

Minería de Datos en Respecto de la Incidencia del Clima sobre el Desperfecto en el Alumbrado Público

Sergio Quiroga, Mag. Alejandra Malberti, Mag. Raúl Klenzi

Instituto de Informática / Departamento de Informática / Facultad de Cs. Exactas Físicas y Naturales / Universidad Nacional de San Juan

Av. Ignacio de la Roza 590 (O), Complejo Universitario "Islas Malvinas", Rivadavia, San Juan, Teléfonos: 4260353 4260355 - 4260394 - 4264721 - 4234129, Fax 0264-4234980, <http://www.exactas.unsj.edu.ar>
{sergiooquiroga,amalberti,rauloscarklenzi}@gmail.com

Resumen

El presente estudio corresponde a una extensión de un trabajo previo “Minería de datos en la detección de desperfectos en el alumbrado público” (WICC 2015), propone extraer información que permita descubrir la incidencia de factores climáticos en la cantidad y tipología de fallas habituales que se producen en el alumbrado público. Los datos provienen de reclamos de usuarios, y otros inherentes a condiciones climáticas particulares registradas en las fechas tratadas. Previo al análisis de los datos por medio de Minería de Datos se consideran aspectos tales como ruido, datos ausentes, y volatilidad, entre otros. Las direcciones correspondientes a los reclamos se normalizan para poder realizar tareas de geolocalización de Google, con la finalidad de detectar zonas de conflicto. En esta propuesta se aplica la metodología CRISP-DM y se utiliza la herramienta Knime analytics, en especial la API de Google Maps.

Palabras clave:

Minería de datos -Geolocalización- Knime – Google Maps

Contexto

La línea de investigación que permite elevar la presente propuesta está contenida en el proyecto bianual “La Ciencia de Datos en grandes colecciones de datos”, aprobado por CICITCA-UNSJ y sujeto a evaluación externa. Los datos procesados provienen de “reclamos” generados en el call-center de una empresa local, que se encarga del mantenimiento del alumbrado público en la ciudad de San Juan. Estos reclamos son generados por los vecinos que detectan fallas en el alumbrado público y solicitan, vía telefónica, su reparación. Cada reclamo es transcripto en una planilla Excel por el operador receptor del llamado, quien registra la dirección postal o catastral correspondiente a la ubicación de la luminaria, o conjunto de luminarias, que presentan desperfectos. Los datos recepcionados son tratados por medio de la aplicación de técnicas de minería de datos.

Introducción

El presente trabajo involucra a una empresa que se dedica al mantenimiento del alumbrado público en diversos departamentos de la provincia de San Juan. Esta empresa se ha propuesto mejorar el servicio que brinda a sus usuarios, especialmente en lo que refiere

erróneos, por tanto, es necesario entender el modo de trabajo de Google Maps. Si bien es limitado en cuanto a lo que se pueda desarrollar internamente, hay que comprender que para realizar posicionamiento se maneja con coordenadas a través de su latitud y longitud.

Google Maps utiliza para funcionar los lenguajes HTML CSS y Javascript. Los mapas son solo imágenes que están colocadas en una página HTML. El objetivo de la API es enviar información de las coordenadas al mapa a través de AJAX¹ y devuelve como resultado la imagen con la posición.

El API consiste de archivos JavaScript que contienen las clases, métodos y propiedades que se usan para el comportamiento de los mapas.

Las coordenadas están expresadas usando números decimales separados por coma. La latitud siempre precede la longitud. Google maps no se limita a cierta cantidad de decimales, sin embargo, según unas pruebas realizadas en la web <http://ojalesa.mbnet.fi>² concluye que:

5 a 6 decimales: es el máximo que debemos usar para ser específicos

4 decimales: para algún detalle en el mapa

3 decimales: es bueno para centrar ciudades

2 decimales: es apropiado para centrar países o estados, tal vez 3 por Mónaco

La abundante cantidad de datos recogidos son tratados por medio de técnicas adecuadas, entre las que se encuentra la Minería de Datos (MD). Para Witten (Witten, 2005) la Minería de Datos es el proceso de extraer

conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos. Este proceso conlleva técnicas de análisis de datos destinadas a extraer patrones, describir tendencias y regularidades, predecir comportamientos y, en general, aprovechar un gran volumen de información digitalizada. Para tal fin se utiliza el software Knime Analytics, una plataforma open source (de código abierto), *...diseñado para descubrir el potencial oculto de los datos, usando la minería para obtener nuevas ideas o predecir nuevos futuros...*³.

En la Figura 4 se muestra el proceso para visualizar ubicaciones de reclamos en Google Maps, realizado en Knime.

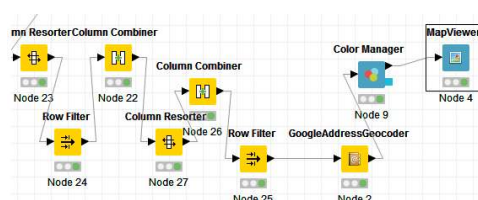


Figura 4 : proceso Knime para visualizar las ubicaciones de los reclamos en Google Maps

Líneas de Investigación, Desarrollo e Innovación

En el marco del laboratorio se llevan adelante diferentes trabajos de investigación aplicada, caracterizada por el tipo de datos observados y analizados. Según la variedad y tipología de datos se está trabajando en análisis de series temporales, aplicaciones en Data Streammining, y reconocimiento de perfiles de usuarios, intentando llevarlas a plataformas paralelas en cluster de computadoras.

pequeños paquetes de datos. Fuente: aprenderaprogramar.com Copyright 2006-2017

² Página de ejemplos y testeos de la API de Google Maps: <http://ojalesa.mbnet.fi/exam/decimal.html>

