Journal of Engineering Science and Technology
Vol. 12, No. 8 (2017) 2154 - 2174

A MULTI-AGENT BASED SOCIAL CRM FRAMEWORK FOR EXTRACTING AND ANALYSING OPINIONS

ABDELAZIZ EL FAZZIKI[1,*], FATIMA ZOHRA ENNAJI[1], ABDERRAHMANE SADIQ[1], DJAMAL BENSLIMANE[2], MOHAMED SADGAL[1]

[1]Computing Systems Engineering Laboratory (LISI),
Cadi Ayyad University, Marrakech, Morocco
[2] Informatics Laboratory in Image and Information Systems,
Claude Bernard University Lyon 1, France
*Corresponding Author: elfazziki@uca.ma

Abstract

Social media provide a wide space for people from around the world to communicate, share knowledge and personal experiences. They increasingly become an important data source for opinion mining and sentiment analysis, thanks to shared comments and reviews about products and services. And companies are showing a growing interest to harness their potential, in order to support setting up marketing strategies. Despite the importance of sentiment analysis in decision making, there is a lack of social intelligence integration at the level of customer relationship management systems. Thus, social customer relationship management (SCRM) systems have become an interesting research area. However, they need deep analytic techniques to transform the large amount of data "Big Data" into actionable insights. Such systems also require an advanced modelling and data processing methods, and must consider the emerging paradigm related to proactive systems. In this paper, we propose an agent based social framework that extracts and consolidates the reviews expressed via social media, in order to help enterprises know more about customers' opinions toward a particular product or service. To illustrate our approach, we present the case study of Twitter reviews that we use to extract opinions and sentiment about a set of products using SentiGem API. Data extraction, analysis and storage are performed using a framework based on Hadoop MapReduce and HBase.

Keywords: Big data, Social CRM, Hadoop, Sentiment analysis, Multi-agent system.

Nomenclatures

E_{aij}	Explicit link between two users
I_{aij}	Implicit link between two users
IS	Influence score
KS	Knowledge score
RS	Rate score
S	Number of posts
S_n	Negative score
S_p	Positive score
s_i	A post
u_i	A user

Abbreviations

AS	Authority Score
AML	Agent Modelling Language
API	Application Programming Interface
BDI	Belief-Desire-Intention
BI	Business Intelligence
CRM	Client Relation Management
DAA	Data Analysis Agent
DEA	Data Extraction Agent
DMA	Data Management Agent
JADEX	Java Agent Development Extension
MA	Manager Agent
MAS	Multi-Agent System
SCRM	Social Client Relation Management
SN	Social Network

1. Introduction

Modern organizations are increasingly using customer relationship management (CRM) in order to maintain long-term relationships with customers. This approach leads to increase firms' benefits in terms of customers' loyalty, profitability and increased sales [1].

For many years, data warehouse technology, data mining techniques, and online analytical processing tools were the key technologies used to build management information systems. However, these tools end up being inadequate for gathering and processing data from different and distributed open data sources, such as social media.

The use of social networks is continuously increasing, especially among young people. In terms of numbers: 40% of the population looks for information about a product or a service via social networks, 77% of consumers pay attention to the comments written about a product, and 75% of users trust the recommendation provided by social media rather than personal recommendations [2], as the reviews express the subjective attitudes, evaluations, and speculations of people in natural language. These facts encourage companies to use social networks to draw attention to their products, services and brands with the aim of building up customer relationship and increase demand.

Integrating data from social media sites such as Twitter, LinkedIn, YouTube and Facebook in the CRM system allows firms to collect and analyses costumer's opinions about a company, a brand, a product or a service.

The main idea of this work is to propose a multi-agent approach for modelling and developing CRM systems. In this approach, a number of autonomous agents are responsible of different tasks to perform. Each agent tries to achieve a particular goal and can make real-time decisions according to the system needs. In order to demonstrate the effectiveness of the system, we have conducted an experiment on data retrieved from Twitter using SentiGem API.

The remainder of this paper is organized as follows. The Section 2 describes related works and studies. Section 3 provides an overview of the system and section 4 presents the system process. The system development and the resulting multi-agent system (MAS) are described in Sections 5 and 6. Sections 7 and 8 are devoted to data modelling and analysis processes. In Section 9 we present some experimental results of the proposed system followed by a conclusion and perspectives in Section 10.

2. Related Works

Nowadays, the World Wide Web, has become a popular tool for the dissemination of information [3]. Its philosophy allows a rapid distribution of information in all over the world. The diversity and the quantity of information present in the Web generate several challenges, such as: Discovering new knowledge, finding pertinent information, and having a thought regarding customers, business partners and other entities (products, services, events, …) [4].

Social network extraction methods using search engines were the subject for several researches, and many applications and frameworks were introduced to resolve this issue: REFERRAL WEB [5], Flink [6], Polyphonet [7] and [8] which propose models for a Business Intelligence (BI) framework intended for heterogeneous data sources. [9], also proposed the use of co-occurrence to address the issues related to the evaluation of a relationship significance (is significant or not).

Ting et al. [10] addressed the issue of multi-source extraction, required in case of data extraction from different sources, such as: Social networks, Blogs, email or web sites. Also, Wang et al. [11] proposed an approach to extract data and build the social network graph from emails and instant messages.

The use of social media's content in decision making by individuals and organisations is increasing with the explosive growth of these computer-mediated tools. Over the last few years, social and political systems have been greatly impacted by opinionated reviews on social networks. These postings have helped influence public sentiments and reshape businesses. Thus, many research-oriented applications have been proposed. For example, Liu et al. [12] proposed a sentiment model to predict sales performance and McGlohon et al. [13] used reviews to rank products and merchants.

In addition, in [14, 15] Twitter sentiments are linked with public opinion polls, and applied to predict election results. Twitter data, movie reviews and blogs were used in [16-18] in order to predict box-office revenues. Miller et al. [19] discussed how far sentiments can be propagated in a large hyperlinked network (sentiment flow). Also, in [20], emotions in mail and books (novels and fairy tales) are tracked. Likewise, in order to predict the stock market, Bollen et al. [21] made use of Twitter moods.

Using Hadoop as a data hub to optimize data processing and decision making process is a new emerging strategy for information system enhancement and Big Data management. Many research works have proposed methods to exploit the Hadoop framework by effectively integrating it to the existing data warehouse: Das and Mohapatro [22] for example, proposed a study on big data integration with data warehouse built using relational technology mainly for operational sources. And Conejero et al.[23], Hadoop is used to extract sentiment data about businesses from social media conversations, and use it to make targeted, real-time decisions.

Sokolova et al. [24] proposed an agent-based decision support system development approach where the software agents use data mining methods for knowledge discovery, that will be used as a foundation for decision making and recommendation generation. Our MAS development approach is based on this work. This system provides all the necessary steps for a standard decision making procedure using intelligent agents [25].

