Chapter 69
Social Big Data Mining:
A Survey Focused on Sentiment Analysis

Anisha P. Rodrigues
https://orcid.org/0000-0002-3050-4555
NMAM Institute of Technology, Nitte, India

Niranjan N. Chiplunkar
NMAM Institute of Technology, Nitte, India

Roshan Fernandes
NMAM Institute of Technology, Nitte, India

ABSTRACT

Social media is used to share the data or information among the large group of people. Numerous forums, blogs, social networks, news reports, e-commerce websites, and many more online media play a role in sharing individual opinions. The data generated from these sources is huge and in unstructured format. Big data is a term used for data sets that are large or complex and that cannot be processed by traditional processing system. Sentimental analysis is one of the major data analytics applied on big data. It is a task of natural language processing to determine whether a text contains subjective information and what information it expresses. It helps in achieving various goals like the measurement of customer satisfaction, observing public mood on political movement, movie sales prediction, market intelligence, and many more. In this chapter, the authors present various techniques used for sentimental analysis and related work using these techniques. The chapter also presents open issues and challenges in sentimental analysis landscape.

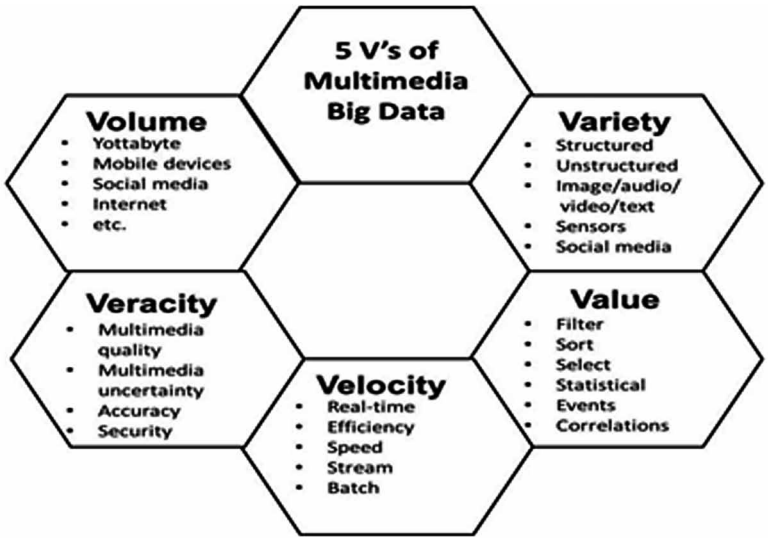
INTRODUCTION

These days use of smart devices and the high speed Internet has led to lots of people to engage in social media sites like Twitter, Facebook, and Instagram. Due to the high social interaction, the data produced by these sites increases drastically. The number of active social media users keeps growing. According

DOI: 10.4018/978-1-6684-6303-1.ch069

to the Global WebIndex statistics the number of people using mobile phones has reached 3.7 billion (Chaffey, D., 2016). According to Facebook, the number of active users has reached 1.59 billion (Tan, W., Blake, M. B., Saleh, I., & Dustdar, S., 2013). A large amount of data is produced by many Internet users while using online social networking media. The term big data is defined as a data with huge volume, complex in nature and inundates business on a day-to-day basis. Big Data is a term that describes data with "3V"s. They are Volume, Variety and Velocity. Volume signifies huge amount of data. Variety signifies various forms of data that is structured, semi-structured and unstructured generated from various sources. Velocity represents the speed at which the data generated. These characteristics were first identified by Doug Laney (Laney, D., 2001). More recently, two additional V's are added to the description of bigdata, namely, Veracity and Value. Hence big data is a data with "5V"s. Figure 1 shows 5vs of big data. Due to these characteristics of data, traditional processing system is unable to process it.

Figure 1. 5vs of big data (Pouyanfar, S., Yang, Y., Chen, S. C., Shyu, M. L., &Iyengar, S. S.,2018)



Big Data mining or analytics is the task of analyzing data of 5V's to extract the hidden interesting patterns, market trends, and unknown relations-associations, analyze customer behavior, their preferences and other useful business intelligence information. People share their views on various topics using social media platform. Earlier, before the invention of the Internet or Web, the companies used the techniques like surveys, polls to collect opinions of the people. Now, due to increase in the use of the Web, people openly discuss their ideas on social media and the companies can easily collect people's opinion using the Web. In today's world, there exists a huge competition among the organizations. Within a competitive market, sentiment analysis helps to understand the customer needs. Bigger organizations makes use of social media for promoting their business and marketing purpose. Mining the data generated by social media is called Social Big data mining. Due to the abundance of social media sentiments and emotions, analyzing these sentiments has become challenging. Sentiment analysis attempts to derive the sentiment expressed by an author against an entity. This chapter highlights various techniques used to mine these sentiments.