

Discovering Business Intelligence from the Subjective Web Data

Ranjit Bose, University of New Mexico, USA

ABSTRACT

The online word-of-mouth behavior that exists today in the Web represents new and measurable sources of information. The automated discovery or mining of consumer opinions from these sources is of great importance for marketing intelligence and product benchmarking. Techniques are now being developed to effectively and easily mine the consumer opinions from the Web data and to timely deliver them to companies and individual consumers. This study investigates this emerging field named 'opinion mining' in terms of what it is, what it can do, and how it could be used effectively for business intelligence (BI). A rigorous review of the research literature on opinion mining is conducted to explore its current state, issues and challenges for its use in developing business applications for competitive advantage. The study aims to assist business managers to better understand the current opportunities and challenges in using opinion mining for deriving BI. Future research directions for further development of the field are also identified.

Keywords: Business Intelligence, Competitive Advantage, Consumer Opinions, Opinion Mining, Sentiment Analysis

1. INTRODUCTION

Consumer opinions used to be very difficult to obtain before the Web was widely available to and used by every day citizens. To find consumer opinions about their products and those of their competitors, companies usually conducted consumer surveys or engaged external consultants. Getting consumer feedback meant asking them to participate in filling out survey instruments. This cumbersome approach had many problems such as making a survey for each product or feature; the format, distribution and timing of

the survey (asking to send a form right after purchase might not be very informative); and the reliance on the goodwill of consumers to take the survey. Additionally, consumers shared their views primarily through spoken word of mouth. Thus, irrespective of how the consumer opinions were collected, they lead to a long period of latency before a brand's weakness or strengths were known to all – the producers and consumers.

With the rapid expansion of e-commerce in recent years, more and more products are being sold on the Web, thus more and more consumers are buying products online. In recent years, the Web has dramatically changed the way consum-

DOI: 10.4018/jbir.2011100101

ers express opinions on products that they have purchased and used, or on services that they have received from various companies. Opinions and reviews are easily posted on the Web, such as in merchant sites, review portals, blogs, internet forums, and social networking sites (Wang *et al.*, 2009). These data are commonly referred to as user-generated content or user-generated media. Both the product manufacturers as well as the potential customers find this online “word of mouth” very useful (Chen *et al.*, 2008). The product manufacturers receive information on their customers’ likes and dislikes, as well as the positive and negative comments on their products whenever available, giving them better knowledge of their products’ limitations and advantages over their competitors. The potential customers on the other hand receive useful and “first hand” information on the products and/or services which assist them in their future purchase decision making process.

As more and more common users are becoming comfortable with the Web, an increasing number of them are writing reviews. Consequently, the number of reviews that a product receives is growing rapidly. Popular products can get hundreds or even thousands of reviews at some large merchant sites. Many of these reviews are long and have only a few sentences containing opinions on the product. This makes it hard for a potential customer to read them all to make an informed decision on whether to purchase the product. If the potential customer only reads a few reviews, they may get a biased view. Similarly, the large number of reviews also makes it hard for product manufacturers to keep track of customer opinions of their products. Moreover, the product manufacturer face additional difficulties since many merchant sites may sell the same product and the manufacturer normally produces many kinds of products.

The online word-of-mouth behavior that exists today in the Web represents new and measurable sources of information. Mining these consumer opinions is of great importance for marketing intelligence and product benchmarking. Techniques are now being developed

to effectively and easily mine the consumer opinions from these sources and to timely deliver them to companies and individual consumers. In the past few years many researchers have been actively involved in developing this nascent and highly potent field of study, which is called *opinion mining* or *sentiment analysis* (Pang *et al.*, 2008b; Liu, 2010).

Opinion mining is a very new field of study that attempts to develop automatic systems to determine human opinion from text written in natural language. It is a discipline at the crossroads of information retrieval and computational linguistics. The task of opinion mining is technically challenging because it requires natural language processing, which itself is a tedious job (Kao *et al.*, 2007). The research and development that has taken place and is currently taking place in this field have been performed by the computer science and computational linguistics communities. Consequently, the published results are not friendly to or are easily understood by business professionals unless they are technically trained. The purpose and contribution of this research is twofold. First, it summarizes and presents the state of the art in opinion mining in a way that is easily comprehensible by business professionals encompassing a wide range of backgrounds or expertise. Second, it categorizes the existing literature into a classification framework that is intended to help management information systems researchers to study this field in further details for future business benefits.

The textual information that is found in the Web is broadly classified into two main categories, *facts* and *opinions*. Facts are objective statements about entities and events in the world. Opinions are subjective statements that reflect peoples’ sentiments or perceptions about the entities and events. Much of the existing research on text information processing has been, almost exclusively, focused on mining and retrieval of factual information, e.g., information retrieval, Web search, and many other text mining and natural language processing tasks. Until only recently, work has begun on the processing of opinions. Yet, opinions are so important that

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

www.igi-global.com/article/discovering-business-intelligence-subjective-web/60241?camid=4v1

This title is available in InfoSci-Journals, InfoSci-Journal Disciplines Business, Administration, and Management.

Recommend this product to your librarian:

www.igi-global.com/e-resources/library-recommendation/?id=2

Related Content

The Future Talent Shortage Will Force Global Companies to Use HR Analytics to Help Manage and Predict Future Human Capital Needs

Carey W. Worth (2011). *International Journal of Business Intelligence Research* (pp. 55-65).

www.igi-global.com/article/future-talent-shortage-will-force/60245?camid=4v1a

Measuring Effectiveness: A DEA Approach Under Predetermined Targets

Heinz Ahn and Ludmila Neumann (2014). *International Journal of Business Analytics* (pp. 16-28).

www.igi-global.com/article/measuring-effectiveness/107067?camid=4v1a

Scheduling of Extract, Transform, and Load (ETL) Procedures with Genetic Algorithm

Vedran Vrbani and Damir Kalpi (2015). *International Journal of Business Analytics* (pp. 33-46).

www.igi-global.com/article/scheduling-of-extract-transform-and-load-etl-procedures-with-genetic-algorithm/126832?camid=4v1a

Robust Supply Chain Risk Management

Amir H. Ansaripoor and Fernando S. Oliveira (2014). *Encyclopedia of Business Analytics and Optimization* (pp. 2093-2103).

www.igi-global.com/chapter/robust-supply-chain-risk-management/107396?camid=4v1a