3. The System Overview

In this system we propose to use a distributed processing approach for the social CRM framework development. This approach is based on the Hadoop MapReduce paradigm [26] in addition to HBase for data storage [27]. Data is extracted using the targeted social media extraction API (e.g., Twitter4J [28]), then used in the Social network construction which will be processed by using a MapReduce job in order to efficiently perform the opinion extraction. The system ensures several features such as:

- Data Extraction from a chosen social media website
- A social network construction based on the social media provided API
- Data Processing using Hadoop tools
- Sentiments Analysis
- OLAP Analysis
- A simple user front-end interface

Figure 1 illustrates an overview of the proposed system.

4. The Opinion Extraction Process

The social intelligence mechanism has been developed in order to easily find opinions expressed in social networks comments, which allows for example to know more about users' assessments of a product and compare it to competitor products. This can be achieved by collecting the strengths and weakening points of the current product or earlier products.

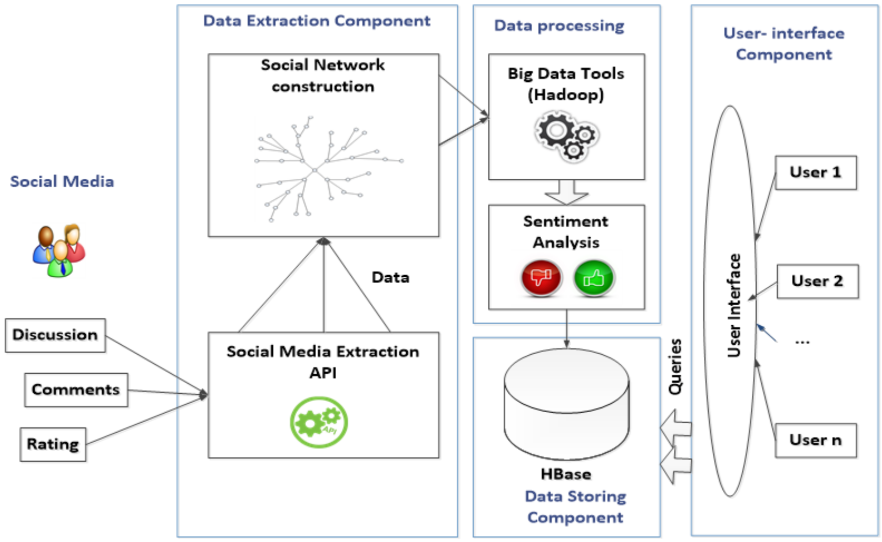

Fig. 1. The system overview.

The opinion extraction process starts by choosing the targeted product keyword (for example the brand name), then the system extracts the opinions related to the product category from social media. The collected online data are organized into a network of reviewers, and rating scores.

In this study, we choose to extract opinions from Twitter. The components included in the system are described as follows (Fig. 2). The data extraction steps are:

- Data collection
- Spam detection
- Knowledge analysis
- Authority analysis
- Sentiment analysis

4.1. Data collection and social network extraction

Social data collection refers to all the methods that have as an objective, collecting social interactions between entities. The social network is constructed by linking users with implicit and explicit bridge. The implicit link is built if a user comments on a post. The explicit link expresses the social link between two users.

4.2. Spam detection

During the spam detection stage we perform a first analysis process in order to eliminate the invaluable tweets based on the following elements:

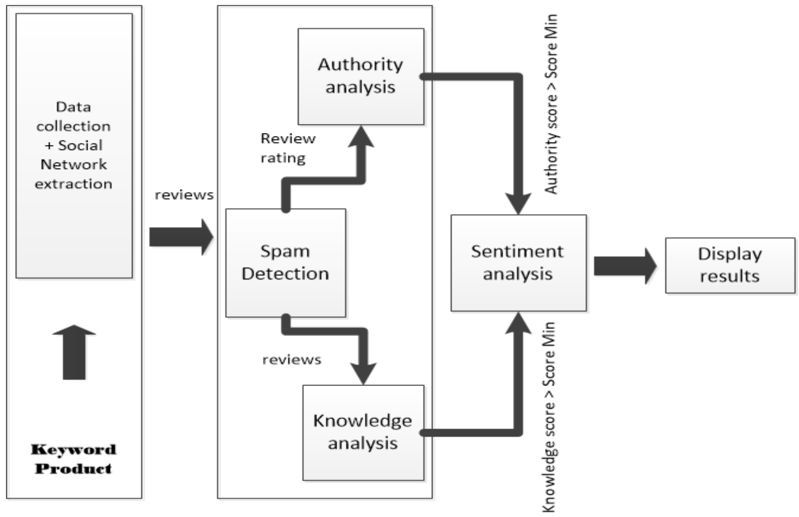

Fig. 2. The Opinion extraction process.

4.2.1. URLs

Since Twitter only allows a message with a maximum length of 140 characters, we can consider that the presence of an URL in a tweet means that it's a spam message. URLs are detected using regular expression functions (Java Regex API).

4.2.2. Replies/Mentions

Spammers often abuse this feature by including @usernames in their twitter's replies or mentions. These posts are considered as unsolicited if a user includes too many mentions and can then be modelled as spam.

4.3. Knowledge analysis

This module includes the quality and the significance assessment according to the author's expertise level. This score means how many times a user has published a post about this topic and its value. In addition to the number of reviews a user has written, this system evaluates the quality of each post by calculating its rate score. The knowledge score (KS) of a user u is:

$$KS(u) = \frac{\sum_{s_i \in S} s_i \times pertinence(s_i, u)}{|S|} \tag{1}$$

where |S| is the number of posts, s_ia post and $pertinence(s_i, u)$ is a score that define the quality of the post and how much this review is interesting. This score is equal to the number of times that the keyword and are repeated.

4.4. Authority analysis

A hundred of users publish and give their opinion about a subject, a product or a service. But the question is: Does that article influence the other user's opinion?

The authority analysis process is based on the influence score and the rate review.

