A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives

Article · January 2017

4 authors, including:

Said Salloum
British University in Dubai
3 PUBLICATIONS **2** CITATIONS
SEE PROFILE

Mostafa Al-Emran
Al Buraimi University College
17 PUBLICATIONS **36** CITATIONS
SEE PROFILE

Khaled Shaalan
British University in Dubai
151 PUBLICATIONS **1,561** CITATIONS
SEE PROFILE

ASTES Advances in Science, Technology and Engineering Systems Journal Vol. 2, No. 1, 127-133 (2017)

www.astesj.com

Special Issue on Computer Systems, Information Technology, Electrical and Electronics Engineering

ASTESJ

ISSN: 2415-6698

A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives

Said A. Salloum[1, 2 *], Mostafa Al-Emran[3], Azza Abdel Monem[4], Khaled Shaalan[1]

[1]*Faculty of Engineering & IT, The British University in Dubai, UAE.*

[2]*University of Fujairah, UAE.*

[3]*Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Malaysia.*

[4]*Faculty of Computer and Information Sciences, Ain Shams University, Egypt.*

ARTICLE INFO

Article history:
Received: 12 December, 2016
Accepted: 06 January, 2017
Online: 28 January, 2017

Keywords :
Text Mining
Social Media
Facebook
Twitter

ABSTRACT

Text mining has become one of the trendy fields that has been incorporated in several research fields such as computational linguistics, Information Retrieval (IR) and data mining. Natural Language Processing (NLP) techniques were used to extract knowledge from the textual text that is written by human beings. Text mining reads an unstructured form of data to provide meaningful information patterns in a shortest time period. Social networking sites are a great source of communication as most of the people in today's world use these sites in their daily lives to keep connected to each other. It becomes a common practice to not write a sentence with correct grammar and spelling. This practice may lead to different kinds of ambiguities like lexical, syntactic, and semantic and due to this type of unclear data, it is hard to find out the actual data order. Accordingly, we are conducting an investigation with the aim of looking for different text mining methods to get various textual orders on social media websites. This survey aims to describe how studies in social media have used text analytics and text mining techniques for the purpose of identifying the key themes in the data. This survey focused on analyzing the text mining studies related to Facebook and Twitter; the two dominant social media in the world. Results of this survey can serve as the baselines for future text mining research.

1. Introduction

As we know that there are various social networking sites available, Facebook and Twitter are considered as the most crowded ones [1], [2]. These networking sites have made it easy to communicate with friends and family members without making any much effort [3], [4]. People related to different values come closer to each other by sharing their ideas, interests, and knowledge [5]. These days, it becomes very easy for anyone to meet the people of their interests for learning and sharing precious information [6], [7].

The advancement in technology has shrunk the world. The distances look closer and sharing information looks easier [8]. Through these social networks, people can easily and confidently share their point of views [9], [10] regarding various global issues by uploading their posts, text comments and blogs [11]. A study by [12] claimed that social media including Google Apps facilitate the way people learn, collaborate, and share ideas with each other. Moreover, social media has been incorporated by several learning forms such as e-learning and m-learning [13], [14]. Whatever the scenario is, people don't like to use structured sentences, correct grammar and spellings [6]. Not matter, whether they are searching something on the site, posting any comment or connecting people through various discussion forums. People use irregular data patterns to convey their messages. It seems like they have a shortage of time but due to the use of this unstructured language, it is not an easy task to bring out the correct and regular data patterns. On different social networking sites, the most common method of interaction with each other is through text. People share their knowledge and information through blogs, posts, and chats by writing in their own languages. The basic use of the text mining

*Corresponding Author: Said A. Salloum, University of Fujairah, UAE.
Tel: +971507679647 Email: Salloum78@live.com

methods is to make the text clear to make it easy for anyone to write or search in the most appropriate manner [15].

As people write words or sentences with errors, so in order to let them write or search with proper grammar and structured sentences, text mining approach [16] is used. Text mining means the extraction of the data which is not familiar to anyone. If we compare web searching with text mining then both the terms are vastly different from each other. If we talk about web searching, then you are fully aware of what you are going to search. But in the case of text mining, the main focus is to bring out the most appropriate data in accordance with the written text, no matter whether it is structured or not. This technique only requires a particular alphabet in order to dig out the data which is then further transformed into different suggestions and expectations. Text mining seems to grasp the entire automatic natural language processing. For instance, exploration of linkage structures, references in academic writing and hyperlinks in the Web writing are important sources of data that lie outside the conventional area of NLP. NLP is one of the hot topics that concerns about the interrelation among the huge amount of unstructured text on social media [17], besides the analysis and interpretation of human-being languages [18], [19].

Several research articles were collected from various databases in order to be analyzed and used in this survey. The search terms include "Text mining with social media", "Text mining with Facebook", and "Text mining with Twitter". This survey is categorized as follows: section 2 provides a complete background about the text mining field. Other related studies are addressed by section 3. Conclusion and future perspectives are presented in section 4.

