

APLICACIÓN DE LA TÉCNICA DE MINERÍA DE DATOS PARA LA EXPLORACIÓN DE CONOCIMIENTO IMPLÍCITO Y PREDICCIÓN DEL CLIMA

— Beatriz Eugenia Marín y Luis Alejandro López
Semillero ITMEDIA
Grupo de Investigación en Tecnologías de la Información y Comunicación -
(GRINTIC)
Institución Universitaria Antonio José Camacho

Resumen

Este artículo pretende contextualizar el uso de la minería de datos para el hallazgo de información implícita del clima, teniendo en cuenta qué herramientas o técnicas fueron usadas en otros proyectos en cuestión de algoritmos, metodologías o modelos favorables para la extracción de conocimiento implícito de las variables del clima.

Palabras clave

Minería de datos, algoritmos, predicción, clima, exploración, técnica.

Abstract

This article intends to contextualize the use of data mining for the discovery of implicit information of the climate, taking into account that tools or techniques were used in other projects regarding algorithms, methodologies or models favorable for the extraction of implicit knowledge of the variables of the weather.

Keywords

Data mining, algorithms, prediction, weather, exploration, technique.

Introducción

La minería de datos unida a procesos de KDD para la extracción de conocimiento pertenece a una serie de nuevas técnicas y tecnologías que están produciendo notables contribuciones. Aplicaciones como la minería web, inteligencia de negocios, predicción del mercado, el filtrado de spam, entre otras, se desarrollan con la ayuda de técnicas de minería de datos. Las técnicas de minería de datos se ofrecen para analizar los datos históricos y obtener de ellas patrones de comportamiento. Dichos patrones se utilizan para identificar similitu-

des entre datos en la tarea de clasificación, o para hacer predicción. El proceso de KDD en el campo de la minería de datos se refiere al proceso no-trivial de descubrir conocimiento e información primaria útil dentro de los datos contenidos en algún repositorio de información (WMC, 2011). No es un proceso automático, es un proceso iterativo que exhaustivamente explora volúmenes muy grandes de datos para determinar relaciones. Es un proceso que extrae información de calidad que puede usarse para dibujar conclusiones basadas en relaciones o modelos dentro de los datos. (Avalos Flores, 2010).

Al utilizar un proceso para la extracción de información mediante la técnica de minería de datos se puede realizar un modelo de predicción del clima. En este orden de ideas, la técnica de pronóstico meteorológico analiza los datos meteorológicos históricos y de variables o atributos correspondientes, variables tales como la humedad, rayos ultra violeta, dirección del viento, entre otros. A partir de este análisis y evaluación de variables e identificando patrones similares, se realiza la previsión meteorológica. Como se discutió anteriormente, las técnicas de minería de datos son útiles para diversos tipos de extracción de patrones, reconocimiento, objetivo y predicción. Estos patrones significativos extraídos constituyen los datos que ayudan a la toma de decisiones, inteligencia de negocios y otras tareas tecnológicas (Yavad et al, 2016).

Proyectos relacionados con la relación y predicción del clima en Colombia

En la Universidad Nacional de Manizales estudiantes del departamento de Informática y Computación se encuentran adscritos a semilleros y grupos de investigación dirigidos por el profesor Néstor Darío Duque.

Con respecto al tema de minería de datos con variables del clima, la universidad actualmente cuenta con el apoyo del Instituto de Hidrología, Meteorología y Estudios Ambientales de Manizales (IDEAM). Este vínculo ha permitido el suministro de información a estudiantes de sistemas para proyectos de minería de datos en el ambiente de desarrollo de conocimiento de la climatología.

Es así como se han presentado algunos proyectos en los que se le da un valor agregado a la información que aporta la red de estaciones que maneja el IDEAM para hallar nuevas características del clima, dando lugar a nuevos temas de estudio e investigación en áreas como gestión del riesgo, au-

tomatización, ingeniería en electrónica entre otros (Johana Gonzales, 2015).

El IDEAM maneja estaciones de tipo Meteorológica, Hidrológica e Hidrometeorológica, que se encuentran publicadas en la plataforma web del IDEAM, a través de la cual es posible realizar solicitudes de información sobre sus estaciones. A continuación se presenta una vista del mapa de estaciones del IDEAM que se encuentran sobre el territorio colombiano y la cantidad de información que de ellas se puede tomar para realizar estudios en diferentes campos y sobre diferentes regiones colombianas.