4.4.1. Influence scores

The influence score evaluates whether the person has a wide network (friends, followers, comments …). This network is constructed based on different types of links (implicit and explicit) depending on the relationships nature. If a user comments or evaluates an article, an implicit link is created. The explicit link means several relations (followers, subscriber…).

$$IS(u) = \frac{\sum_{u \epsilon SN}(\mathrm{E_{aij}} + \mathrm{I_{aij}})}{|SN - \{u\}|} \tag{2}$$

where, Eaij (Iaij) stands for the explicit (implicit) link between two users. Eaij=1 (Iaij=1) if the link exists; Eaij=0 (Iaij=0), otherwise. |SN − {u}| represents the total number of users on the social network (SN) except user u. The influence score is calculated based on the algorithm shown below.

```
double InfluenceScore ( userNetwork , user ) {
        double influence = 0;
        int i;
        for ( i = 0; i < userNetwork.Length; i++ ) {
             if ( relationShip ( user, userNetwork[i] ) ) {
                         influence++;
}
        }
        return influence;
}
```

4.4.2. Rate review

Most of social networks nowadays, offer rating tools on publications which are used in order to generate a Rate Score (RS). This information is useful in our case and it can be calculated as follows:

$$RS(u_i) = \frac{\sum_{u_k \neq u_i \in SN, r_j \in \varphi_i} rate(r_j, u_k) \times KS(u_k)}{|SN - \{u_i\}| \times |\varphi_i|} \tag{3}$$

This module was a subject of several studies. The aim of sentiment analysis is to collect users' opinions about a product or a service (detection of the authors' feelings). Two expressions are highlighted: Sentiment Analysis and Opinion Mining. These two terms are interchangeable and they express a mutual meaning, although there is a slight difference. Opinion Mining extracts and analyses people's opinion about an entity while Sentiment Analysis identifies the sentiment expressed in a text then analyses it. To do so, we classify collected sentences by using SentiGem [29]. This tool classifies sentences into three groups: negative, positive and neutral opinion. As alternatives of SentiGem, there are: SentiStrenght [30] and SentiWordNet [31].

5. The System Development

Multi-agent systems actually constitute a choice technology for the design and implementation of distributed and cooperative applications. The duality between agent as a standalone and adaptive entity and multi-agent system as a cooperative decentralized organization was privileged to address aspects increasingly dynamic and distributed in applications such as a social CRM framework. The Agent technology provides an adequate solution in terms of adaptability, flexibility and reactivity. The system is composed of a set of autonomous agents which collaborate with each other via an interaction protocol in order to achieve the main objective of the system.

5.1. The development process

For the agent modelling, we first build a use case model and a goal hierarchy diagram in order to define the requirements and capture goals [32]. Then, we use the AML language (Agent Modelling Language) [33] which is a semi-formal visual modelling language widely used for modelling systems that incorporate concepts related to MAS theory [33]. The AML based modelling process is divided into five main phases (as shown in Fig. 3): Modelling MAS Entities, Modelling Social Aspects, Modelling Interactions, Modelling Capabilities and Behaviour and Modelling MAS Deployment and Mobility.

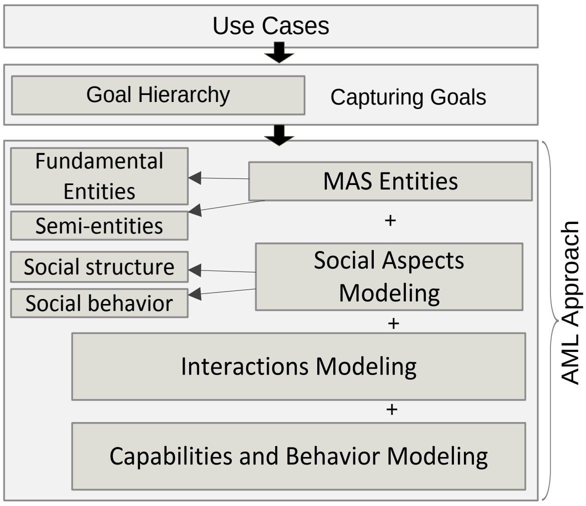

Fig. 3. The development approach.

In the following section we present the modelling stages and several resulting diagrams obtained based on the AML development approach.

5.2. Agent modelling

5.2.1. The domain analysis

The first modelling stage is based on the global domain analysis and the identification of the main elements that have an impact on the quality of the system and should be taken into account in this stage, such as use cases and actors. These elements are used in the goals (external and internal goals) capturing stage. Figures 4 and 5 show the resulting goal hierarchy diagram and the UML use case diagram.

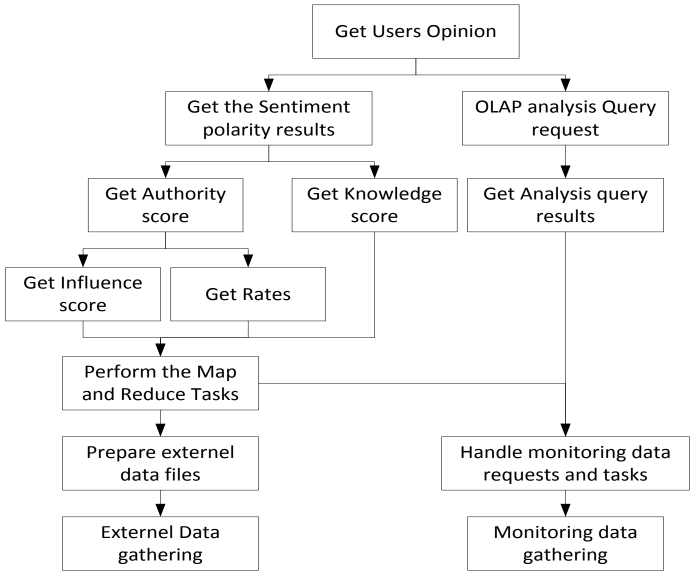

Fig. 4. Goal hierarchy diagram of the MAS.

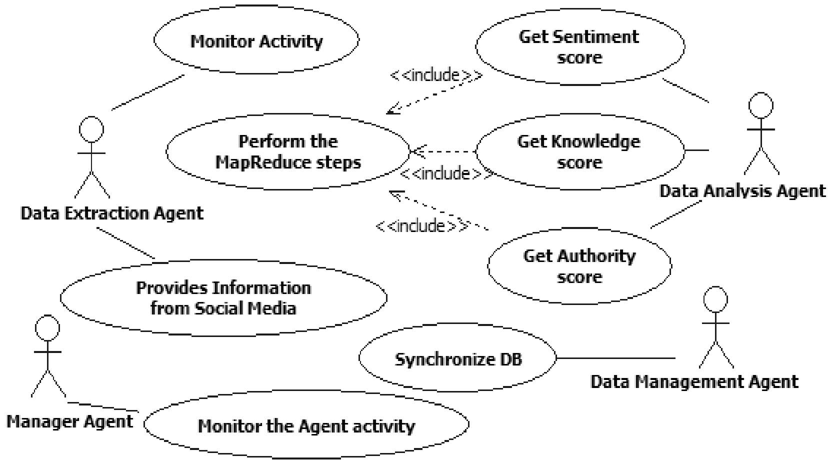

Fig. 5. The system use case diagram.

5.2.2. MAS entities modelling

A specialized UML classes are used for representing the main entities that compose MASs and specify a coherent set of features, logically grouped according to the aspects of the system. The resulting diagram is completed by the other details related to the social aspects.

