An Analysis of Algorithms Used by Business Intelligence Software

Kuo Lane Chen  
School of Accountancy and information Systems  
University of Southern Mississippi  
Hattiesburg, MS 39406  
Kuo.chen@usm.edu

Huei Lee  
Department of Computer Information Systems  
Eastern Michigan University  
Ypsilanti, MI 48197  
Tel: (734)487-4044  
Huei.lee@emich.edu

Chen-Chi Shing  
Information Technology Department  
Radford University  
Box 6933  
Radford, VA 24142  
cshing@radford.edu

Jiaqin Yang  
College of Business  
Georgia College & State University  
Milledgeville, Georgia 31061  
Tel: (912) 445-2573  
jiaqin.yang@gcsu.edu

Abstract: The purpose of this paper is to examine the algorithms used by popular BI software. An initial informal survey shows that BI software packages now include most of the algorithms described by the literature. Also, one of the major differences between BI and traditional statistical packages is that BI software has better report features.

INTRODUCTION

Business intelligence (BI) has gradually become a popular information systems terminology. There are a variety of BI software packages in the market today although it is actually a combination of data mining, statistical analysis, and advanced reporting features. Data mining searches for hidden patterns from a huge data warehouse so it can help managers to make business decisions. To discover the hidden pattern from a huge data warehouse, BI software packages use a variety of algorithms to find out the relationship between different data and variables. While most of the BI algorithms seem to be similar to traditional statistical techniques, software companies are enthusiastic to market BI software as a new decision tool. Therefore, the purpose of this paper is to examine the algorithms used by popular BI software.

The algorithms used by BI software include regression analysis, naive bay, decision tree, association, and cluster analysis. By using association rules, it also creates a market basket analysis. Market basket analysis allows a business to know the related products or services if a customer purchases a product. For example, if a customer purchases an airline ticket, then he is likely to rent a car and make a hotel reservation. The term, “Business Intelligence,” is first used by the Gartner Group in the 1990 (Turban et. al, 2008). A literature study shows there are many textbooks and articles in the last five years (Larson, 2009; Turban et. al., 2008).

BUSINESS INTELLIGENCE ALGORITHM

According to several scholars (Turban et. al., 2008; Nemati & Barko, 2001), the algorithms used by data mining can be classified as the three levels:
1) Simple algorithms: SQL queries
2) Intermediate algorithms: regression analysis, decision trees
3) Complex algorithm: neural network, and orthogonal partitioning clustering

Microsoft SQL Server (2009) lists the following algorithms for business intelligence:
1) Classification algorithms such as decision trees.
2) Regression algorithms such as time series.
3) Segmentation algorithms such as clustering.
4) Association algorithms.
5) Sequence analysis algorithms.

**BUSINESS INTELLIGENCE SOFTWARE**

BI software includes most ERP-related software such as SAP, Oracle, and Microsoft Dynamics. Traditional statistical software packages also provide BI functions. For example, a traditional statistical software package SAS now includes BI features.

Both SAP R/3 and Oracle are ERP software are for large corporations. SAP R/3 will link its software to a separate package call Business Warehouse (BW), which is the software for business intelligence also offered by SAP. Because its strength in data warehousing, Oracle also provides complex algorithms in business intelligence. Microsoft SQL Server 2005/2008 is the database software packages for small-size and medium-size companies. It has a component called Business Intelligence Development Studio. Figure 2 shows the screenshot for the decision tree model in BI Development Studio in SQL Server. In additional, Microsoft has offered ERP software, Microsoft Dynamics. Microsoft Dynamics actually includes several independent ERP software packages: GP (formally Great Plains), AX, SL, and NAV. Most of them have link to outside components which have BI features. These BI features focus on reporting functions. One of the strength is its ability to work with Microsoft Excel and Dashboard (see Figure 2). While the definitions for BI are different, most software companies seem to believe BI is an important component for ERP and database software.

![Decision Tree Algorithm in MS SQL Server 2005](source: Microsoft Handouts)
RESEARCH QUESTIONS AND RESEARCH METHOD

The following are the research questions which are interesting to us:

1) What algorithms are used by popular BI software?
2) What are popular algorithms used by business people?
3) Which one is not a traditional statistical algorithm?

Research question can be answered by examining current BI software. For questions 2 and 3, we will conduct a survey to the students and instructors. These students have working experiences in the business. Second, the BI packages have the ability to integrate different sources of data.

RESULTS

We have examined three different software packages: SAP, SQL Server, and Oracle. Table 1 shows the algorithms offered by these three different BI software packages.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>SAP BW</th>
<th>Microsoft SQL Server</th>
<th>Oracle (BI Suite)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision Tree</td>
<td>x</td>
<td>x</td>
<td>X</td>
</tr>
<tr>
<td>Linear Regression</td>
<td>x</td>
<td>x</td>
<td>X</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>x</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Clustering</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Association Rules (market basket analysis)</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Sequence Clustering</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Time Series</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Neural Network</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Logistic Regression Algorithm</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Support Vector machine</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Enhanced k-Means Clustering</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Orthogonal Partitioning Clustering</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Non-negative Matrix Factorization</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Table 1: Comparison of BI software packages
An initial informal survey shows that most algorithms described by the literature have been used by BI software. Many algorithms are similar to traditional statistical analysis. One of the major difference between BI and traditional statistical packages appears to be that BI software has better report features.

REFERENCES