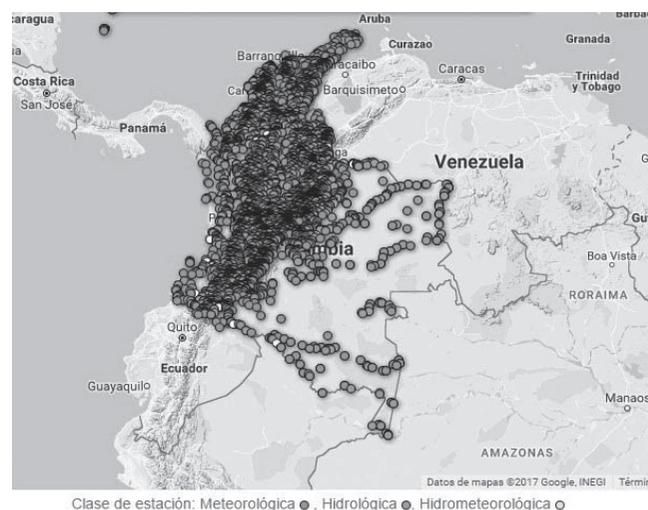


Figura 1 No. Red de estaciones del IDEAM en Colombia (IDEAM, 2014).

Según el ingeniero Duque, la técnica de minería de datos se apoya en estadísticas fuertes o en el uso de inteligencia artificial para encontrar estándares ocultos o tendencias sobre las que se podría hacer pronósticos de diversa índole (Agencia de noticias UN, 2011). Debido a esto los estudiantes encontraron una oportunidad de estudio del clima sobre la información que aporta la red de estaciones del IDEAM, siendo una fuente primaria de información con respecto a variables del clima en Colombia.

Proyectos relacionados a la exploración y predicción del clima en el exterior

En el estado del arte se reconocen algunos trabajos realizados sobre el tema en cuestión fuera de Colombia, entre ellos solo se mencionarán algunos que tengan características relacionadas con la exploración y predicción de información con respecto al clima y que representen un avance en el campo.

- En Nigeria, *Application of Data Mining Techniques in Weather Prediction and Climate Change Studies* por Forolunsho Olaiya et al (y otros), es un trabajo en donde se usa el proceso y sigue las fases de KDD para lograr un modelo de extracción junto a algoritmos de redes neuronales, árboles de decisión con datos meteorológicos del año 2000 al 2009 de la ciudad de Ibadan, Nigeria. Estos datos fueron usados para comprobar que esta información era útil para poder predecir el clima y fueron comparados con los datos capturados de los días que fueron predichos para comprobar la veracidad de la información hallada. (Forolunsho Olaiya et al, 2012)
- En España, *Aportaciones de la redes bayesianas en meteorología. Predicción probabilística de predicción*, por Rafael Ancell Trueba et al, es una investigación en la que se describe brevemente el problema de la minería de datos desde la perspectiva de los modelos de redes de probabilísticas (las redes bayesianas en particular). En primer lugar, se muestra la capacidad de estos modelos para codificar en un grafo todas las asociaciones/dependencias relevantes entre un conjunto de variables, y también para transferir de forma automática esta estructura de dependencia a una dependencia de probabilidad conjunta, calculando de esta manera y en forma eficiente una probabilidad posteriori. (Rafael Ancell Trueba et al, 2013).
- En Cuba por Gonzalo Joya-Caparrós et al, *Aplicación de la técnica de minería de datos SOM* (Mapas auto-organizados, por sus siglas en inglés), utiliza el lenguaje en R en datos climáticos, para describir el comportamiento climático a partir de los datos observados con el fin de buscar relaciones y agrupamientos entre ellos. Para lograrlo usaron la técnica de minería de datos con mapas auto-organizados (SOM) y se confeccionó un paquete en R para facilitar el análisis. En este trabajo de logró una descripción del comportamiento climático de Cuba en el año 2011, el cual se caracterizó por presentar dos estaciones bien definidas en periodos concretos y se obtuvo una descripción más detallada de cada uno de estos períodos. (Gonzalo Joya-Caparrós, 2013-2014)
- En India por Dr. R. Satya Prasad et al, *Use of Data Mining Techniques for Weather Data in Basra City*. En este trabajo Satya y otros investigadores, proponen extraer conocimiento útil del historial de datos recopilados localmente en el tiempo del sector de la ciudad de Basra, para ello recopilan datos meteorológicos para la ciudad y luego usan la tecnología necesaria para hacer minería de datos. Los datos incluyen un período de nueve años [2004-2013]. Después de procesar los datos y revisar anomalías en los datos, se aplican reglas en los algoritmos de minería de datos, como: clustering, predicción, clasificación y asociación, a partir de los cuales pretenden extraer conocimiento y describir a la minería de datos en su importancia en el campo del tiempo. (Dr. R. Satya Prasad et al, 2015).
- En India por Rohit Kumar Yadav y Ravi Khatri, *A Weather Forecasting Model using the Data Mining Technique*. El trabajo pretende investigar sobre el patrón de condiciones climáticas y su modelo de predicción. Por otro lado, la técnica de minería de datos permite en este trabajo analizar los datos y extraer los patrones valiosos.