³ Cita de <https://www.knime.org/>

¹ Para agilizar los tiempos de espera al servidor web surgió Ajax (inicialmente Asynchronous JavaScript And XML, aunque hoy día ya no es una tecnología ligada a XML con lo cual no pueden asociarse las siglas a estos términos). Tecnología que busca evitar las demoras propias de las peticiones y respuestas del servidor mediante la transmisión de datos en segundo plano usando un protocolo específicamente diseñado para la transmisión rápida de

Se utiliza la metodología CRISP-DM4, este modelo del proceso de Minería de Datos, para realizar tratamiento de grandes cantidades de datos, con implementaciones de algoritmos de minería de datos,

Se utiliza la metodología CRISP-DM5, este modelo del proceso de Minería de Datos, se divide en seis pasos Comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue

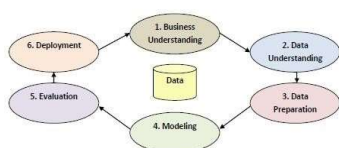


Figura 5: Modelo conceptual de CRISP-DM

En nuestro caso, se aplican tareas de: 1) ingreso de datos, ajustando los parámetros de los diferentes módulos de Knime al formato de los datos de entrada 2) preprocesamiento, filtrados, ajustes y adaptación de formatos de datos, como de diferentes estrategias, siendo estas: segmentación, clasificación entre otras.

Objetivos y Resultados

El objetivo que se persigue es proporcionar zonas en las que se detectan fallas en el alumbrado público debido a las inclemencias, a partir de la aplicación de estrategias de minería de datos, análisis de factores climáticos e incorporación y procesamiento de información relativa a la geolocalización.

Al momento de la elaboración de este documento, se han obtenido los siguientes resultados:

- Profundización en el conocimiento de las tareas de minería de datos, y de plataformas open source.

- Tratamiento de bases de datos provenientes de distintas fuentes.

- Detección y solución de inconsistencias en los datos.

- Estudio de técnicas y herramientas de geolocalización.

- Geolocalización de direcciones de reclamos en Google Maps.

Así mismo, se ha trabajado con los módulos contenidos en KNIME GoogleAddressGeocoder y MapzenGeocoder permitiendo éste último una mayor carga de registros por unidad de tiempo, aunque con una precisión de ubicación menor asociadas a las variables latitud y longitud de una referencia cartográfica.

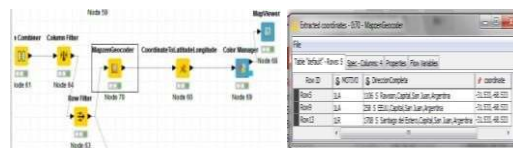


Figura 6

Esto hace, como se puede apreciar en la figura 6, que puntos con diferente dirección postal sean ubicados en un mismo punto de coordenadas.

Por el contrario, el módulo Google AddressGeocoder genera la salida que se puede apreciar en la Figura 7 y donde cada uno de las direcciones postales posee su correspondiente coordenada.

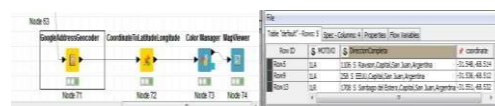


Figura 7

Por otro lado, la herramienta permite desde un módulo de combinación de columnas de datos, alcanzar el formato necesario con que deben ser ingresadas las direcciones postales en cada uno de

⁴Cross Industry Standard Process for Data Mining

⁵Cross Industry Standard Process for Data Mining

los módulos o APIs que transforman este dato a un sistema de coordenadas de latitud y longitud como muestra la figura 8.



Figura 8

Formación de Recursos Humanos

Las tareas desarrolladas en el ámbito del proyecto han permitido la conclusión y desarrollo actual de muchos trabajos finales de grado pertenecientes a alumnos de las carreras contenidas en el ámbito del Departamento Informática de la Facultad de Ciencias Exactas Físicas y Naturales.

Referencias

- Mark Salvador, Ron Resmini (auth.), Guido Cervone, Jessica Lin, Nigel Waters (eds.) **-Data Mining for Geoinformatics. _ Methods and Applications-** Springer New York (2014)
- Harvey J. Miller, Jiawei Han, **Geographic Data Mining and Knowledge Discovery**, Second Edition (Chapman & Hall CRC Data Mining and Knowledge Discovery Series)-CRC Press (2009)
- Hernández Orallo, José; Ramírez Quintana, Ma José; Ferri Ramírez, César. (2004) **Introducción a la Minería de Datos**. Edit. Pearson educación
- Microsoft (2013) Prueba y validación (minería de datos) <http://msdn.microsoft.com/esAR/library/ms174493.aspx>

- Molina L. (2002) **Data mining: torturando a los datos hasta que confiesen**.

<http://www.uoc.edu/web/esp/art/uoc/molina1102/molina1102.html>.

- Pérez López, César; Santín González, Daniel. (2006) **Minería de Datos Técnicas y Herramientas**. Edit. Alfaomega Grupo Editor

- Pyle, Dorian; Kaufmann, Morgan. (1999) **Data Preparation for Data Mining**. Edit. Morgan Kufmann Publishers Inc.

- Witten, Ian H.; Eibe, Frank. (2005) **Data Mining Practical Machine Learning Tools and Techniques**. Edit. Elsevier Inc.

- **Geolocalización**. (s.f.). En Wikipedia. Recuperado el 1 de febrero de 2017 de <https://es.wikipedia.org/wiki/Geolocalizaci%C3%B3n>

North, M. (2012). *Data mining for the masses* (pp. 91-100). Global Text Project.