5.2.3. Social aspects modelling

The objective of the social aspect modelling is to represent the structural and behavioural aspects of the MAS by using numerous modelling elements. These elements are designed to explicitly represent diverse abstractions of society within the system. Figure 6 shows a part of the resulting social model.

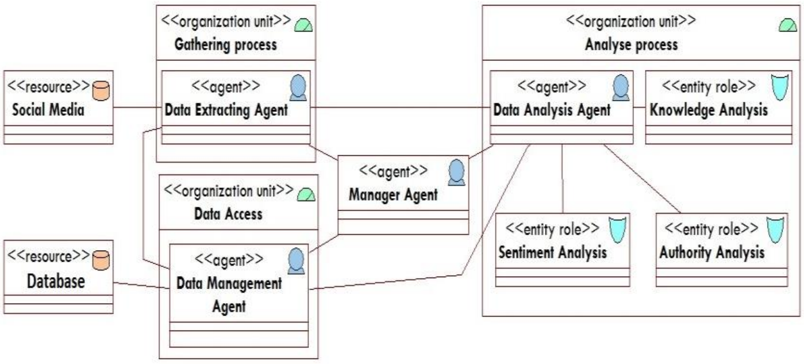

Fig. 6. Part of the system social model.

5.2.4. Interactions modelling

The aim of this stage is to represent interactions within the MAS. AML provides a number of UML extensions, such as a generic extension to UML interactions. In this stage, we model the interactions between entities, by representing interactions, changes, messages, etc. The interactive behaviour can be represented by several diagrams such as sequence diagrams. Figure 7 illustrates an overview of the proposed system sequence diagram.

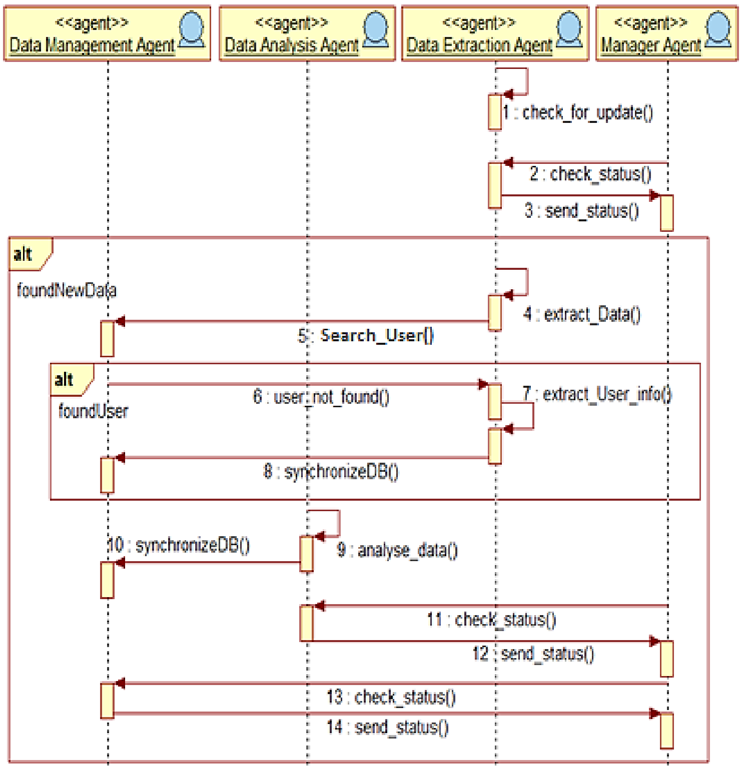

Fig. 7. Sequence diagram.

5.2.5. Capabilities and behaviour modelling

In addition to the capacities that UML offers, AML has also the ability to abstract and decompose behaviour by two additional modelling elements: capability which is an abstract specification of a behaviour which allows reasoning with and operations on that specification. The second element is a specialized behavioural type called behaviour fragment. It is used to model a coherent re-usable fragment of behaviour and related structural and behavioural features.

5.2.6. Modelling MAS deployment and mobility

The aim of this stage is modelling the multi-agent system deployment and agent mobility by representing the physical infrastructure onto which the resulting entities are deployed, the entities hosting property as well as the relationship between deployed entities and the deployment nodes.

5.3. Agent implementing

The system implementation is made by using the JADEX (Java Agent Development framework Extension) platform [34], which is an hybrid agent architecture for representing mental states in JADE agents in accordance with the BDI (Believe, Desire, Intention) model. The Jadex agents follow the concepts of beliefs, goals and plans which are Java objects that can be created and managed within the agent at execution time. The Jadex kernel is a BDI reasoning engine for intelligent agents. It can be used together with different agent which provide basic capabilities such as a communication infrastructure and management facilities [35]. Thus, the use of Jadex for implementing the system is suitable in our case.

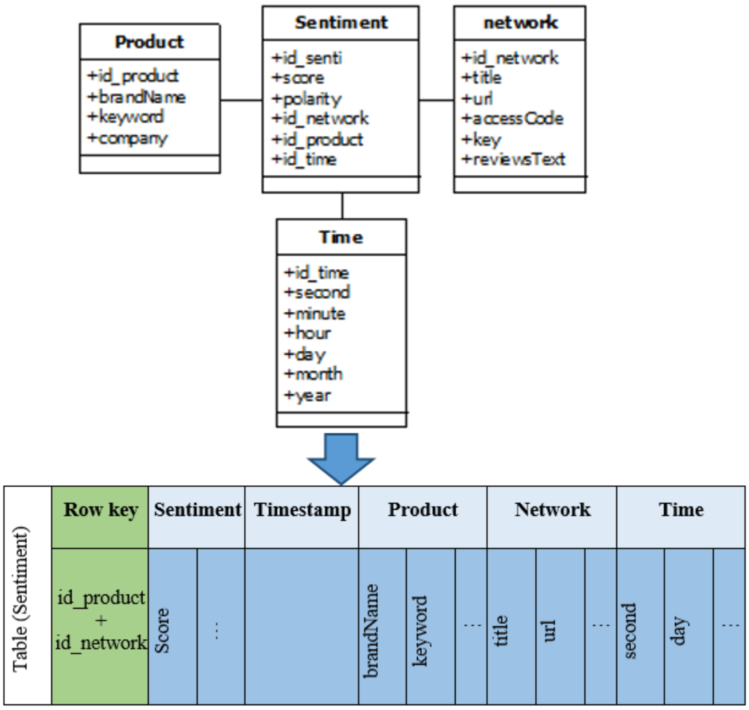

Fig. 11. Social Media Database schema.

