Chapter XXII

A Framework for Organizational Data Analysis and Organizational Data Mining

Bernd Knobloch University of Bamberg, Germany

ABSTRACT

This chapter introduces a framework for organizational data analysis suited for datadriven and hypotheses-driven problems. It shows why knowledge discovery and hypothesis verification are complementary approaches and how they can be chained together. It presents a methodology for organizational data analysis including a comprehensive processing scheme. Employing a plug-in metaphor, data analysis process engineering is introduced as a way to set up data analysis processes based on taxonomies of tasks that have to be performed during data analysis and on the idea of re-using experience from past data analysis projects. The framework aims at increasing the benefits of data mining and other data analysis approaches, by allowing a wider range of business problems to be tackled and by providing the users with structured guidance for planning and running analyses.

Copyright © 2004, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

INTRODUCTION

The way from raw data to business intelligence is often long and difficult to take, and, sometimes, it does not even lead to where you are heading. Since the time when data mining became popular, lots of methods and algorithms have been proposed, enabling efficient mining of large data sets. However, there are more steps to take than merely putting an algorithm to work on some data. The process embracing all these steps is known as knowledge discovery in databases (KDD). Some KDD tool vendors have covered process issues and provide features for defining sequences of processing steps tailored to meet the specific needs of individual mining cases. However, some, if not most of these approaches ignore conceptual questions, such as which business goal is to be pursued, which data are necessary for producing the desired information, which aspects of data organization have to be taken into account, etc. The need for a human-centered, process-oriented view on knowledge discovery has long been neglected. Analysts need more support in understanding how to do knowledge discovery (Brachman & Anand, 1996). As Smyth (2001) puts it, users "will almost certainly not say that they need a slightly more accurate decision tree algorithm, or a slightly faster association rule algorithm. Instead their most pressing problem is that of managing the whole process." It's well understood that methods are elementary, but they are of very limited use, if you don't know how to properly employ them to solve your business problems.

Moreover, discussion has almost always been restricted to data mining analyses. Data mining, however, is simply one of several approaches to data analysis. Focusing solely on data mining bears the risk of not applying the one analytical approach that fits the current business objective best. Why rule out the potentials of other approaches, if data mining is not the ultimate choice? Why not combine the strengths of different instruments for data analysis to obtain the best results possible? Why not think big and put the pieces of the puzzle together, constructing an overall framework for business intelligence?

This chapter introduces an organizational framework that helps mend some of the shortcomings mentioned above. That framework provides structured guidance to analysts for putting data analysis to work and covers different approaches to organizational data analysis. It may also serve as reference for documentation of project experience.

BACKGROUND

User support for data analysis is crucial in terms of efficiency. Data analysis is profitable only if the return gained from newly discovered insights is higher than the costs produced by the analysis effort. The efficiency issue has received surprisingly little attention from KDD research. However, some progress has been made.

Brachman and Anand (1996) investigate the interactions between humans and data during analysis, which are important for development of knowledge discovery support tools. Knobloch and Weidner (2000) have taken a problem-oriented approach to infer requirements for analysis tools that help reduce inefficient iterations in the analysis process.

Skalak (2001) suggests three areas for additional research in the data mining field to make data analysis processes more efficient. These areas are modeling the mining

Copyright © 2004, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: <u>www.igi-</u> <u>global.com/chapter/framework-organizational-data-analysis-</u> organizational/27925

Related Content

Issues in Feathers Application in the Seoul Metropolitan Area

Won Do Lee, Chang-Hyeon Joh, Sungjin Choand Bruno Kochan (2014). *Data Science and Simulation in Transportation Research (pp. 71-85).* www.irma-international.org/chapter/issues-in-feathers-application-in-the-seoul-metropolitanarea/90066

Extending LINE for Network Embedding With Completely Imbalanced Labels

Zheng Wang, Qiao Wang, Tanjie Zhuand Xiaojun Ye (2020). *International Journal of Data Warehousing and Mining (pp. 20-36).* www.irma-international.org/article/extending-line-for-network-embedding-with-completelyimbalanced-labels/256161

A Survey of Open Source Tools for Business Intelligence

Christian Thomsenand Torben Bach Pedersen (2009). International Journal of Data Warehousing and Mining (pp. 56-75).

www.irma-international.org/article/survey-open-source-tools-business/3896

Mining Partners in Trajectories

Diego Vilela Monteiro, Rafael Duarte Coelho dos Santosand Karine Reis Ferreira (2020). *International Journal of Data Warehousing and Mining (pp. 22-38).* www.irma-international.org/article/mining-partners-in-trajectories/243412

Research on Fraud Detection Method of Financial Data of Listed Companies Based on HMCRAN

Fubing Liand Huaqiong Duan (2024). *International Journal of Data Warehousing and Mining (pp. 1-29).*

www.irma-international.org/article/research-on-fraud-detection-method-of-financial-data-oflisted-companies-based-on-hmcran/356510