



IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-<http://www.idea-group.com>

ITB9387

Chapter IX

Mining Meaning: Extracting Value from Virtual Discussions

William L. Tullar
University of North Carolina at Greensboro, USA

ABSTRACT

This chapter focuses on the pattern detection and extraction step in text data commonly called text data mining. I examine some of the literature on natural language processing and propose a method of recovering value from the text of virtual group discussions based on methods derived from the communication field. Then, I apply the method in a case using data from 216 different groups from a virtual group experiment. The results from the case show that higher performing groups are characterized by higher frequencies of acts of dominance and higher frequencies of terms concerning cognition, communication and praise. Higher performing groups were also characterized by lower frequencies of acts of equivalence and lower frequencies of leveling terms and numerical terms. Ways to use this knowledge to improve the groups' performance are discussed.

INTRODUCTION

Text mining may be considered a subspecialty of the broader domain of Knowledge Discovery from Data. This in turn can be defined as the computational process of extracting useful information from huge amounts of data by mapping low-level data into richer, more abstract forms and by detecting meaningful patterns implicitly present in the

data. Knowledge Discovery from Data is typically conducted on structured, relational databases, and data mining is one of its subtasks. When the data are ill-structured text data, the data mining process becomes somewhat more difficult. Moreover, data mining itself is only one of the steps within the Knowledge Discovery from Data process. The full process usually includes data storage and access, data cleansing, pattern detection and extraction and data interpretation. Data mining refers more narrowly to the particular step of applying specific algorithms for detecting and extracting patterns (Liddy, 2000).

This chapter focuses only on the pattern detection and extraction step in text data. It proposes one method of recovering value from the text of virtual group or team discussions and deliberations. Thus, mining meaning in the text of group discussions requires an approach that is iterative between human and machine. It is my purpose here to outline a strategy for recovering valuable meaning from this data.

MINING TEXT DATA

Large text databases potentially contain a great wealth of knowledge. However, text represents factual information (and information about the author's communicative intentions) in a complex but opaque manner. Unlike numerical and fixed field data, it cannot be analyzed by standard statistical data mining methods. Relying on human analysis results in either huge workloads or the analysis of only a tiny fraction of the database (Nasukowa & Nagano, 2001). Mining text data is the science and art of extracting meaningful factual information from masses of text, usually by means other than the statistical approaches that have produced value in numerical and fixed-field data.

Many of the most successful text mining systems are focused on document retrieval. But in order to begin extracting the full value out of text, a truly useful system must go beyond simple retrieval. It must provide a broad range of information access and analytic capabilities. The way that text mining can accomplish this goal is through reliance on Natural Language Processing (NLP). NLP consists of a range of computational techniques for analyzing and representing naturally occurring texts at all levels of linguistic analysis to achieve human-like language processing that can support this kind of analysis. A fully featured text miner should be one that combines both information retrieval and text mining capabilities (Liddy, 2000). Such a technology would:

- detect the specific sources that contain information worth mining;
- recognize and extract meaningful entities that convey valuable knowledge;
- produce a semantic interpretation of the information;
- store the semantically interpreted information in an efficient data structure; and
- provide means for easy access and utilization of this knowledge base for new insights or for utilization in decision-making tasks.

Several firms (e.g., IBM and SAS) are hard at work at producing text miners that can do all these things. SAS has recently acquired the rights to use Inxight's LinguistX Platform, a natural-language text solution for analyzing words, phrases and sentences, and Inxight's Thing Finder, which identifies and extracts from documents key content, such as company names, products, people, addresses and dates. SAS offers a product

14 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: www.igi-global.com/chapter/mining-meaning-extracting-value-virtual/27912

Related Content

A Lattice-Based Framework for Interactively and Incrementally Mining Web Traversal Patterns

Yue-Shi Lee (2009). *Handbook of Research on Text and Web Mining Technologies* (pp. 448-467).

www.irma-international.org/chapter/lattice-based-framework-interactively-incrementally/21740

On-Demand ELT Architecture for Right-Time BI: Extending the Vision

Florian Waas, Robert Wrembel, Tobias Freudenreich, Maik Thiele, Christian Koncilia and Pedro Furtado (2013). *International Journal of Data Warehousing and Mining* (pp. 21-38).

www.irma-international.org/article/demand-elt-architecture-right-time/78285

A Survey of Open Source Tools for Business Intelligence

Christian Thomsen and 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

Understanding the SNN Input Parameters and How They Affect the Clustering Results

Guilherme Moreira, Maribel Yasmina Santos, João Moura Pires and João Galvão (2015). *International Journal of Data Warehousing and Mining* (pp. 26-48).

www.irma-international.org/article/understanding-the-snn-input-parameters-and-how-they-affect-the-clustering-results/129523

Big Data Warehouse Automatic Design Methodology

Francesco Di Tria, Ezio Lefons and Filippo Tangorra (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 454-492).

www.irma-international.org/chapter/big-data-warehouse-automatic-design-methodology/150179