Estos patrones identificados a partir de los datos históricos permiten aproximar las condiciones climáticas próximas y sus resultados. Para diseñar y desarrollar un modelo de datos tan preciso, se revisan varias técnicas y se recogen los enfoques más prometedores. El modelo de datos propuesto incorpora el modelo de Markov oculto para la predicción y para la extracción de las observaciones de la condición meteorológica, utilizando la agrupación de K-means. Para predecir las condiciones nuevas o futuras, el sistema debe aceptar los escenarios actuales de las condiciones climáticas.

La implementación de la técnica propuesta se realiza con la tecnología Java. Adicionalmente para la justificación del modelo propuesto se utiliza el estudio comparativo con el algoritmo ID3 tradicional. Para comparar ambas técnicas, la precisión, la tasa de error y la complejidad de tiempo y espacio se estima como los parámetros de rendimiento. De acuerdo con los resultados obtenidos, el rendimiento de la técnica propuesta se encuentra mejorada en comparación con la técnica basada en ID3 disponible. (Rohit Kumar Yadav & Ravi Khatri, 2016)

Ventajas y desventajas del uso de Data mining para la predicción del clima

La alerta temprana del comportamiento del clima es la aplicación de la ciencia y tecnología para predecir el estado de la atmósfera para una localización determinada. Antiguos métodos de previsión del clima, utilizados por los babilonios en la época de 650 a.C., predecían el tiempo a través de patrones de nubes, la forma más puntual de predecir el clima en la antigüedad se daba para la observación de eventos de patrones, también determinados como patrones de reconocimiento. Por ejemplo, usualmente cuando en la puesta de sol el cielo se encuentra particularmente rojo, el siguiente día

trae un buen clima; era una creencia popular en la antigüedad que no siempre acertaba con la predicción de la lluvia. Con la invención del telégrafo en 1837, avanzó el estado al reporte del clima, debido a que con el telégrafo se podía llegar a zonas más alejadas avisando del comportamiento del clima, así inició el reporte del clima conociendo las condiciones regionales (Wiki, 2011).

Hoy en día existen diferentes métodos, técnicas y hasta aparatos electrónicos que buscan predecir el clima en el menor margen de error, algunos de ellos son muy eficientes y otros no tanto, debido al porcentaje en el que acierta la minería de datos en la predicción del clima. Esta técnica se pone en práctica debido a que se puede obtener información válida para la predicción de forma sistemática. Sin embargo, se pueden encontrar ventajas y desventajas; las ventajas pueden ser que el usuario que desee utilizar el software de predicción del clima, lo pueda hacer de manera fácil y sencilla, y desde luego el software provea de información oportuna y necesaria al usuario. Por otro lado, las desventajas se podrían evidenciar con los resultados que no pueden ser siempre precisos y es necesario información previa para que el software pueda encontrar un patrón para lograr predecir el clima (WFUM, 2015).

Conclusiones

En resumen, las previsiones meteorológicas son cada vez más precisas y útiles, y sus beneficios se extienden ampliamente en ramas como la agricultura, riesgos climáticos, catástrofes, entre otras. Aunque se ha conseguido avances significativos en las previsiones meteorológicas, todavía queda mucho por mejorar.

La comunidad de pronósticos está trabajando estrechamente con múltiples partes interesadas para asegurar que las previsiones y advertencias satisfa-

gan sus necesidades específicas. Simultáneamente, están desarrollando nuevas tecnologías y redes de observación que pueden mejorar la habilidad del pronosticador y el valor de sus servicios para sus usuarios (AMS, 2015).

Referencias bibliográficas

Agrawal R, Srikant R (1994) Fast algorithms for mining association rules. En: Proceedings of the 20th VLDB conference, pp 487–499.