2. Background

Businesses have identified data-driven approaches as the ideal blueprint for their growth. It is easier to understand this theory. After all, wouldn't it benefit a company to get an idea about the perception of its products in the market without having to consult individual reviews from everyone? Wouldn't it be better if they could gauge which political candidate is ideal for their public image without having to analyze them all individually? This is why market study and research are some of the most highly invested fields in the world right now. Social networking sites like Twitter and Facebook are ideal for this purpose. Posts or messages shared by people on these platforms with their friends remain freely accessible or are kept confidential. They give businesses the chance to scoop up public sentiments [20], [21] about topics that they are interested to share by a large group of people.

The processing of surveys and public impressions using specially designed computational systems is a shared objective of inter-connected fields like subjectivity analysis, opinion mining [22], and sentiment analysis. Creating problem-solving techniques or methods to define the structure and precedence or for summarizing opinionated messages for particular topics [23], occasions or products is another target of the survey. For example, these methods could be used for gauging support for particular occasions or items, or determining thumbs down or thumbs up votes for specific movies based on their reviews.

2.1. Text Mining

Text mining makes it easy to obtain a meaningful and structured data from the irregular data patterns [24], [25], and [26]. It is really not an easy task for the computers to understand the unstructured data [27], [28] and make it structured. Human beings can perform this task without any further efforts due to the availability of different linguistic techniques. However, human beings are limited in terms of speed and space as comparing to computers. That is, computers are much better than humans to do these tasks. Most of the existing data in any organization is represented in a text format, so if we compare data mining with text mining then text mining is more important [29]. But as text mining is used for structuring the unstructured text data then this task is more demanding as compared to data mining. In general, the data related to social media sites is not collected for the research purpose [30], it is mandatory to change the structure of the data coming from the social media. 80% of the available text on the web is unstructured while only 20% is structured [31].

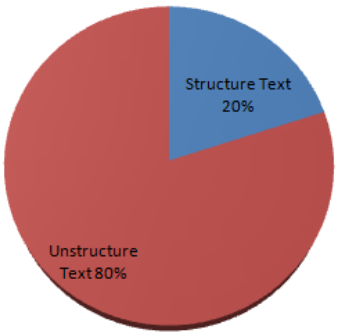

Figure 1: Text available on the web.

2.2. Text Mining vs. Data Mining

In the case of posting comments on any post on different social networking sites, there is not a single structured technique available which causes problems in the direct usage of the data. Data available in the text format has much more importance and that is why text mining is generating much business value [32]. A study by [33] stated that data mining represents the derivation of a meaningful pattern or principles from a spatial database for determining a particular issue or issues. Data mining is different from text mining [34]. A study by [35] pointed out that text mining is much more complex than data mining because it contains irregular and unstructured data patterns, whereas data mining is dealing with the structured sets of data. The tools that were used in data mining were only dealing with structured data [34]. Text mining is like an intelligence system which is extracting proper words or sentences from the improper words and then transforming those words into the particular suggestions. Text mining is basically a new field having the main purpose of data recovery, machine learning, information mining and computational linguistics [36].

2.3. Text mining in social networks

The importance of text mining has been increased due to the significant contributions in the field of technology. Data mining

as reported by [37] is also important but due to the advancement, text mining is taking its place. It is really a big effort to convey valuable information and knowledge [38] through powerful handling and mining processes from the irregular form of information. In this era, structured data has lost its importance and the unstructured data has gained the popularity. Most of the organizations are going towards text mining and forgetting the concept of data mining [39]. Scholars of [40] reported that all the social networking sites are providing a great space to individuals to facilitate interaction and share their views and opinions. The best thing which these sites are doing is that it has become easy for the individuals to understand a particular person depending upon his or her activities. Through all these activities, people related to different customs and values have come closer to each other because of having the better understanding of each other's emotions, perceptions and areas of interest. At this time, user interfaces are going to be equipped with personality based qualities [41]. Personalized designs were used in e-commerce [42], [43], e-learning, and information filtering for enhancing different styles and skills.

3. Text mining efforts in resolving various NLP issues

A study by [44] stated that text mining is responsible for structuring the irregular data patterns written in the human language. As most of the people interact with each other in the form of text so for those people who are not able to share structured form of data, text mining is the best technique to handle these situations. Among others, NLP is considered as the most amazing research field. The main goal of NLP is to seek information regarding how the computer systems are examining and getting information from the languages of human beings to create applications of high quality [17]. The art of sharing meaningful information with the help of uncommon and meaningless data is truly a good thing. Text mining technique as described by [45] examines the content for extracting the meaningful data which can be used for particular purposes. It looks like text mining that is going to include the overall NLP scheme [46] in its system in order to effectively examine the human language and to structure the unstructured data patterns accordingly. As the technology is advancing day by day, text mining system will get better and better and this is what all people are looking for.

