Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 112 (2017) 906–916

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France

Big data analysis to Features Opinions Extraction of customer

Ammar Mars[a*], Mohamed Salah Gouider[a,b]

[a] *Université de Tunis, ISG , SMART Lab , 2000, Cité Bouchoucha, Le Bardo, Tunisia*
[b] *Université de Tunis, ESSECT, 1089 , Montfleury ,Tunisia*

Abstract

Opinion mining refers to extract subjective information from text data using tools such as natural language processing (NLP), text analysis and computational linguistics. Micro-blogging and social network are the most popular Web 2.0 applications , like Twitter and Facebook which are developed for sharing opinions about different topics. This kind of application becomes a rich data source for opinion mining and sentiment analysis. This information is crucial for managers, who should improve the quality of a product based on customers' opinions. Concerning the characteristic of a product as mobile phone, it is particularly difficult to identify the features being commented on (e.g., camera quality, battery life, price, etc).
In our work, we present a new method that able to extract product features opinions of customer from social networks using text analysis techniques. This task identifies customers opinions regarding product features. We develop a system for retrieving tweets about a product from Twitter and detect product features opinions and their polarity. To validate the effectiveness of this approach , we used a dataset published by Bing Lius group in our approach experimentation. This dataset contains many notated customer reviews of five products such as Canon G3 and Nokia 6610. Next, we test this method with tweets retrieved from Twitter about Nokia,Samsung and Iphone features products.

Peer-review under responsibility of KES International

Keywords: Big data; Sentiment Analysis; Opinion mining ; Feature mining ;

1. Introduction

Big data mining is considered as the core component of the modern decision support systems related to social networks and other data sources. "What other people think" has been an important piece of information decision-making processes. People are accustomed to frequently give their opinions about a subject available in different social media like Facebook and Twitter. As well as other Web sources such as: product reviews, forums, discussion groups, and blogs.
It is important for the decision maker of a company to study the opinion of customers about a product. However, due to the large amount of information collected from these data sources, it is essentially impossible for a supplier or

* Corresponding author.
E-mail address: ammar.mars@gmail.com

1877-0509
Peer-review under responsibility of KES International
10.1016/j.procs.2017.08.114

decision maker to read all this data collected and make a decision. For this reason, the ”big data opinion mining” has become an important issue for research in Web information. It aims to extract user opinions from a huge volume of data texts, automatically or semi-automatically.
It is a challenge for us to provide effective information extraction systems able to process a large mass of data and extract customer opinions from data available on the Web. It is important to study users' opinions about a merchandise, a product or a brand. e.g. studying users' opinions about a Samsung brand.
It is necessary to develop a system that helps to exploit users opinions from different data sources like open data. In this work, we aim to propose a big data architecture for analyzing data and extracting customers opinions about product features using a lot of search domain such as Machine Learning(ML), Natural Processing Language(NLP) and big data. This architecture is used to make decision. We consider Twitter as a data source: It is a microblogging service addressed in our work, a communication means and a collaboration system that allows users to share short text messages, which do not exceed 140 characters with a defined group of users called followers.
In this work, we aim to design a new approach able to identify features costumers opinions and their polarity about a specific product from Twitter. We present the related work in section 1. In section 2, we discuss our proposed approach for features opinions extraction from social networks. Next, We present the found experimentation and results founded in section 3. Finally, we give some conclusions in section 4.

2. Related Works

Feature Extraction (FE) is a delicate task in sentiment analysis .In this work, we will propose big data framework able to extract features opinions of customer about a product from opened data.In this work we use three main concepts :”Opinion Mining” , ”Big Data” and ”Opinion feature mining”. Several researchers have worked and many work has been published about these concepts . We present in this section the related work.

2.1. Big Data

To face the explosion data volume , a new field of technology was born that called ”Big Data”[7]. It was iInvented by the Wweb giants like Google, Facebook and Twitter with solutions designed to provide a real-time access to large volumes of data., We can mention that Google processes hundreds of pPetabytes (PB), Facebook generates data of more than 10 petabytes per month, and Twitter generates 10 terabytes daily.
It is used to describe a large mass of data. The difference between traditional data and this enormous mass of data is that the latter contains structured and unstructured data. Big data give us an opportunity and a challenge ,at the same time, to process a very large mass of data which are unstructured. Also, these data could not be acquired, managed and treated by traditional information technology.
The concept of big data has been characterized by ”3V”: volume, velocity and variety[8] . The value chain of Big data can be generally divided into four phases: the generation, acquisition, storage and data analysis[9].Data analysis is the latest and most important phase in the chain of Big Data[8] . It is used to extract useful to provide fatherly suggestions or decisions. The challenge of this phase is to quickly extract information from massive data. There are several treatment methods, and the most used method is the ”Parallel Computing”.
We cite MapReduce[25] and Dryad[26] as parallel programming model.MapReduce[25] is a simple but powerful programming model for large-scale computing using a large number of clusters of commercial PCs to achieve automatic parallel processing and distribution. In MapReduce, the computational workload are caused by inputting key-value pair sets and generating key-value pair sets. The computing model only has two functions, i.e., Map and Reduce, both of which are programmed by users.

2.2. Opinion Mining

An important part of our information-gathering behavior is always to find out ”what other people think ?”. With the emergence of Web 2.0, a lot of opinion resources are available such as online review sites and personal blogs. They are a new opportunities and challenges to use information technologies to seek out and understand the opinions