6. The Resulting Agent System for Social CRM

According to the modelling process we can assign each agent to the suitable component. Table 1, shows the different agents and their corresponding components and Figure 8 illustrates an overview of the different agents and their interactions.

Table 1. The system agents.

Components	**Agents**
Data integration component	Data Extraction Agent
Analysis Component	Data Analysis Agent
Data Management Component	Data Management Agent
Interacts with all components	Manager Agent

The multi-agent social CRM system was developed in order to demonstrate the use of a multi-agent approach for designing and developing a social CRM.

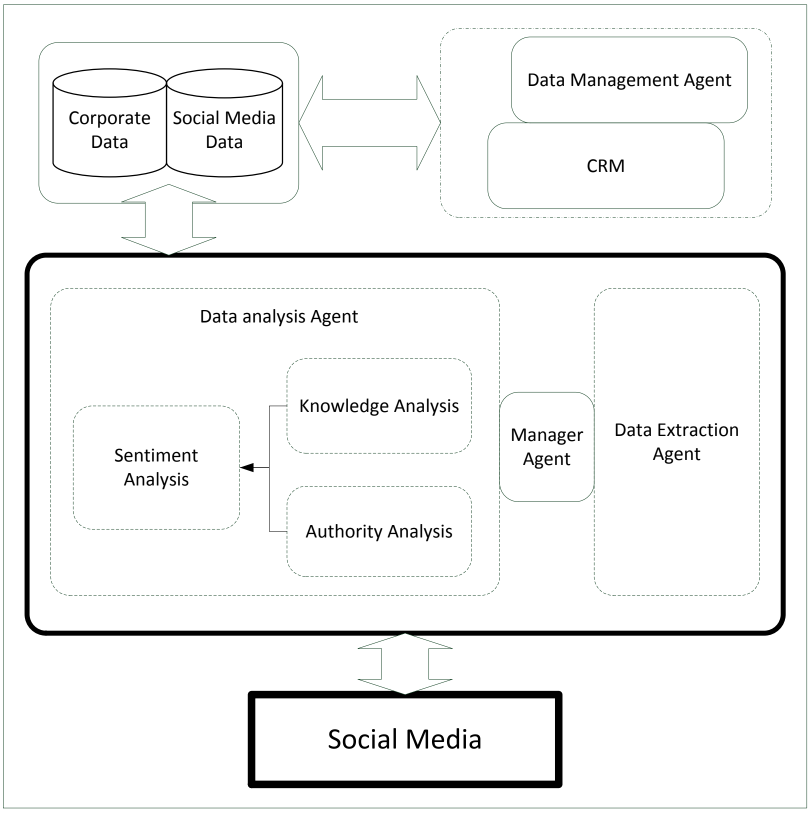

Fig. 8. The multi-agent social CRM system architecture.

6.1. Data extraction agent

The Data Extraction Agent (DEA) handles data gathering from social media. As a result, it has two main capacities: i) Data extraction from social media, and ii) Real time detection of users' activities that contain the chosen key words (see Fig. 9).

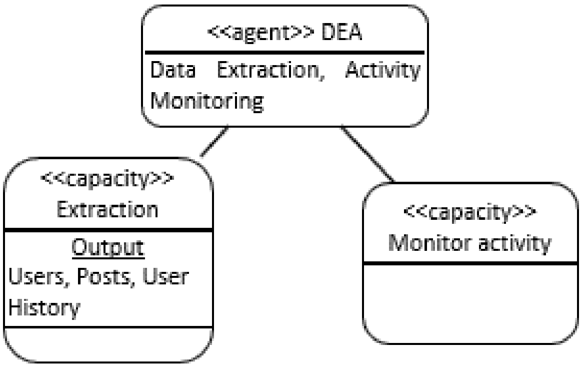

Fig. 9. Representation of the DEA capacities.

6.2. Data analysis agent

The Data Analysis Agent (DAA) analyses extracted data and updates stored results. This agent evaluates the knowledge and Authority scores (AS) of users, if these scores filled out conditions (the score exceeds the threshold), then the agent can perform the sentiment analysis stage. Data analysis is based on a MapReduce model for distributed computations which is implemented using the Hadoop framework. Thus, this agent uses the Hadoop MapReduce framework provided resources [36] in order to perform the needed analysis. Figure 10 illustrates this agent functioning.

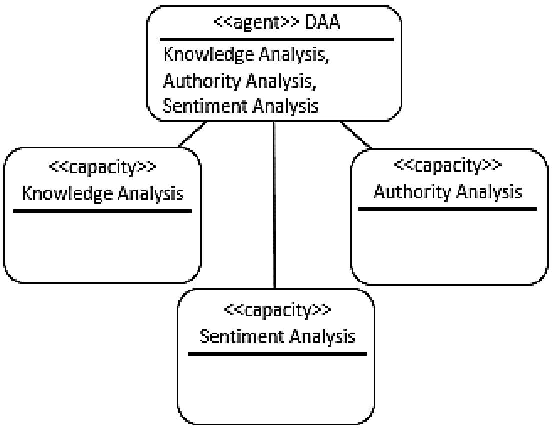

Fig. 10. Representation of the DAA capacities

6.3. Data management agent

The Data management Agent (DMA) is responsible of storage synchronization. Every time that an entry is evaluated, DMA intervenes to refresh the results stored in the database. This agent must ensure enough performance to minimize the offline time during updates operations.

6.4. Manager agent

The Manager Agent (MA) is responsible for the reliability of the whole system and manages the operation of the individual agents. It sends a request for the data extraction agent which search and send the requested information to the data analysis agent. The Manager agent prepares the analysis results for the data management agent which loads them into the appropriate target.

7. Data Modelling

In this work, all gathered data are stored into a data warehouse using Apache Hadoop HBase [27] which is a distributed non-relational database with column-oriented storage. Its goal is hosting very large tables with billions of rows. An HBase table is organized as key-value and each table contains a series of records. Figure 11 represents the data warehouse conceptual schema and the corresponding HBase table structuring. In our case, the fact is the “Sentiment” entity which has three dimensions: “Product”, “Network” and “Time”. Each dimension is mapped into a column family in the HBase table. The fact measures are also grouped into a column family (with the same name as the Fact entity) in the target table.