3.1. Text mining in Facebook

The social networks are growing at a rapid rate without a break. Most importantly, the unstructured data is being stored on these networks as they act as a large pool and this data pertains to a host of domains containing governments, businesses, and health. Data mining techniques tend to transform the unstructured data for its placement within a systematic arrangement [47]. Nowadays, Facebook is one of the most popular social media. This media is used by a large number of people on earth for expressing their ideas, thoughts, sorrows, pleasures and poems [48]. Researchers had chosen a number of Facebook variables that were expected to develop the right situation for carrying out our investigations. The valuable statistics of user's personality is provided by the Facebook profiles and activities, which exposes the actual objects instead of projected or idealized character [49]. The digital data has currently witnessed an enormous growth. The key area of interest among professionals is now data mining and knowledge discovery. Moreover, a strong need has been felt to transform such data into useful knowledge and information. A number of applications like business management and market analysis have realized the benefits from the information and knowledge extracted out of large scale data. Information is stored in text form across various applications so one of the up-to-date areas for research is text mining. The hard issue is extracting the user required information. The knowledge discovery process has an important step which is believed to be the Text Mining. The hidden information is extracted from unstructured to semi-structured data in this process. Extracting information from a number of written resources and its automatic discovery is called as Text mining. Moreover, computers are also used for the needful and to meet this goal.

Scholars of [50] illustrated the text mining techniques, methods, and challenges. These successful techniques would be described to give usefulness over information acquisition during text mining. The study discussed the situations where each technology could be beneficial for a different number of users. A number of business organizations would be examined by mining data that has been exposed by their employees on LinkedIn, Facebook, and other openly available sources. A network of informal social connections among employees is extracted through web crawler developed for this purpose. According to the findings, leadership roles can be identified within the organization and this could be achieved absolutely by using machine learning techniques besides centrality analysis. Clustering the social network of an organization and collecting available information within each cluster can result in the valuable non-trivial perceptions. A key asset or a considerable threat to the primary organization can be the knowledge about the network of informal relationships. Besides analyzing social networks of the organizations, algorithms and methods used to gather data from freely available sources would be presented by this paper. A web crawler was developed to obtain profiles of employees from six targeted organizations and this was done by collecting the Facebook data. A social network topology was created for each organization, and machine-learning algorithms and centrality measures were implemented so that the hidden leadership positions within each company could be discovered. Moreover, the social community clusters inside these organizations were also revealed by the algorithms, which gave us understanding about the communication network of each company in addition to the structure of the organization.

According to a study by [51], it has become clear that social media data is simply susceptible to misuse. The scheme encompasses structured approach and its application. Furthermore, it entails performing a statistical cluster analysis in addition to the comprehensive analysis of social media comments so that researchers could determine the inter-relationships among key factors. The qualitative social media data can be quantified by these schemes and subsequently cluster them based on their similar features, and then they can be used as decision-making tools. The SAMSUNG Mobile Facebook page, where Samsung smartphones were introduced, was used for the data acquisition process. The comment published by Facebook users on the captioned Facebook page is referred to as the "Data". In a period of 3 months, almost 128371 comments were downloaded. The English comments only were undergone through the analysis process. Afterward, the

conceptual analysis was used by the content analysis and ultimately statistical cluster analysis was performed by carrying out relational analysis. Hence, social media data is integrated by applying the statistical cluster analysis and it is performed based on the output of the conceptual analysis. The researchers are consequently enabled to categorize a large dataset into many subsets, at times, referred to as objects. One of the disciplines of its application is marketing. Factors that can be manageable in some cases are also minimized by these types of techniques.

A study by [52] explored the social data as a systematic data mining architecture. Findings indicated that Facebook as a social networking site is the major source of data. Besides this approach, information on "my wall" post regarding myself, age and comments from the Facebook all are emphasized by the author. It has been taken as a raw data, which is applied later to study and monitor the analytical tactics. In addition, the study investigated images for the advertisement of their products and for the decision-making process. A number of data mining techniques precede the coercion of intellectual knowledge from social data. Mainly, it organizes the key information and other applied activities in which users are attributed regarding their colleagues on social networking sites (i.e. Facebook). For the recovery on Facebook user database, Facebook API performs Application Secret key and Facebook API Key are executed by Facebook API. As a result, WEKA files and data mining techniques are supported to collect certain data into the secondary database, while the text data is represented by the detached data.

Researchers of [41] explored the applicability of representing user's personality based on the extracted features from the Facebook data. The classification techniques and their utilities were completely analyzed with regard to the inspirational research outcomes. A sample of 250 user instances from Facebook formed the research study and this sample was from about 10,000 status updates, which was delivered by the My Personality project [53]. The study has the following two interconnected objectives: (1) having knowledge about the pertinent personality-correlated indicators that presents user data implicitly or explicitly in Facebook, and (2) identifying the feasibility of prognostic character demonstration so that upcoming intelligent systems could be supported. The study emphasized on the promotion of pertinent features in a model, through which the enhanced output of the classifiers under evaluation could be observed.

