, lo que resulta en muchos algoritmos útiles para encontrar patrones predictivos en los datos. Los datos utilizados en las tareas de clasificación se modelan en una representación estándar basada en casos, donde cada caso incluye un conjunto de variables de características y una variable de clase única. La tarea de predicción es entonces predecir la variable de clase en función de las variables de característica. Para aprovechar las técnicas existentes, es necesario modelar los datos usando la representación basada en casos. Lamentablemente, los conjuntos de datos de muchas tareas no se expresan naturalmente en este formato estándar. Los datos de series temporales, por ejemplo, no se expresan naturalmente en el formato estándar basado en casos y no se pueden convertir trivialmente en este formato. Las series de tiempo pueden contener una o más variables; la serie temporal de alarmas de telecomunicaciones que se analiza más adelante en esta sección contiene varias variables.

RESULTADOS

De la tabla 1, podemos observar los modelos de cada antena, su número de serie y su identificador, entre otros datos. Esto evita la presencia de técnicos, que usualmente cotejaban la antena manualmente. El sistema permite recopilar esta información más rápidamente.

Los resultados claves de la base de conocimiento son los comandos del OSS (por ejemplo, para solicitar que se ejecuten los diagnósticos) y las notificaciones a los técnicos de problemas que requieren mayor atención. Las mismas vienen en forma de boleto y alertan al técnico para su diagnóstico.

Tabla 1
'RET' Remote Electrical Tilt Alarm

SiteID	eTilt	Antenna ModelNum	AntennaSerialNumber	onUnit/Uniqueid	Sector	operational State	eTILT
PRO0470	90	QS665122_R1_A	QS665122V52715975	RFA114QU08212	Alpha	ENABLED	20
PRO0470	50	QS665122_R2_A	QS665122V52715975	RFA114QU08211	Alpha	ENABLED	20
PRO0470	20	QS665122_Y1Y2_A	QS665122V52715975	RFA114QU08224	Alpha	ENABLED	20
PRO0470	80	QS665122_B1B2_A	QS665122V52715975	RFA114QU08235	Alpha	ENABLED	20

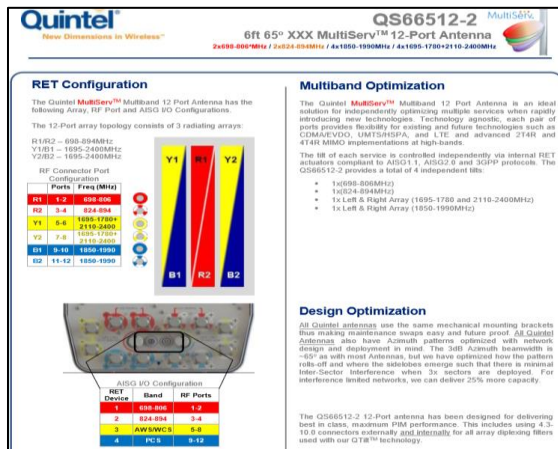


Figura 2
Modelo Antenas Quintel

Las notificaciones se clasifican en baja y alta prioridad que notifican a los técnicos de problemas que requieren mayor atención. Las alertas se envían a los técnicos de vigilancia en los centros de control de red NOC (*Network Operation Center*), que luego deciden qué medidas se van a tomar. Pueden ignorar la alerta, retrasar el procesamiento de la alerta en espera de información adicional, o crear manualmente un ticket de "orden de trabajo" para enviar al técnico de turno. Con el sistema experto

OSS se puede ver una colección de los modelos de las antenas, valores y no requerirán la intervención humana.

CONCLUSIÓN

Este artículo ha descrito varias tecnologías y aplicaciones de telecomunicación inteligentes. Se destacaron dos tecnologías: sistema experto y minería de datos. El sistema experto es la tecnología más exitosa, con muchos éxitos comerciales en las últimas dos décadas. La descripción de OSS de este artículo muestra cómo la tecnología inteligente, una tecnología que goza de gran éxito comercial [5], se puede integrar con tecnología basada en reglas para facilitar el desarrollo y el mantenimiento de un sistema experto.

Tanto el sistema experto basado en reglas como las tecnologías de minería de datos ayudan a la industria de las telecomunicaciones a lidiar con las grandes cantidades de datos disponibles. Por ejemplo, el sistema experto OSS procesa miles de mensajes de alarma diariamente y puede diagnosticar problemas en el equipo [6]. En este caso, el conocimiento contenido en el sistema experto se adquirió mediante programación de algoritmos. La tecnología de minería de datos, sin embargo, permite que el conocimiento útil sea adquirido automáticamente y directamente de los datos. Esto fue demostrado por varias de las aplicaciones; es un intento de complementar las reglas adquiridas manualmente de la aplicación OSS con reglas adquiridas automáticamente a través de minerías de datos.

Si bien es poco probable que la minería de datos elimine por completo la necesidad de adquirir conocimiento de expertos humanos, la misma está haciendo contribuciones significativas a la industria de las telecomunicaciones y esperamos que estas contribuciones se aceleren a medida que se avance en la tecnología de sistema experto.

REFERENCIAS

- [1] S. Weiss & N. Indurkha, *Predictive Data Mining: A Practical Guide*, Morgan Kaufmann, San Francisco, 1998.

- [2] F. Hayes-Roth, D. Waterman & D. Lenat, *Building Expert Systems*, Addison-Wesley, Reading, MA, 1983.
- [3] K. Hatonen, M. Klemettinen, H. Mannila, P. Ronkainen & H. Toivonen, "Knowledge discovery from telecommunication network alarm data bases," in *12th International Conference on Data Engineering (ICDE '96)*, New Orleans, Louisiana, Feb. 1996.
- [4] T. Cooper & N. Wogrin, *Rule-Based Programming with OPS5*, Morgan Kaufmann Pub., San Mateo, CA, 1988.
- [5] J. Rumbaugh, M. Blaha, W. Premerlani, F. Eddy & W. Lorensen, *Object-Oriented Modeling and Design*, 1st ed., Prentice Hall, 1991.
- [6] H. Hanrahan. (2007, March). *Network Convergence: Services, Applications, Transport, and Operations Support* [Online]. Available: www.hit.bme.hu/~jakab/edu/litr/TMN/oss.pdf.