Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 120 (2017) 664–670

Procedia Computer Science

www.elsevier.com/locate/procedia

9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, ICSCCW 2017, 24-25 August 2017, Budapest, Hungary

Applying fuzzy logic for sentiment analysis of social media network data in marketing

Karen Howells[a]*, Ahmet Ertugan[a]

[a]*Near East University, Faculty of Economics and Administrative Sciences, POBOX:99138, Nicosia, North Cyprus, MERsin 10 Turkey*

Abstract

The amount of data freely available from social networking grows on an hourly basis. Much of this data concerns consumers perceptions and opinions of organizations, and as such is of interest to business intelligence gatherers in marketing, for customer relationship management and customer retention. With the use of soft computing, specifically fuzzy logic, it will be possible to design, create and build social bots that can analyse consumer comments in social media networks. Further programming would allow these social bots to interact with consumers, and carefully produced social bots would be able to disseminate marketing campaigns. This paper proposes a model for sentiment analysis of social media network data.

Peer-review under responsibility of the scientific committee of the 9th International Conference on Theory and application of Soft Computing, Computing with Words and Perception.

Keywords: soft computing application; fuzzy logic; social bots; sentiment analysis; social media networks; customer relationship management.

1. Introduction

The most commonly used social media networks are Facebook, YouTube, Instagram and Twitter. Currently, there are approximately 500 million tweets on Twitter per day, that's 6000 every second. (Internetlivestats, 2017). The exact numbers of users are shown in Table 1 below. Marketers use social media as part of their marketing strategy. Before planning a social media campaign, marketers need to know something about the consumer and their environment. They need to know the personality of the consumer, their past experiences, their responses to

* Corresponding author: Karen Howells Tel: +90-392-675-1000(3125); fax: +90-392-675-1051
E-mail address: karen.howells@neu.edu.tr

1877-0509
Peer-review under responsibility of the scientific committee of the 9th International Conference on Theory and application of Soft Computing, Computing with Words and Perception.
10.1016/j.procs.2017.11.293

campaigns, brand loyalty, and perception of brands. This is very important as marketers base decisions on what they know about their target audience (Tuten & Solomon, 2014). The more we understand as marketers, the better decisions we can make. Gathering up information about consumers to create insight into markets, and to analyse competitive intelligence are very important steps in creating a marketing strategy for the organisation. More information means more informed decisions. Social media networks have given us an additional source of information on consumers, their demographics, their perceptions, and their thoughts. The content is created by the users of social media networks, and includes their opinions, experiences, and this data can be in the form of binary input ('likes' on a Facebook post), text, audio or visual. These items are shared, sometimes across social media platforms, and a conversation will build around them.

Table 1. Social media metrics.

Social media network	No of users	Daily metrics
Facebook	1.94 billion	
YouTube	1 billion	
Instagram	700 million	
Twitter	313 million	500 million tweets
Reddit	250 million	
Viber	249 million	
Vine Camera	200 million	
Pinterest	150 million	
Snapchat	200 million	

(Kallas, 2017)

Facebook “likes” allow users of social networks to express their support of a post by a click of the mouse. The notion of “liking” has been extended to "love," "haha," "wow," "sad," and "angry". The emoji selected corresponds to a feeling of the use at that point in time. The notion of emojis began in Japan in 1999, where a graphical response or post could be coded as one of 722 Unicode character codes (Perticone et al., 2015). Many of these emojis represent nuances of Japanese culture, and have often been misunderstood or used in a different way by Western users of social media. Many users choose to use emojis for communication, rather than text, and this data can also be collected as business intelligence.

With current technology it is very simple to gain access to the millions of public tweets through Twitter, and to the databases of other social media such as Facebook. Software, known as social bots, can be set up to trawl through these massive databases, using key words to lead data analysts to possible data relating to their organisation.

In the different types of social media, Twitter is known as microblogging, as it has a limited character size for tweets of 140 characters. It is easy to search through the millions of tweets tweeted everyday, and the users' use of hashtags makes the available data easy to categorise.

While the body of literature contains examples of ideas on how to use artficial intelligence and neural networks to analyse tweets, there is little in the form of analysing tweets through fuzzy logic.