3.2. Text mining in Twitter

A significant size of research has been occupied by the Twitter data analysis over the last couple of years [54]. Large spectrums of domains are using this data, some of which are using it for academic research and others for applications [55]. New improvements regarding twitter data are presented by this section. The document collection from various resources triggers the "Text Mining" process. A particular document would be retrieved by Text mining tool and this document is pre-processed by checking the character sets and format [56]. Subsequently, a text analysis phase would monitor the document. Semantic analysis is used to derive high-quality information from text; this is referred to "Text analysis". The market has a lot of text analysis techniques. Professionals can use combinations of techniques subject to the goal of the organization. Researchers tend to repeat the text analysis techniques till the time information is acquired. A management information system is capable of incorporating the resulting information, and as a result, significant knowledge is produced for the user of that information system [57]. A key issue in text mining is intricacy of natural language. The ambiguity problem is much dense in the natural language. There are multiple meanings of a single word and multiple words can possess same meaning. Ambiguity is referred to as the understanding of a word which has more than one possible meaning. Noise has emerged in extracted information as a result of this ambiguity. Since usability and flexibility are the main parts of ambiguity, it cannot be removed from the natural language. One phrase or sentence can have multiple understandings, so there is a chance we can obtain a number of meanings. The work is still undeveloped and a particular domain is correlated with the suggested approach while the experts have attempted to resolve the ambiguity problem by performing a number of research studies. As there is uncertainty/vagueness in the semantic meanings of many discovered words, so it is very difficult to answer the requirements of the user.

Scholars of [58] developed and formulated an automatic classification technique through which potentially abuse-indicating user posts could be identified and evaluating the likelihood of social media usage as a source for automatic monitoring of drug medication abuse. In this regard, Twitter user posts (tweets) were collected and these were linked with three commonly abused medications (Oxycodone, Adderall, and Quetiapine). Besides interpreting a control medication (metformin), which is not the subject of abuse due to its process, nearly 6400 tweets were manually annotated, where these three medications were pointed out. The annotated data was qualitatively and quantitatively analyzed to determine as to whether or not signals of drug medication abuse are presented in Twitter posts. To sum up, Twitter's value was assessed in exploring the patterns of abuse over time and an automatic supervised classification technique was also designed, in which the purpose was to observe and separate the posts containing signals of medication abuse from those that do not. According to the findings of investigations, Twitter posts have yielded clear signals of medication abuse. As compared to the proportion for the control medication (i.e., metformin: 0.3 %), there is a very high ratio of tweets containing abuse signals for the three case medications (Adderall: 23 %, oxycodone: 12 %, quetiapine: 5.0 %). In addition, almost 82 % accuracy (medication abuse class recall: 0.51, precision: 0.41, F-measure: 0.46) has been achieved through the automatic classification approach. The Study demonstrated how the abuse patterns over time can be analyzed by using the classification data and its goal is to illustrate the effectiveness of automatic classification. As a result, it is found that abuse-related information for medications can be significantly acquired from social media, and the research indicates that natural language processing and supervised classification are the automatic approaches that have potentials for future monitoring and intervention assignments. With respect to supervised learning, the lack of sufficient training data is believed to be the largest shortcoming of the study. Both annotation and automatic classification are hindered by the lack of context and ambiguity in tweets. During the course of annotations, many ambiguous tweets were found and services of pharmacology expert were hired to address these issues. As a result of these ambiguities, the undefined situation is observed in the binary classification process and this inadequacy will continue until the time fine-tuned annotation rules could be specified by the future annotation rules.

A study by [59] applied the text mining approach on a large dataset of tweets. The complete Twitter timelines of 10 academic

libraries were used to collect the dataset for this research. Nearly 23,707 tweets formed the total dataset, where there were 7625 hashtags, 17,848 mentions, and 5974 retweets. Inconsistency among academic libraries is found in the distribution of tweets. "Open" was the most repeated word that was used by the academic libraries in different perspectives. It was observed that "special collections" was the most frequent bigram (two-word sequence) in the aggregated tweets. While "save the date" was the most recurrent tri-gram (three-word sequence). In the semantic analysis, words such as "insight, knowledge, and information about cultural and personal relations" were the most frequent word categories. Moreover, "Resources" was the most widespread category of the tweets among all the selected academic libraries. The significance of data and text-mining approaches are reported within the study and their purpose is to gain an insight with the aggregate social data of academic libraries so that the process of decision-making and strategic planning could become facilitated for marketing of services and patron outreach. The 10 academic libraries from top global universities have undergone the text mining approach. The study aimed to illustrate their Twitter usage and to examine their tweet content.