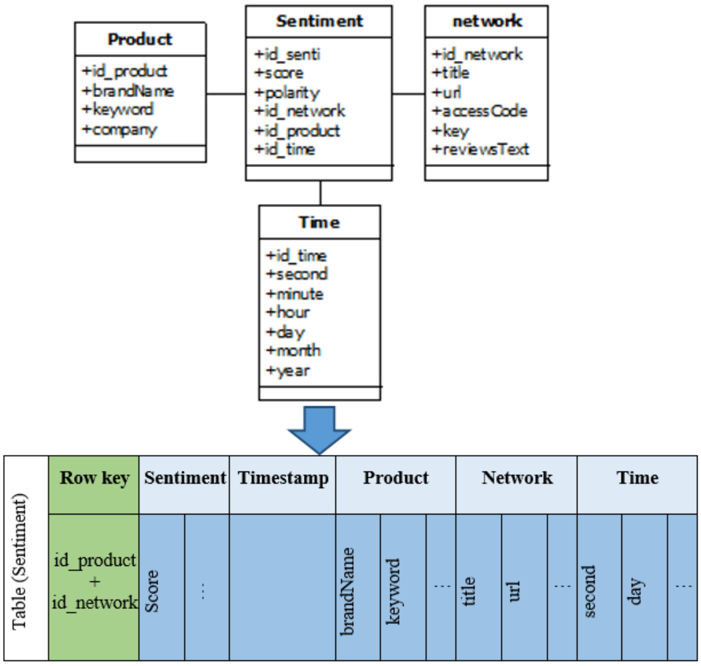

Fig. 11. Social Media Database schema.

8. Data Analysis

MapReduce is a programming model used to handle vast amounts of data and easily write application which process huge data sets in parallel [26]. It is particularly adapted to distribute computing on a set of clusters. In our case; this model will be used to process the social networks’ data. This data is divided into a set of independent fragments which are processed by the map tasks in parallel.

The outputs are then sorted and passed to the reduce job. The input and output data are stored in a Hadoop HBase during the map and reduce jobs (Fig. 12).

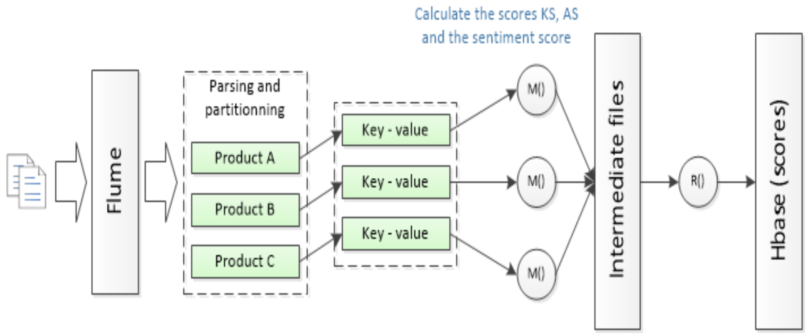

Fig. 12. The MapReduce steps data processing.

At the Map stage, each Tweet message is evaluated by the previous scores (AS and KS). These two scores are used to remove the invaluable tweets (opinions). If they satisfied the conditions, they are processed by SentiGem in order to get the sentiment scores (negative, positive or neutral). Each result is generated independently, comprising the Tweet identifier and its associated positive, negative and neutral sentiment scores. With this architecture, the Map algorithm can be easily adapted to perform different analysis on individual Tweet messages by replacing SentiGem with another analysis package. The Reduce stage outputs the results obtained by the Mapper. The map and reduce functions are shown below.

```
Map (String file_name, String tweet, User user) {
        String ProductName = getProduct (tweet);
        int as = authorityAnalysis (user);
        intks = knowledgeAnalysis(user);
        If (as >minAs&&ks>minks) {
                String score = Sentigem (tweet);
                file_name = Emit ((String) ProductName, (String) score);
    }
}
Reduce (String ProductName, String score) {
        Emit ((String) ProductName, (String) score);
}
```

9. Case Study

9.1. Description

In this study, we choose to extract opinions from Twitter, one of the most popular social networking site. To verify the effectiveness of the proposed framework, we compare the extracted results of three different car brands: Product A, Product B and Product C. The system collects the data from twitter using the Twitter4J API [28] and removes the spam tweets.

The experimental setup comprises a multi-node virtual Hadoop cluster of six nodes. The infrastructure is built on 12 Intel Xeon 2.83 GHz cores and 32 GB memory. The used operation system is Ubuntu 12.04.

9.2. Application

Tweets can be transformed to spam tweets for several reasons, for example, if the tweet contains a long link or a lot of hash tags and usernames. But after collecting the data, other spam types can appear (e.g. sales proposals). Table 2 exposes some examples of spam tweets.

Table 2. Spam types on Twitter.

Type of spam	Example
Long URL	http://car-us.com/car/volkswagen-touran/39883 …#volkswagen 2010 VW Touran 2.0 TDi SE
Hashtagsand username	#nouveau#Mini#Clubman long commeune#Volkswagen#Golf@MINI_FR@MyMiniParis@lookatmyminihttp://urlc.fr/AdTaEM
Sales proposals	volkswagen.cars-stock.com/volkswagen-passatstdi/28694 …#volkswagen 2003 VOLKSWAGEN PASSAT S TDI SILVER IN VERY GOOD CONDITION

When the network is built, a knowledge and authority analysis is done to assign scores to users and their tweets. Finally the sentiment analysis detects negative, positive and neutral tweets after cleaning messages (remove punctuations, hashtags, ..). Figures 13 to 15 illustrate the experimental results for the three products (A, B and C).

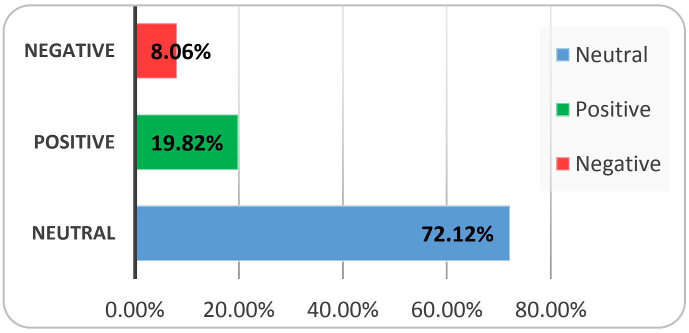

Fig. 13. Opinion dispersion for “Product A”.

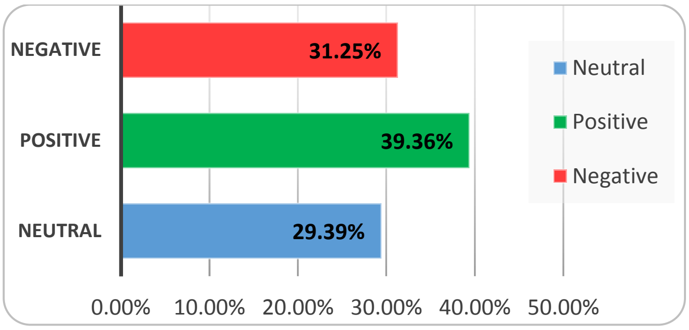

Fig. 14. Opinion dispersion for “Product B”.

