

Analysis of Algorithms Used by the BI Software

Análise de Algoritmos Usados pelo Software para BI

Análisis a los Algoritmos Utilizados por el Software para BI

Curtis Cooke

Oregon State University

Curtis.Cooke@oregonstate.edu

(Artículo de INVESTIGACIÓN. Recibido el 22/02/2011. Aprobado el 17/05/2011)

Abstract

The purpose of this work is to examine which are the algorithms that use the most popular packages of software for BI. An initial informal inquiry show that current BI software package including the majority of algorithms described in the literature of the traditional statistical packages. Also, one of the main differences between BI software and the traditional statistical packages is that the first one has better characteristics for reports.

Keywords: *Business intelligence, algorithms, statistical packages.*

Resumo

O objetivo deste trabalho é analisar quais são os algoritmos que usam os pacotes de software popular para BI. Um levantamento inicial mostra que informal pacotes de software de BI atuais incluem a maioria dos algoritmos descritos na literatura tradicional de pacotes estatísticos. Além disso, uma das principais diferenças entre o BI e software para as tradicionais pacotes estatísticos é que o primeiro tem características melhores para os relatórios.

Palavras-chave: *Business intelligence, algoritmos, pacotes estatísticos.*

Resumen

El propósito de este trabajo es examinar cuáles son los algoritmos que utilizan los paquetes de software más populares para BI. Una encuesta inicial informal muestra que los actuales paquetes de software de BI incluyen a la mayoría de algoritmos descritos en la literatura de los paquetes estadísticos tradicionales. Además, una de las principales diferencias entre el software para BI y los paquetes estadísticos tradicionales es que el primero tiene mejores características para los reportes.

Palabras clave: *Inteligencia de negocios, algoritmos, paquetes estadísticos.*

1. Introducción

Inteligencia de Negocios se ha convertido gradualmente en un término popular en los sistemas de información. Actualmente existe una variedad de paquetes de software de BI en el mercado a pesar de que en realidad son una combinación de minería de datos, análisis estadístico, y funciones avanzadas de presentación de reportes. La minería de datos busca patrones ocultos en los *data warehouse* por lo que puede ayudar a los gerentes a tomar decisiones de negocio. Para descubrir esos patrones ocultos, los paquetes de software de BI utilizan una variedad de algoritmos para determinar la relación entre los datos y las variables. Si bien la mayor parte de esos algoritmos parecen ser similares a las tradicionales técnicas estadísticas, las empresas de software están promocionando al software de BI en el mercado como una herramienta de decisión nueva. Por lo tanto, el propósito de este trabajo es examinar los algoritmos utilizados por el popular software de BI.

Los algoritmos utilizados por el software de BI incluyen el análisis de regresión, los árboles de decisión, la asociación y el análisis de *cluster*. Mediante el uso de reglas de asociación, crean también un análisis a la canasta de compras. Este análisis le permite a una empresa conocer los productos o servicios relacionados cuando un cliente compra un determinado producto. Por ejemplo, si un cliente compra un pasaje de avión, entonces es probable que alquile un auto y haga una reserva de

hotel. El término "*Business Intelligence*", se utilizó por primera vez por Gartner Group en 1990 [1]. Un estudio de la literatura muestra que en los últimos años se han publicado varios libros de texto y artículos acerca de esta temática [1], [2].

2. Algoritmos para BI

De acuerdo con algunos académicos [1], [3], los algoritmos utilizados por la minería de datos se pueden clasificar en tres niveles:

1. Algoritmos simples: SQL queries
2. Algoritmos intermedios: regression analysis, decision trees
3. Algoritmos complejos: neural network, orthogonal partitioning clustering

Microsoft SQL Server [4] lista los siguientes algoritmos para BI:

1. Algoritmos de clasificación: decision trees.
2. Algoritmos de regresión: time series.
3. Algoritmos de segmentación: clustering
4. Algoritmos de asociación
5. Algoritmos de análisis de secuencia

3. Software para BI

Software para BI incluye a la mayoría de programas ERP relacionados como SAP [5], Oracle [6] y Microsoft Dynamics [7]. Los paquetes tradicionales de software estadístico también proporcionan funciones de BI, por

ejemplo, un paquete tradicional de software estadístico SAS ahora incluye funciones para BI.