As far as social media is concerned, decision-making is supported and user-generated text is analyzed through text mining and content analysis [60]. By employing an archiving service (twimemachine.com) in December 2014, the complete Twitter timelines of 10 academic libraries were taken into account to collect the dataset for this research. The libraries of 10 highest-ranking universities from the global Shanghai Ranking were chosen for that purpose. The language of the university must be English-based, which was the condition for selection and selection was restricted to only one library if there was more than one library in the university. Certain weaknesses were found in the study, for example, all of the libraries are English-language libraries in the sample and only 10 academic libraries were considered for the analysis. This gap must be filled in future by applying the analysis to a dataset from diversified academic libraries, including non-English language libraries. Consequently, a complete understanding of tweet patterns would be acknowledged. The future inquiry can also incorporate the international or cross-cultural comparisons. Any discrepancy among libraries in their tweets' content affected by the number and interaction of followers could be highlighted by the analysis and its findings. The accuracy of the tweet categorization tool has yielded the inadequate findings, and the said tool needs to be substantiated through other machine-learning models along with their applications.

Researchers of [55] demonstrated in a smoking cessation nicotine patch study an innovative Twitter recruitment system that is deployed by the group. The study aimed to describe the methodology and used to address the issue of digital recruitment. Furthermore, designing a rule-based system with the provision of system specification besides representing the data mining approaches and algorithms (classification and association analysis) using Twitter data. Twitter's streaming API captured two sets of streaming tweets, which were collected for the study. Ten search terms, (i.e. quitting, quit, nicotine, smoking, smoke, patches, cig, cigarette, ecig, cigs, marijuana) were used to gather the first set. The second set of tweets contains 30 terms, in which the terms from the first set were included. Moreover, the second set is a superset of the first one. A number of studies have been conducted to review the information gathering methods. As unstructured data sets are in the textual format, the use of various procedures of text mining has been tackled by many research studies. Nonetheless, the data sets on the social networking websites are not mainly discussed by these studies. A study by [50] applied various text mining techniques. The study would describe the application of these strategies in the social networking websites. In the field of intelligent text analysis, the latest improvements would also be examined in the survey. The study focused on two key techniques pertaining to the text mining field, namely classification and clustering. Usually, they are operated for the study of the unstructured text accessible on the extensive scale frameworks. Prior to the start of World Cup, a total of approximately 30,000 tweets were used by [61]. Moreover, an algorithm was used for integrating the consensus matrix and the DBSCAN algorithm. Consequently, the concerned tweets on those prevailing topics were available to him. Afterward, the clustering analysis was applied to seek the topics discussed by the tweets. The tweets were grouped utilizing the k-means [62], Non-Negative Matrix Factorization (NMF), and a popular clustering algorithm. After that, the results were compared. Similar results were delivered by both algorithms. However, NMF became faster and the researchers could easily interpret the outcomes.

A study by [1] initiated a workflow to gain an insight into both the large-scale data mining methods and qualitative analysis. Twitter posts of engineering students were the primary concern. The basic goal was to identify their issues in their academic experiences. The study conducted a qualitative analysis of samples obtained from around 25,000 tweets that were associated with the engineering students and their college life. The encounter troubles of engineering students were discovered during the study. For example, a large volume of study, sleep deprivation and lack of social engagement. Considering these outcomes, a multi-label classification algorithm was implemented to categorize tweets in lieu of students' queries. The algorithm was applied on approximately 35,000 tweets streamed at the geo-location of Purdue University. At the first instance, the concerned authorities have addressed the experiences and issues of the students and social media data was used to expose the issues. Moreover, a study by [1] also developed a multi-label classifier so that tweets founded within the content evaluation phase could be organized. A number of renowned classifiers are significantly consumed in machine learning domain and data mining process. With Comparison to other state-of-the-art multi-label classifiers, the Naïve Bayes classifiers were found proficient on the dataset.

A study by [63] discussed the clustering technique, the execution of correlation and association analyses to social media. The investigation of insurance Twitter posts was carried out to assess this matter. Consequently, recognizing theories and keywords in the social media data has become an easy task, due to which the information by insurers and its application would be facilitated. After having a detailed analysis, client queries and the potential market would be proactively addressed with usefulness and the findings of the analysis are to be effectively implemented in suitable fields. According to this evaluation, the overall 68,370 tweets were utilized. Two additional kinds of evaluation need to be applied to the data. The first is the clustering analysis, through which the tweets depending on their similarities or dissimilarities would be merged. An Association Analysis is the second one whereas the occurrences of particular composed words were discovered.

Authors of [64] stated that sentiment analysis through social media usage has witnessed a huge interest from scholars in the last few years. In that, the authors discussed the influence of tweets'

sentiment on elections and the impact of the elections' results on web sentiment.