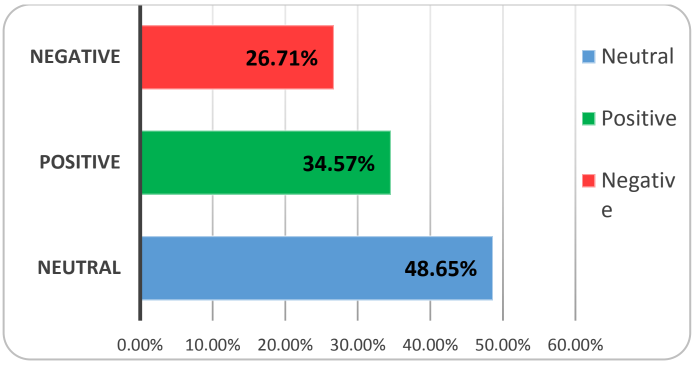

Fig. 15. Opinion dispersion for “Product C”.

We note that the most important area is always occupied by the neutral reviews. It is due to the large number of tweets that contains only the sales proposals which include the characteristics of the product to sell, in addition of the real neutral opinions.

In order to overcome this issue, we choose to neglect the neutral opinion of each product, and then recalculate the positive and negative opinion. The results are illustrated in Fig. 16.

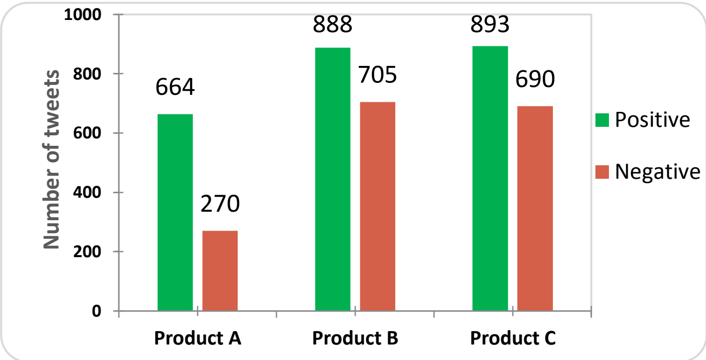

Fig. 16. Positive and negative opinion for each product.

As a result, each product has a positive and a negative score, noted respectively as $S_P(Product)$ and $S_N(Product)$. The following formula calculates the final score:

$$S(Product) = S_P(Product) - S_N(Product) \tag{4}$$

As we can notice in Fig. 17, the product A comes first, followed by product C, then B with a small difference.

To compare this result with real statistics, Fig. 18 illustrates the number of selling units in January 2014.

As we can notice the product (A) comes first in both, obtained and real world statistics, but for the second and the third product they were switched. Thus, “Product B” comes in the second place instead of “Product C”. This goes back to the small difference between these two products.

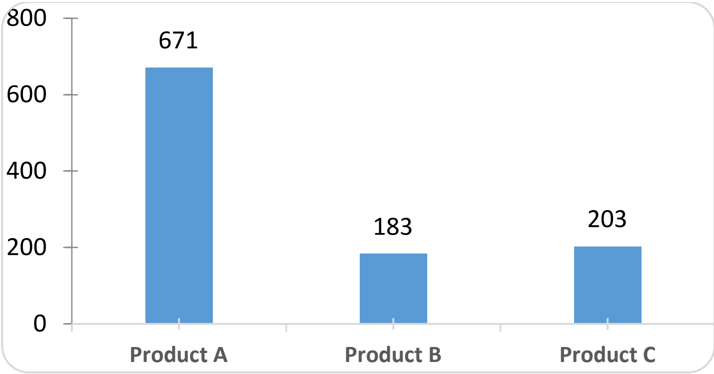

Fig. 17. Final scores for each product.

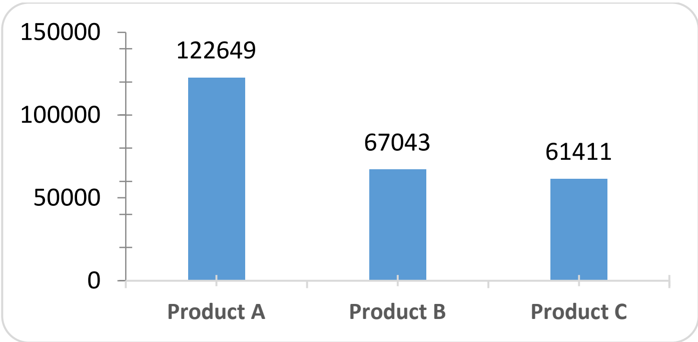

Fig. 18. The number of selling units in January 2014.

9.3. Discussion

In the proposed multi-agent based social CRM framework, social networks are used as input data sets. Collecting them makes this application suitable for big data analytics. The processing of such increasingly larger datasets must take into account the very strict constraints related to storage and time. To overcome this issue, we proposed the use of Hadoop and MapReduce, which might be the best solution due to its ability to process large volumes of structured and unstructured data. Also, it can easily be integrated to the existing information system, due to its flexibility and ability to adapt to other technologies.

Furthermore, the system is dedicated to be an analytical tool to help firms extracting and consolidating public opinion expressed via social media. Twitter, one of the most famous social media; has been adopted in the presented case study to evaluate the system's effectiveness and the results relevance. In order to readapt this system to be used for another social media, we have just to change the Twitter API by another API (e.g. Graph API for Facebook). And to do so, we only should intervene in the first part of the process (data extraction) the rest of the process is not related to a particular API.

The Multi-agent technology has been used in the system development process. It allows increasing the level of agility, interoperability and flexibility of distributed and dynamic applications. The resulting Multi-Agent System is built utilizing the AML language.

10. Conclusion

The fact that the majority of customers are nowadays using social media to express their feelings and opinions with freedom and ease, make these websites an open source of valuable information. Integrating customer's opinions into the business as drivers for decision making is gradually becoming an important practice. Using collective reviews from social media in order to provide insights for new products development will greatly enhance companies' competitiveness. By adopting social intelligence technologies; enterprises can have better planning, steering and analysis based on online social networks.

This study demonstrates that the use of multi-agent approach is an interesting and useful method for SCRM designing, since it allows creating an autonomous application that contains a group of intelligent agents which collaborate and provide needed features for the proposed system.

The extracted data are categorized as "Big Data"; which obviously leads to use Hadoop to ensure a fast data loading, fast query processing and an efficient storage. The Hadoop highly efficient fault tolerant nature, flexibility, extensibility, efficient load balancing and the platform-independent are also useful features for the development of the system. Our future work consists essentially of the improvement of the data warehouse model, the integration of an Extract-Transforming-Loading (ETL) process and OLAP tools, to improve the data integration and analysis and allow the use of social OLAP.