Tanto SAP R-3 como Oracle son software ERP para grandes corporaciones. SAP R-3 vincula su software a un paquete separado llamado Business Warehouse – BW–, que es el software para inteligencia de negocios también ofrecidos por SAP. Debido a su fortaleza en data warehousing, Oracle también ofrece complejos algoritmos de inteligencia de negocios. Microsoft que SQL Server 2005/2008 es un paquetes de software de base de datos para pequeñas y medianas empresas que tiene un componente llamado BI Development Studio. Además, Microsoft ofrece un software ERP llamado Microsoft Dynamics, que actualmente incluye varios paquetes de software ERP independientes: Great Plains GP, AX, SL, y NAV. La mayoría de ellos se conecta a los componentes externos que tienen características de BI. Estas características se centran en funciones de presentación de informes, y una de sus fortalezas es la capacidad para trabajar con Microsoft Excel y Dashboard. Si bien las definiciones de BI son diferentes, la mayoría de las compañías de software parecen creer que BI es un componente importante para el software ERP y de base de datos.

4. Preguntas y método de investigación

Las siguientes son las preguntas de investigación que se aplicaron:

- 1) ¿Cuáles son los algoritmos utilizados por el software para BI?
- 2) ¿Cuáles son los algoritmos utilizados por los paquetes tradicionales de software estadístico?
- 3) ¿Qué es un algoritmo estadístico?

Estas preguntas fueron contestadas mediante un examen al actual software de BI. Para las preguntas 2 y 3, se realizó una encuesta a estudiantes, con experiencia de trabajo en negocios, y a instructores. En segundo lugar, se verificó si los paquetes de BI tenían la capacidad de integrar diferentes fuentes de datos.

5. Resultados

Se examinaron tres paquetes de software diferentes para BI: SAP, SQL Server y Oracle. La Tabla 1 muestra los algoritmos que ofrecen estos paquetes.

Una encuesta inicial informal muestra que la mayoría de algoritmos descritos en la literatura han sido utilizados por el software de BI, y que muchos de esos algoritmos son similares para el análisis estadístico tradicional. Una de las principales diferencias entre el software para BI y los paquetes estadísticos tradicionales es que parece ser que el software de BI tiene mejores características para los informes.

Un algoritmo estadístico es aquel que utilizan las propiedades estadísticas inscritas en la fuente para mejorar su codificación. Su objetivo es aprovechar la información redundante de la fuente para conseguir la compresión de datos o la información necesaria para ofrecer resultados a las consultas [8]. Los algoritmos estadísticos más populares son los de Huffman, de Shannon-Fano, los Aritméticos y los Predictivos.

TABLA 1
Comparación de los paquetes de software para BI

Algoritmo	SAP BW	Microsoft SQL Server	Oracle (BI Suite)
Decision Tree	X	X	X
Linear Regression	X	X	X
Naïve Bayes		X	X
Clustering	X	X	
Association Rules (market basket analysis)	X	X	
Sequence Clustering	X	X	
Time Series	X	X	
Neural Network	X	X	
Logistic Regression	X	X	
Support Vector machine			X
Enhanced k-Means Clustering			X
Orthogonal Partitioning Clustering			X
Nonnegative Matrix Factorization			X

Fuente: Autor

REFERENCIAS

- [1] E. Turban, R. Sharda, J. E. Aronson, & D. King. "Business Intelligence: A Managerial Approach". London: Prentice Hall, 284 p. 2007.
- [2] B. Larson. "Delivering Business Intelligence with Microsoft SQL Server 2008". USA: McGraw Hill, 792 p. 2008.
- [3] H. R. Nemati & C. D. Barko. "Issues in organizational data mining: A survey of current practices". *Journal of Data Warehousing*, Vol. 6, No. 1, pp. 25-36. 2001.
- [4] Microsoft. [http://msdn.microsoft.com/en-us/library/ms175595\(SQL.90\).aspx](http://msdn.microsoft.com/en-us/library/ms175595(SQL.90).aspx). [Dec. 2010].

- [5] K. McDonald, A. Wilmsmeier, D. C. Dixon, W. H. Inmon. "Mastering the SAP Business Information Warehouse". USA: Wiley, 516 p. 2002.
- [6] Oracle. Oracle Database 11g for Data Warehousing and Business Intelligence, <http://www.oracle.com/us/solutions/datawarehousing/039574.pdf>, [Nov. 2010].
- [7] Microsoft. http://www.microsoft.com/dynamics/gp/product/business_intelligence.msp, [Nov. 2010].
- [8] R. B. McCall. "Fundamental Statistics for the Behavioral Sciences". New York: Harcourt Brace Jovanovich, 512 p. 1998. Ω