4. Conclusion and Future work

The method of communication with each other has now completely changed due to the progress in the field of social media. Nowadays, modernization can be seen everywhere and based on that; the information production is touching the altitudes. Currently, the new companies are moving forward to take an active part in transforming the communication method [65]. The keywords and phrases' particularization can become helpful to different companies in order to shape their future. In the present study, we have highlighted the state-of-the-art research work regarding the implementation of text mining in the most dominant social media (Facebook and Twitter). From the point of view of several scholars, Text mining was explained through various models. Moreover, different authentic references are also provided to support the research work. As a result, text mining can be classified into text clustering, text categorization, association rule extraction and trend analysis according to applications. With the passage of time, text mining is going to be progressed well.

We can observe from the surveyed literature that Arabic text in social media is overlooked from the point of view of several text mining studies. As a result, this gap opens the door for many text mining scholars to bridge that gap through conducting various studies in the field of text mining in the Arabic language context. A study by [66] argued that researchers analyzing the Arabic post are seldom found, focusing on the text mining of English, albeit the Arabic post on social media is present in bulk amount. Scholars of [67] outlined its strange and peculiar characteristics as the reasons behind this attitude. From the surveyed literature, we have observed that researchers have paid less attention to sentiment analysis in the Arabic text. The sophisticated tasks of parsing and sense disambiguation fortify production of target lists of the most recurrent grammatical structures and senses of polysemous words, and the potential for syntactic and semantic ambiguity is found to be high [68]. As a future work, we are highly interested in examining the text mining techniques on Arabic textual data from Facebook and Twitter. In addition, future research should take sentiment analysis of Arabic text into consideration. The Arabic language is convoluted morphologically, possesses free word order, punctuation seldom found and short vowels are avoided in the written form of Standard Arabic. Hence, context is essential to eradicate prevailing ambiguity from apparently identical forms which is significant in recognizing opinions.

References

[1] Chen, X., Vorvoreanu, M., & Madhavan, K. (2014). Mining social media data for understanding students' learning experiences. *IEEE Transactions on Learning Technologies*, *7*(3), 246-259.

[2] Buettner, R. (2016). Predicting user behavior in electronic markets based on personality-mining in large online social networks. *Electronic Markets*, 1-19.

[3] Baumer, E. P., Sinclair, J., & Tomlinson, B. (2010, April). America is like Metamucil: fostering critical and creative thinking about metaphor in political blogs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1437-1446). ACM.

[4] Stieglitz, S., Dang-Xuan, L., Bruns, A., & Neuberger, C. (2014). Social media analytics. *Wirtschaftsinformatik*, *56*(2), 101-109.

[5] Kasture, N. R., & Bhilare, P. B. (2015, February). An Approach for Sentiment analysis on social networking sites. In *Computing Communication Control and Automation (ICCUBEA), 2015 International Conference on* (pp. 390-395). IEEE.

[6] Sorensen, L. (2009, May). User managed trust in social networking-Comparing Facebook, MySpace and Linkedin. In *Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, 2009. Wireless VITAE 2009. 1st International Conference on* (pp. 427-431). IEEE.

[7] Naaman, M. (2012). Social multimedia: highlighting opportunities for search and mining of multimedia data in social media applications. *Multimedia Tools and Applications*, *56*(1), 9-34.

[8] Evans, B. M., Kairam, S., & Pirolli, P. (2010). Do your friends make you smarter?: An analysis of social strategies in online information seeking. *Information Processing & Management*, *46*(6), 679-692.

[9] Li, J., & Khan, S. U. (2009, November). MobiSN: Semantics-based mobile ad hoc social network framework. In *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE* (pp. 1-6). IEEE.

[10] Liu, X., Wang, M., & Huet, B. (2016). Event analysis in social multimedia: a survey. *Frontiers of Computer Science*, 1-14.

[11] Tsytsarau, M., & Palpanas, T. (2012). Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, *24*(3), 478-514.

[12] Al-Emran, M., & Malik, S. I. (2016). The Impact of Google Apps at Work: Higher Educational Perspective. *International Journal of Interactive Mobile Technologies (iJIM)*, *10*(4), 85-88.

[13] Al-Emran, M. N. H. (2014). Investigating Students' and Faculty members' Attitudes Towards the Use of Mobile Learning in Higher Educational Environments at the Gulf Region.

[14] Al-Emran, M., & Shaalan, K. (2015, August). Learners and educators attitudes towards mobile learning in higher education: State of the art. In *Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on* (pp. 907-913). IEEE.

[15] Irfan, R., King, C. K., Grages, D., Ewen, S., Khan, S. U., Madani, S. A., ...& Tziritas, N. (2015). A survey on text mining in social networks. *The Knowledge Engineering Review*, *30*(02), 157-170.

[16] Berry Michael, W. (2004). Automatic Discovery of Similar Words. *Survey of Text Mining: Clustering, Classification and Retrieval", Springer Verlag, New York*, *200*, 24-43.

