Chapter 10

An Application of Text Mining to Capture and Analyze eWOM: A Pilot Study on Tourism Sector

Taşkın Dirsehan
Marmara University, Turkey

ABSTRACT

Marketing concept has progressed through different phases of evolution in the past. At the moment, customer relationship management is considered as the last era of marketing development. The main purpose of this approach is to build long-term oriented profitable relationships with customers. So, companies should know better their customers. This knowledge can be created through a deeper analysis of companies’ data with data mining tools. Companies which are able to use data mining tools will gain strong competitive advantages for their strategic decisions. Hotel industry is selected in this study, since it provides a warehouse of customer comments from which precious knowledge can be obtained if text mining as a data mining tool is used appropriately. Thus, this study attempts to explain the stages of text mining with the use of Rapidminer. As a result, different approaches according to the customer satisfaction/dissatisfaction are discussed to build competitive advantages.

INTRODUCTION

The rapid growth of web 2.0 applications directed consumers to communicate their experiences online and to read these reviews before choosing new experiences. By this way, huge volumes of textual data —called eWOM— are created by them and these data keep precious information for businesses. A useful tool to reach this kind of information from large data sets is data mining. When the data is in textual form, text mining as an extension of data mining can help them to uncover hidden patterns existing in large data sets. Moreover, these hidden patterns may lead a company to create unique competitive advantages in today’s information era.

DOI: 10.4018/978-1-4666-9449-1.ch010
Without a strong data analysis tool, an abundance of data is still poor in information. Data gathering over time are like raw ore, and the gap between data and information requires systematic data mining tools to pan that ore for gold (Han & Kamber, 2001).

The most common functions of data mining are association rules, artificial neural networks, decision trees, and clustering analysis. Researchers may gather significant information from raw data by using these functions, gaining competitive advantage from it. For instance, new customer segments may be explored and studied with data mining techniques (Köktürk and Dirsehan, 2012).

On the other hand, as a data mining tool, text mining enumerates words and number of repeats in a basic view. This methodology is appropriate for texts in the online environments, and especially for electronic word-of-mouth (eWOM).

Consumers perceive WOM as more trustworthy and persuasive than traditional media (Cheung & Thadani, 2012). The advent of the Internet has given birth to eWOM. Consumers post their opinions, comments and reviews on weblogs, discussion forums, websites and social networks (Cheung & Lee, 2012). In the tourism industry, Jalilvand et al. (2012) indicate that eWOM positively influences the destination image, tourist attitude and travel intention. Moreover, increasing consumer opinion portals on travel services shows that eWOM will continue to play an ever more vital role in travelers’ purchasing decisions. Thus, eWOM becomes an important focus of research in marketing, e-commerce, and e-tourism studies (Filieri & McLeay, 2013). The shift of WOM to the electronic environment has made it also easier for companies to reach raw data. Tweets, online reviews, and blogs have become “Big Data” sources of real sharing behavior (Berger, 2014). Content of texts in such communication is a challenge for researchers to extract meaningful information from the unstructured data to enhance customer relationship management and marketing (Leong et al., 2004). Since a company may know its customers by revealing the knowledge from these data sets.

This study aims to show the application of text mining to analyze eWOM in a fundamental way. A practical tool for this purpose is RapidMiner, which can be downloaded from the website rapidminer.com freely, with some limitations. Among many functions, the “Text Processing” application has been used in this study. The comments have been retrieved from “booking.com”, one of the famous hotel booking websites. All the steps have been explained in details with screenshots of RapidMiner.

After dividing the comments in two (high satisfaction levels and low satisfaction levels) according to the stars of visitors, the keywords determined for the tourism sector such as “price” have been analyzed for these groups. At the end, a general guideline is provided for researchers and practitioners, to enable them to perform basic text mining by themselves.

BACKGROUND

Text Mining as an Innovative Research Tool

Technological advancements enable new techniques in marketing research. In today’s digital information environment, customer data are recorded and stored in digital data warehouses. A data warehouse is defined by Hair et al. (2009) as a “logical aggregation of information stored in a single location.” A digital one can be defined as an aggregation of customer information collected in digital environments stored in a digital location. The purpose of a single data storage area is that managers may use such data for further analysis and decision making. This process of analyzing data to extract patterns and