

Chapter 8

Text–Driven Reasoning and Multi–Structured Data Analytics for Business Intelligence

Lipika Dey

Innovation Labs, Tata Consultancy Services, India

Ishan Verma

Innovation Labs, Tata Consultancy Services, India

ABSTRACT

Business Intelligence (BI) refers to an organization's capability to gather and analyze data about business operations and transactions in order to evaluate its performance. The abundance of information both within the enterprise and outside of it has necessitated a change in traditional Business Intelligence practices. There is a need to exploit heterogeneous resources. Text data like news, analyst reports, etc. helps in better interpretation of business data. In this chapter, the authors present a futuristic BI framework that facilitates acquisition, indexing, and analysis of heterogeneous data for extracting business intelligence. It enables integration of unstructured text data and structured business data seamlessly to generate insights. The authors propose methods that can help in extraction of events or significant happenings from both unstructured and structured data, correlate the events, and thereafter reason to generate insights. The insights extracted could be validated as cause-effect pairs based on the statistical significance of co-occurrence of events.

INTRODUCTION

In today's knowledge-driven economy organizations are heavily dependent on the Business Intelligence (BI) tools to collect, analyze, and disseminate the right information at the right time to the right people so that the decision makers can make informed decisions. Business decisions

today depend not only on numerical data locked in the Enterprise database, but to a large extent on the data that is out there on the Web. The Internet delivers News about an event almost as soon as it occurs all over the world. It is critical for successful business organizations to be on top of this Information high-way. Social Media presence is as integral to business today as other

DOI: 10.4018/978-1-4666-6477-7.ch008

channels of customer touch-points. The global nature of economy has also made it imperative for the organizations to be aware of all political, financial and social events happening across the world. The Web is not only a tacit store-house of knowledge about potential consumers; but also about competitors, suppliers, partners and all other stakeholders.

Most of the Web data is unstructured in nature. More precisely, most of it is in the form of unstructured text. Unstructured data exist inside the enterprise also in the form of customer call-logs, emails, comments on the discussion forum etc. Advances in Big Data technologies have ensured that organizations need not to be concerned about the availability and harnessing of large volumes and varieties of data that are potentially available for Business Intelligence applications. The last few years have also seen considerable initiative to use textual information gathered from Social Media sites, emails, surveys, call-logs etc. for gaining Business Intelligence. In 2011, the text analytics market was estimated to be growing at the rate of \$1 billion annually by text mining enthusiasts¹. Though it has not really shown that kind of phenomenal growth it has not been far behind also. A survey shows that it is likely to touch around \$2 billion by the end of 2014¹. Analysis shows that the retardation is primarily caused by the disconnect that exists between the text mining applications and the traditional Business Intelligence platforms.

Text mining applications are currently being used by organizations to gain increased consumer insights, improve customer satisfaction, understand performance bottlenecks and also gain new product ideas. Mining customer communications like emails, discussion forums, social media pages or traditional surveys have been adopted by all organizations to understand customer issues with products or services and address them appropriately to gain increased customer satisfaction. Analysis of call-center transcripts or support center emails also provide insights into

the problem solving activities and thereby help in improving efficiency and performance parameters like response time, customer satisfaction etc. Understanding the unified persona of the customer through consumer-generated content independent of the product or service has also gained momentum. The unified persona of a customer is used to push targeted and relevant advertisements for maximum business benefit. Consumer profiling also helps in identifying potential future consumers and generate new product or service ideas.

The two categories of text mining based business intelligence applications described above use considerably different technologies. The business contexts for using these applications are also very different. Understanding and addressing customer problems from consumer-generated text is the concern of customer support divisions. Text-mining applications employed for analyzing customer complaints necessarily depend on availability of domain knowledge to be successful. Gauging public sentiment on social media with respect to an event, product or service is somewhat more generic in nature. Applications built for understanding consumer profiles are also quite generic. There are many off-the-shelf tools available for understanding consumer profiles and computing public sentiments. Placing contextual advertisements for maximum business benefits is also available as a service. We provide a brief overview of the core text mining technologies needed for these applications along with a few successful commercial and academic implementations.

In this chapter, our aim is not to look at the existing and operative text mining techniques used for business applications only. Our aim is rather to analyze and bridge the gap that exists today between the use of text mining for business intelligence and the traditional business intelligence platforms. As mentioned earlier, though text mining based business intelligence applications have achieved some success; these are yet to be integrated into the traditional Business Intelligence platforms. Thus the knowledge extracted from

29 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-driven-reasoning-and-multi-structured-data-analytics-for-business-intelligence/116813

Related Content

Logical Inference and Defeasible Reasoning in N-tuple Algebra

Boris Kulik, Alexander Fridman and Alexander Zuenko (2013). *Diagnostic Test Approaches to Machine Learning and Commonsense Reasoning Systems* (pp. 102-128).

www.irma-international.org/chapter/logical-inference-defeasible-reasoning-tuple/69406

A Cached-Based Stream-Relation Join Operator for Semi-Stream Data Processing

M. Asif Naeem, Imran Sarwar Bajwa and Noreen Jamil (2016). *International Journal of Data Warehousing and Mining* (pp. 14-31).

www.irma-international.org/article/a-cached-based-stream-relation-join-operator-for-semi-stream-data-processing/168484

Discovering Frequent Embedded Subtree Patterns from Large Databases of Unordered Labeled Trees

Yongqiao Xiao and J. F. Yao (2005). *International Journal of Data Warehousing and Mining* (pp. 70-92).

www.irma-international.org/article/discovering-frequent-embedded-subtree-patterns/1752

Computing Join Aggregates Over Private Tables

Rong She, Ke Want, Ada Waichee Fu and Xu Yabo (2008). *International Journal of Data Warehousing and Mining* (pp. 22-41).

www.irma-international.org/article/computing-join-aggregates-over-private/1816

Classification of Failures in Photovoltaic Systems using Data Mining Techniques

Lucía Serrano-Luján, Jose Manuel Cadenas and Antonio Urbina (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 1347-1366).

www.irma-international.org/chapter/classification-of-failures-in-photovoltaic-systems-using-data-mining-techniques/150220