[17] Salloum, S. A., Al-Emran, M., & Shaalan, K. (2016). A Survey of Lexical Functional Grammar in the Arabic Context. *Int. J. Com. Net. Tech*, *4*(3).

[18] Al Emran, M., & Shaalan, K. (2014, September). A Survey of Intelligent Language Tutoring Systems. In *Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on* (pp. 393-399). IEEE.

[19] Al-Emran, M., Zaza, S., & Shaalan, K. (2015, May). Parsing modern standard Arabic using Treebank resources. In *Information and Communication Technology Research (ICTRC), 2015 International Conference on* (pp. 80-83). IEEE.

[20] Alfaro, C., Cano-Montero, J., Gómez, J., Moguerza, J. M., & Ortega, F. (2016). A multi-stage method for content classification and opinion mining on weblog comments. *Annals of Operations Research*, *236*(1), 197-213.

[21] Yang, L., Geng, X., & Liao, H. (2016). A web sentiment analysis method on fuzzy clustering for mobile social media users. *EURASIP Journal on Wireless Communications and Networking*, *2016*(1), 1.

[22] Robinson, R., Goh, T. T., & Zhang, R. (2012). Textual factors in online product reviews: a foundation for a more influential approach to opinion mining. *Electronic Commerce Research*, *12*(3), 301-330.

[23] Rahmani, A., Chen, A., Sarhan, A., Jida, J., Rifaie, M., & Alhajj, R. (2014). Social media analysis and summarization for opinion mining: a business case study. *Social Network Analysis and Mining*, *4*(1), 1-11.

[24] Grimes, S. (2008). Unstructured data and the 80 percent rule. *CarabridgeBridgepoints*.

[25] Hung, J. L., & Zhang, K. (2012). Examining mobile learning trends 2003–2008: A categorical meta-trend analysis using text mining techniques. *Journal of Computing in Higher education*, *24*(1), 1-17.

[26] Feldman, R., & Dagan, I. (1995, August). Knowledge Discovery in Textual Databases (KDT). In *KDD* (Vol. 95, pp. 112-117).

[27] Rajman, M., & Besançon, R. (1998). Text mining: natural language techniques and text mining applications. In *Data mining and reverse engineering* (pp. 50-64). Springer US.

[28] Gök, A., Waterworth, A., & Shapira, P. (2015). Use of web mining in studying innovation. *Scientometrics*, *102*(1), 653-671.

[29] Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2005). Tapping into the power of text mining.

[30] SØRENSEN, H. T., Sabroe, S., & OLSEN, J. (1996). A framework for evaluation of secondary data sources for epidemiological research. *International journal of epidemiology*, *25*(2), 435-442.

[31] Zhang, J. Q., Craciun, G., & Shin, D. (2010). When does electronic word-of-mouth matter? A study of consumer product reviews. *Journal of Business Research*, *63*(12), 1336-1341.

[32] Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. *Journal of emerging technologies in web intelligence*, *1*(1), 60-76.

[33] Zaza, S., & Al-Emran, M. (2015, October). Mining and Exploration of Credit Cards Data in UAE. In *2015 Fifth International Conference on e-Learning (econf)* (pp. 275-279). IEEE.

[34] Navathe, S. B., & Ramez, E. (2000). Data warehousing and data mining. *Fundamentals of Database Systems*, 841-872.

[35] Akilan, A. (2015, February). Text mining: Challenges and future directions. In*Electronics and Communication Systems (ICECS), 2015 2nd International Conference on* (pp. 1679-1684). IEEE.

[36] Sukanya, M., & Biruntha, S. (2012, August). Techniques on text mining. In*Advanced Communication Control and Computing Technologies (ICACCCT), 2012 IEEE International Conference on* (pp. 269-271). IEEE.

[37] Piatetsky-Shapiro, G. (2007). Data mining and knowledge discovery 1996 to 2005: overcoming the hype and moving from "university" to "business" and "analytics". *Data Mining and Knowledge Discovery*, *15*(1), 99-105.

[38] Weninger, T. (2014). An exploration of submissions and discussions in social news: mining collective intelligence of reddit. *Social Network Analysis and Mining*, *4*(1), 1-19.

[39] Chakraborty, G., & Krishna, M. (2014). Analysis of unstructured data: Applications of text analytics and sentiment mining. In *SAS global forum* (pp. 1288-2014).

[40] Stieglitz, S., & Dang-Xuan, L. (2013). Social media and political communication: a social media analytics framework. *Social Network Analysis and Mining*, *3*(4), 1277-1291.

[41] Markovikj, D., Gievska, S., Kosinski, M., & Stillwell, D. (2013, June). Mining facebook data for predictive personality modeling. In *Proceedings of the 7th international AAAI conference on Weblogs and Social Media (ICWSM 2013), Boston, MA, USA*.