References

1. Payne, A.; and Frow, P. (2013). A Strategic Framework for Customer Relationship Management. *Journal of Marketing*, 69(4), 167-176.
2. Ku, Y.C.; Wei, C.P.; and Hsiao, H.W. (2012). To whom should I listen? Finding reputable reviewers in opinion-sharing communities. *Decision Support Systems*, 53(3), 534-542.
3. Arif, T.; and Ali, R. (2014). Social Network Extraction : A Review of Automatic Techniques. *International Journal of Computer Applications*, 95(1), 16-23.
4. Singh, B.; and Singh, H.K. (2010). Web Data Mining Research : a Survey. *Proceedings of IEEE International Conference on Computational Intelligence and Computing Research*. Tamilnadu College of Engineering, Coimbatore, India, 1-10.
5. Kautz, H.; Selman, B.; and Shah, M. (1997). The hidden web. *AI Magazine*, 18(2), 27-36.
6. Mika, P. (2005). Flink: Semantic Web technology for the extraction and analysis of social networks. *Journal of Web Semantics*, 3(2-3), 211-223.

7. Matsuo, Y.; Mori, J.; Hamasaki, M.; Nishimura, T.; Takeda, H.; Hasida, K.; and Ishizuka, M. (2007). POLYPHONET: An advanced social network extraction system from the Web. *Journal of Web Semantics*, 5(4), 262-278.
8. Troyano, D.L.R.; López, F.G.; Teresa, M.G.; and Martínez, R. (2010). Competitive Intelligence Based on Social Networks for Decision Making. *International Journal of Software Engineering and Its Applications*, 4(4), 93-104.
9. Jin, Y.; Matsuo, Y.; and Ishizuka, M. (2007). Extracting Social Networks Among Various Entities on the Web. *The Semantic Web: Research and Applications, Lecture Notes in Computer Science*, 4519, 251-266.
10. Ting, I.H.; Wu, H.J.; and Chang, P.S. (2009). Analyzing multi-source social data for extracting and mining social networks. *Proceedings of the 12th IEEE International Conference on Computational Science and Engineering*. Miami, Florida, USA, 815-820.
11. Wang, K.Y.; Ting, I.H.; Wu, H.J.; and Chang, P. S. (2010). A dynamic and task-oriented social network extraction system based on analyzing personal social data. *Proceedings of the International Conference on Advances in Social Network Analysis and Mining*. Odense, Denmark. 464-469.
12. Liu, Y.; Huang X.; An, A.; Yu X.; and Huang, J. (2007). ARSA : A Sentiment-Aware Model for Predicting Sales Performance Using Blogs. *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. Amsterdam, Netherlands, 607-614.
13. McGlohon, M.; Glance, N.; and Reiter, Z. (2010). Star Quality : Aggregating Reviews to Rank Products and Merchants. *Proceedings of Fourth International AAAI Conference on Weblogs and Social Media (ICWSM)*. Washington, DC, USA, 114-121.
14. O'Connor, B.; Balasubramanyan, R.; Routledge, B.R.; and Smith N.A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. Washington, DC, USA, 122-129.
15. Tumasjan, A.; Sprenger, T.O.; Sandner, P.G.; and Welpe, I.M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *Proceedings of Fourth International AAAI Conference on Weblogs and Social Media*. Washington, DC, USA,178-185.
16. Asur, S.; and Huberman, B. (2010). Predicting the future with social media. *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.* Toronto, Canada, 492-499.
17. Joshi, M.; Das, D.; Gimpel, K.; and Smith, A.N. (2010). Movie Reviews and Revenues: An Experiment in Text Regression. *Proceedings of HLT '10, Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Los Angeles, CA, 293-296.
18. Sadikov, E.; Parameswaran, A.; and Venetis, P. (2009). Blogs as predictors of movie success. *Proceedings of the Third International AAAI Conference on Weblogs and Social Media*. San Jose, California, USA, 304-307.
19. Miller, M.; Sathi, C.; Wiesenthal, D.; Leskovec, J.; and Potts, C. (2011). Sentiment Flow Through Hyperlink Networks. *Proceedings of the Fifth*

International AAAI Conference on Weblogs and Social Media. Barcelona, Spain, 550-553.

20. Mohammad, S.M. (2012). From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems*, 53(4), 730-741.
21. Bollen, J.; Mao, H.; and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8.
22. Das, A.; and Mohapatro, K. (2014). A study on big data integration with data warehouse. *Journal of Computer Trends and Technology*, 9(4), 188-192.
23. Conejero, J.; Burnap, P.; Rana, O.; and Morgan, J. (2013). Scaling archived social media data analysis using a hadoop cloud. *Proceedings of IEEE International Conference on Cloud Computing, CLOUD*. Santa Clara, CA, USA, 685-692.
24. Sokolova, M.V.; and Fernández, C.A. (2009). Modeling and implementing an agent-based environmental health impact decision support system. *Expert Systems with Applications*, 36(2), 2603-2614.
25. Gloria, L.P.; and Jain, W. (2007). Recent Advances in Intelligent Decision Technologies. *Knowledge-Based Intelligent Information and Engineering Systems, Lecture Notes in Computer Science*, 4692, 567-571.
26. Dean, J.; and Ghemawat, S. (2008). MapReduce: Simplified Data Processing on Large Clusters. *Communications of the ACM*, 51(1), 107-113.
27. Apache. (2015). Apache HBase. Retrieved September 1, 2015, from http://hbase.apache.org/.
28. Twitter4J. (2007). Twitter4J. Retrieved July 8, 2015, from http://twitter4j.org/.
29. Sentigem. (2013). Sentigem BETA sentiment analysis API. Retrieved July 24, 2015 from http://sentigem.com/.
30. Thelwall, M.; Buckley, K.; and Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1), 163-173.
31. Baccianella, S.; Esuli, A.; and Sebastiani, F. (2010). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Malta, 2200-2204.
32. ElFazziki A.; Sadiq, A.; Ouarzazi, J.; and Sadgal, M. (2015). A Multi-Agent Framework for a Hadoop Based Air Quality Decision Support System. *Proceedings of the 27th conference on Advanced Information Systems Engineering*. Stockholm, Sweeden, 45-59.
33. Trencansky, I.; and Cervenka, R. (2005). Agent Modeling Language (AML): A Comprehensive Approach to Modeling MAS. *Informatica*, 29, 391-400.
34. Activecomponents. (2015). Jadex Documentation. Retrieved July 5, 2015, from https://download.actoron.com/docs/releases/latest/jadex-mkdocs/.
35. Pokahr, A.; Braubach, L.; and Jander, K. (2013). The Jadex Project: Programming Model. *Multiagent Systems and Applications*, 45, 21-53.
36. White, T. (2012). *Hadoop: The definitive guide*. O'Reilly Media, Inc.