Chapter 15

Mining Big Data for Marketing Intelligence

Khadija Ali Vakeel
Indian Institute of Management Indore, India

ABSTRACT

This chapter elaborates on mining techniques useful in big data analysis. Specifically, it will elaborate on how to use association rule mining, self-organizing maps, word cloud, sentiment extraction, network analysis, classification, and clustering for marketing intelligence. The application of these would be on decisions related to market segmentation, targeting and positioning, trend analysis, sales, stock markets and word of mouth. The chapter is divided in two sections of data collection and cleaning where we elaborate on how twitter data can be extracted and mined for marketing decision making. Second part discusses various techniques that can be used in big data analysis for mining content and interaction network.

INTRODUCTION

Big data has transformed the way people interact with each other. It not only impacts the interaction between company and consumer but also between consumers. Big data can be analyzed to take effective decisions by the company for successful marketing strategy. In this chapter, we elaborate on techniques that could be effectively used in textual mining data for marketing intelligence. We explain the complete process of big data mining from data extraction, data cleaning, analysis, and visualization. Specifically, we would elaborate on how to use association rule mining, self-organizing maps, word cloud, sentiment extraction, network analysis, classification, and clustering for marketing intelligence. The application of these would be on decisions related to market segmentation, targeting and positioning, trend analysis, sales, stock markets and word of mouth.

Association rule mining on transactions help us show given X and Y item in the itemset, X leads to Y. Apriori algorithm is used to find frequent itemsets that occur together in transactions given minimum support and confidence.

DOI: 10.4018/978-1-5225-2031-3.ch015
Self-organizing maps or Kohonen maps are algorithms that embed the clusters in a low-dimensional space right from the beginning and proceed in a way that places related clusters close together in that space.

Word cloud is used to describe the importance of each word based on the frequency of the attributes occurring in the total corpus.

Sentiment analysis is performed to extract opinion and subjectivity knowledge from user generated content, formalize this knowledge discovery and analyze it for managerial implications. This type of knowledge discovery covers analyzing product user opinion, appraisals, attitudes, and emotions etc. related to entities, individuals, issues, events, topics, and their own attributes. Sentiment analysis involves techniques like machine learning, information retrieval, and natural language processing to process vast amounts of user generated content.

Topological models for community analysis help in detecting the structural properties of the social network. Average shortest path length, efficiency, clustering coefficient, degree distributions are used to derive the network characteristics. Different models such as random graph models, small-world model, and scale-free models are used for analysis.

Centrality analysis helps us in identifying key nodes of the network such that roles like leader, follower and bridge are discovered for efficient marketing strategy on a social media network. Three main centrality measurements named- degree, betweenness, and closeness are used.

Classification and clustering are a basic technique we explore Latent Dirichlet Allocation (LDA) for clustering. Topic models like LDA have previously been used for a variety of applications, including ad-hoc information retrieval, geographical information retrieval and the analysis of the development of ideas over time in the field of computational linguistics.

To conclude, all these techniques have extensively been used in mining, this chapter would deal with how BI tools can be used in the context of marketing. In the following sections, a framework for big data mining for marketing research is elaborated.

DATA COLLECTION AND CLEANING

Taking an example of twitter, following section shows how data collection and cleaning can be done. Once the communities on the twitter have been identified, user data and tweet information can be collected by the twitter APIs like rest or stream. Every user has unique id through which his profile details can be extracted. The usage pattern of a twitter user can be either personal, group, aggregator, satire or marketing related (Cheong & Lee, 2009). Users can also be categorized as tourist, minglers, devotees and insiders (Kozinets, 2002). Tweet search is generally based on time period or is a keyword search (Twitter Developer). Data about location, user mentions, friends, and followers can also be extracted with the help of these API. Example for pattern detection, retweet, replied and trend keyword can be used (Cheong & Lee, 2009). Re-tweet are special posts where one twitter user forwards the tweet of another user, this kind of tweets start with RT. Similarly, reply to a tweet can be detected as a reply starts with @username where username is the twitter user whom another user is replying too. Hashtags are another feature which starts with #keyword such that it creates an index for the keyword. This user information tweets and metadata can be stored in a relational database for further analysis. Further, not all the extracted tweets might to useful or related to the marketing research question. Classifying tweets as primarily on topic or off topic or primarily social or informational will help to get tweets related to the scope of the question (Kozinets, 2002).
Related Content

A Strategic Analysis of Mixed Channel Structure: Retail Store Ownership
www.igi-global.com/article/a-strategic-analysis-of-mixed-channel-structure/124181?camid=4v1a

Optimize Short to Long Run Supply Chain Planning
www.igi-global.com/chapter/optimize-short-to-long-run-supply-chain-planning/107363?camid=4v1a

An Automatic User Interest Mining Technique for Retrieving Quality Data
www.igi-global.com/article/an-automatic-user-interest-mining-technique-for-retrieving-quality-data/176927?camid=4v1a

Social Media Mining: A New Framework and Literature Review
www.igi-global.com/article/social-media-mining/142781?camid=4v1a