Álvaro Galán Sánchez & Julio Daniel Pérez (2012). Predicción meteorológica, Inteligencia en Redes de Comunicación, la Universidad de Carlos III de Madrid. En: <http://www.it.uc3m.es/jvillena/irc/practicas/05-06/4mem.pdf>

AMS council (2015). Weather Analysis and Forecasting, American Meteorological Society, adopted by AMS council. En: <https://www.ametsoc.org/ams/index.cfm/about-ams/ams-statements/statements-of-the-ams-in-force/weather-analysis-and-forecasting/>

Ashwini Mandale & Jadhawar B.A, Assistant professor, Dr.Daulatrao (2015). Weather forecast prediction a Data Mining application, Aher College of engg. En: <http://pnrsolution.org/Datacenter/Vol3/Issue2/176.pdf>

Avalos Flores Imelda (2010). Definición de algoritmos. En: <http://correo.uan.edu.mx/~iavalos/FP/FP1.html>

Fahad Sheikh S. Karthick, D. Malathi, J. S. Sudarshan and C. Arun (2016). Analysis of Data Mining Techniques for Weather Prediction, Department of Software Engineering, SRM University, Kattankulathur. En: <http://www.indjst.org/index.php/indjst/article/view/101962/74214>

Forolunsho Olaiya & Adesesan Barnabas Adeyemo (2012). Application of Data Mining Techniques in Weather Prediction and Climate Change Studies, University of Ibadan. En: <http://www.mecs-pres.org/ijieeb/ijieeb-v4-n1/IJIEEB-V4-N1-7.pdf>

Han, J.; Kamber M. (2001). Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, USA.

Hincapié Luisa (2011). Minería de Datos para el análisis de Datos Meteorológicos, Universidad Nacional de Colombia Sede Manizales. En: <http://www.docentes.unal.edu.co/morozcoa/docs/Duque2011.pdf>.

IDEAM (2014). Red de estaciones IDEAM en Colombia, Solicitud de información. En: <http://www.ideam.gov.co/solicitud-de-informacion>

Jesús Sáez Gómez Escalonilla (2013). Minería de Datos Predicción Meteorológica, 5º Curso, Inteligencia en Redes de Comunicación, la Universidad de Carlos III de Madrid. En: <http://www.it.uc3m.es/jvillena/irc/practicas/05-06/6mem.pdf>

Nevon Projects Software (2015). Weather Forecasting Using Data Mining, Nevon Projects Software & Embedded Kits. En: <http://nevonprojects.com/weather-forecasting-using-data-mining/>

Rafael Ancel Trueba (2013). Aportaciones de la redes bayesianas en meteorología, Predicción probabilística de predicción, Ministerio de agricultura, alimentación y medio ambiente, Gobernación España. En: http://www.aemet.es/documentos/es/conocermas/recursos_en_linea/publicaciones_y_estudios/publicaciones/redes_bayesianas/Predicc_probab_precipitac.pdf

Web Mining Consultores (2011). KDD: Proceso de Extracción de conocimiento, Web Mining Consultores, En: <http://www.webmining.cl/2011/01/proceso-de-extraccion-de-conocimiento/>

Wikipedia (2011). Historia de los pronósticos del tiempo, Pronostico del tiempo, Wikipedia. En: https://es.wikipedia.org/wiki/Pron%C3%B3stico_del_tiempo.

Xiang Li, Beth Plale, Nithya Vijayakumar, Rahul Ramachandran, Sara Graves, Helen Conover (2008). Real-time storm detection and weather forecast activation through data mining and events processing. En: <http://link.springer.com/article/10.1007/s12145-008-0010-7>

Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, Dan Steinberg (2007). Top 10 Algorithms in Data Mining, Knowl Inf Syst. En: <http://www.cs.uvm.edu/~icdm/algorithms/10Algorithms-08.pdf>

AUTORES

Luis Alejandro Cortés:

Estudiante último semestre de Ingeniería de Sistemas. Integrante Semillero Itmedia de la Institución Universitaria Antonio José Camacho, sede sur. E-mail: gibaleles@gmail.com

Beatriz Eugenia Marín:

Ingeniera de Sistema, con estudios de maestría en Sistemas de Información Geográficos Unigis. Integrante del grupo Grintic de la Institución Universitaria Antonio José Camacho. Coordinadora de Investigación Facultad de Ingeniería. E-mail: bmarin@admon.uniajc.edu.co