[42] Zhang, Y., & Yu, T. (2012). Mining trust relationships from online social networks. *Journal of Computer Science and Technology*, *27*(3), 492-505.

[43] Yan, B. N., Lee, T. S., & Lee, T. P. (2015). Analysis of research papers on E-commerce (2000–2013): based on a text mining approach. *Scientometrics*,*105*(1), 403-417.

[44] Witten, I. H. (2005). Text mining. *Practical handbook of Internet computing*, 14-1.

[45] Steinberger, R. (2012). A survey of methods to ease the development of highly multilingual text mining applications. *Language Resources and Evaluation*, *46*(2), 155-176.

[46] Schoder, D., Gloor, P. A., & Metaxas, P. T. (2013). Special Issue on Social Media. *KI*, *27*(1), 5-8.

[47] Injadat, M., Salo, F., &Nassif, A. B. (2016). Data mining techniques in social media: A survey. *Neurocomputing*.

[48] Kamal, S., & Arefin, M. S. (2016). Impact analysis of facebook in family bonding. *Social Network Analysis and Mining*, *6*(1), 1-14.

[49] Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological science*.

[50] Fire, M., & Puzis, R. (2012). Organization mining using online social networks. *Networks and Spatial Economics*, 1-34.

[51] Chan, H. K., Lacka, E., Yee, R. W., & Lim, M. K. (2014, December). A case study on mining social media data. In *2014 IEEE International Conference on Industrial Engineering and Engineering Management* (pp. 593-596). IEEE.

[52] Rahman, M. M. (2012). Mining social data to extract intellectual knowledge. *arXiv preprint arXiv:1209.5345*.

[53] Celli, F., Pianesi, F., Stillwell, D., & Kosinski, M. (2013, June). Workshop on computational personality recognition (shared task). In *Proceedings of the Workshop on Computational Personality Recognition*.

[54] Reips, U. D., & Garaizar, P. (2011). Mining twitter: A source for psychological wisdom of the crowds. *Behavior research methods*, *43*(3), 635-642.

[55] Hamed, A. A., Wu, X., & Rubin, A. (2014). A twitter recruitment intelligent system: association rule mining for smoking cessation. *Social Network Analysis and Mining*, *4*(1), 1-19.

[56] Dey, L., & Haque, S. M. (2009). Opinion mining from noisy text data. *International Journal on Document Analysis and Recognition (IJDAR)*, *12*(3), 205-226.

[57] Gaikwad, S. V., Chaugule, A., & Patil, P. (2014). Text mining methods and techniques. International Journal of Computer Applications, 85(17).

[58] Sarker, A., O'Connor, K., Ginn, R., Scotch, M., Smith, K., Malone, D., & Gonzalez, G. (2016). Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from Twitter. *Drug safety*, *39*(3), 231-240.

[59] Al-Daihani, S. M., & Abrahams, A. (2016). A Text Mining Analysis of Academic Libraries' Tweets. *The Journal of Academic Librarianship*, *42*(2), 135-143.

[60] Abrahams, A. S., Fan, W., Wang, G. A., Zhang, Z. J., & Jiao, J. (2015). An integrated text analytic framework for product defect discovery. *Production and Operations Management*, *24*(6), 975-990.

[61] Godfrey, D., Johns, C., Meyer, C., Race, S., & Sadek, C. (2014). A case study in text mining: Interpreting twitter data from world cup tweets. *arXiv preprint arXiv:1408.5427*.

[62] Velez, D., Sueiras, J., Ortega, A., & Velez, J. F. (2015). A method for K-Means seeds generation applied to text mining. *Statistical Methods & Applications*, 1-23.

[63] Mosley Jr, R. C. (2012). Social media analytics: Data mining applied to insurance Twitter posts. In *Casualty Actuarial Society E-Forum*, *Winter 2012 Volume 2* (p. 1).

[64] Kermanidis, K. L., & Maragoudakis, M. (2013). Political sentiment analysis of tweets before and after the Greek elections of May 2012. *International Journal of Social Network Mining*, *1*(3-4), 298-317.

[65] Wu, J., Sun, H., & Tan, Y. (2013). Social media research: A review. *Journal of Systems Science and Systems Engineering*, *22*(3), 257-282.

[66] Habash, N. (2010). Introduction to Arabic natural language processing. Synthesis Lectures on Human Language Technologies, 3(1), 1-187, Morgan & Claypool Publishers.

[67] Shaalan, K., Abo Bakr, H., & Ziedan, I. (2007). Transferring Egyptian Colloquial into Modern Standard Arabic, International Conference on Recent Advances in Natural Language Processing, PP. 525-529, Bulgaria.

[68] Farghaly, A., Shaalan, K. (2009). Arabic Natural Language Processing: Challenges and Solutions. ACM Transactions on Asian Language Information Processing (TALIP), ACM, 8(4), 1-22.