Chapter VI

Text Mining in Business Intelligence

Dan Sullivan, The Ballston Group, USA

ABSTRACT

As the demand for more effective Business Intelligence (BI) techniques increases, BI practitioners find they must expand the scope of their data to include unstructured text. To exploit those information resources, techniques such as text mining are essential. This chapter describes three fundamental techniques for text mining in business intelligence: term extraction, information extraction, and link analysis. Term extraction, the most basic technique, identifies key terms and logical entities, such as the names of organizations, locations, dates, and monetary amounts. Information extraction builds on terms extracted from text to identify basic relationships, such as the roles of different companies in a merger

or the promotion of a chemical reaction by an enzyme. Link analysis combines multiple relationships to form multistep models of complex processes such as metabolic pathways. The discussion of each technique includes an outline of the basic steps involved, characteristics of appropriate applications, and an overview of its limitations.

INTRODUCTION

Traditionally, business intelligence has focused on analyzing data gathered from transaction processing systems, such as enterprise resource planning (ERP), customer relationship management (CRM), sales force automation (SFA), claims processing, and other structured data sources. Structured data sources implement well-defined, but relatively limited data models. Relational and object-oriented databases are commonly used to implement these models. Somewhat less structured and more flexible are semi-structured data sources, such as XML-based models. Like structured data models, semi-structured models have well-defined structures that are relatively limited in scope but allow more flexibility with the range and ordering of data elements. At the far end of the data structure spectrum is unstructured data. As the name implies, there is no formal schema for such data. Free-form text, audio, and video are the most common forms of unstructured data. (The term "unstructured" is something of a misnomer when referring to text since language is highly structured according to linguistic principles, but the term is widely used nonetheless, and convention will be followed here.) Unstructured data is abundant in most organizations but to date has not been tapped as a source of business intelligence.

Recent advances in computational linguistics as well as Web and enterprise search make the integration of unstructured data into a business intelligence infrastructure feasible and effective. Together, these advances are broadly considered text mining, which is defined as analysis of natural language text to extract key terms, entities, and relationships between those terms and entities. These extracted elements are used for several purposes, including:

- categorizing and classify documents;
- generating summaries of texts;
- providing data for visualization tools for navigating large text databases; and
- mapping multistep relationships between series of entities.

11 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/text-mining-business-intelligence/6067

Related Content

Stock Markets in Changing Times: A Study of Select Global Market Indices and Indian Banks
Meru Sehgaland Shruti Gupta (2021). *International Journal of Business Analytics (pp. 14-25).*
www.irma-international.org/article/stock-markets-in-changing-times/279627

Profit Maximization Modeling for Supply Chain Planning
Tan Millerand Renato de Matta (2014). *Encyclopedia of Business Analytics and Optimization (pp. 1910-1921).*
www.irma-international.org/chapter/profit-maximization-modeling-for-supply-chain-planning/107379

Stock Market Responses to the COVID-19 Health Crisis: Evidence From the World's Largest Economies
Abdul Wajidand Kanishka Gupta (2022). *International Journal of Business Analytics (pp. 1-19).*
www.irma-international.org/article/stock-market-responses-to-the-covid-19-health-crisis/303114

A Timeline Optimization Approach of Green Requirement Engineering Framework for Efficient Categorized Natural Language Documents in Non-Functional Requirements
K. Mahalakshmi, Udayakumar Allimuthu, L Jayakumarand Ankur Dumka (2021). *International Journal of Business Analytics (pp. 21-37).*
www.irma-international.org/article/a-timeline-optimization-approach-of-green-requirement-engineering-framework-for-efficient-categorized-natural-language-documents-in-non-functional-requirements/269485

Application of Six Sigma to Achieve Sales Lifecycle Optimization
Victoria Nacarelli (2014). *International Journal of Business Intelligence Research (pp. 1-13).*
www.irma-international.org/article/application-of-six-sigma-to-achieve-sales-lifecycle-optimization/126894