Fig. 1. Current popular social media network icons

When analysing short, informal texts, such as those used in social media networks, the emojis used provide further information and context on the data. In 2015 it was reported that over half the posts on Instagram contained emojis (Novac et al., 2015). In addition to emojis, certain abbreviations can be analysed in terms of emotion, such as the acronym LOL meaning Laugh Out Loud, and could be categorised as positive.

One issue of analysing the vast amount of data from social media networks such as Twitter, is that the English used is very informal. Colloquialisms are used, as well as slang, and other unconventional language and shortened forms or abbreviations. While a human can read and understand a particular tweet, based on context and prior knowledge, it is likely that software would not be able to organise a tweet into a particular category.

From a technical point of view, there are a number of challenges to analysing the data from users in social media networks, including Twitter, in that this social media contains rich data sets or metadata. There are tags, #hashtags, or labels using words that users make up themselves by joining a number of words together, as well as emojis, and the above mentioned issues with text.

1.1. Sentiment analysis

Sentiment analysis is the determination of a piece of text to be positive, neutral or negative in meaning. When applied to the comments of social media network users, we can attribute their views as being positive, neutral or negative. This is an important piece of detective work when we are an organisation wondering what our customers, or stakeholders are saying about us.

1.2. Fuzzy Logic

Fuzzy Logic, or fuzzy thinking, has its roots in the early days of computer science, first suggested by Zadeh's work on fuzzy sets (Zadeh, 1965). In this ground breaking paper, Zadeh posits that the real world cannot be described in a binary fashion, as it is far too complex, they are many grey areas, as well as data that can be described as black and white. While classical logic allows for only two values, true or false, 1 or 0, the Boolean values; fuzzy logic allows us to extend this to describe variables that are at best vague. It allows us approximate reasoning.

1.3. Applying Fuzzy Logic to Marketing in Social Media Networks

Marketing is often defined as both a science and as an art. There is theory, and there is creativity. In this sense, fuzzy logic is an ideal tool to apply to problems in marketing. The continual posting of opinions and decisions of consumers to social media networks creates a vast opportunity for organisations to collect information in relation to consumer perceptions and ideas about the organisation. More importantly, this information is timely, it reflects the mood at this moment it time. This huge data set of information can be seen as a grey area, as customer behaviour cannot always be squeezed into a value of yes or no, it varies, on a scale of greyness. The advantage of using fuzzy logic is that we can use linguistics to phrase a set, and this is based on if-then rules. The if-then rules use fuzzified variables, and there are different methods of showing this.

In marketing, fuzzy logic has been applied to a number of customer behaviour and customer satisfaction models, including product and service evaluation (Enache, 2015). Analysing the data from social media networks will allow marketers to gain new insights into their relationships with consumers. The fuzzy approach can elicit business intelligence.

It is possible to use a computer program to trawl through the huge amounts of data available in the Internet to find knowledge, using a mathematical model (ter Veen et al., 2011). Large databases of relevant information can be created by using web crawls through the available data of Twitter, Facebook, blogs, and RSS Feeds. Natural language processing tools such as OpenCalais can be used to identify the comments that are related to the information being sought (ter Veen et al., 2011). Secondary tools such as NetMiner can be used to analyse the data, to understand its network characteristics, such as is the data cohesive, or central, or distance of data relations (ter Veen et al., 2011).

2. Methodology

The objective of this study is to develop a model that can analyse the content of a microblog (such as tweet in Twitter) and be able to analyse customer feedback or perception. This model would then form the basis of a computer application which an organisation could use in confidence for marketers to develop a timely, and strong representation of customers' perceptions on a product or service.

The initial step in creating a fuzzy logic model is to transform the linguistic variables into membership functions that describe the variable in a fuzzy way. Each membership function represents a linguistic variable, and defines a fuzzy set that is associated with a particular parameter.

Once the fuzzy set is defined, the next step is to define the if-then rules that describe how the fuzzy sets interact. Fuzzy operators are introduced to create connections between the fuzzy sets, namely max (which represents *and*) and min (which represents *or*). On completion of the connections, all the propositions are aggregated into the final fuzzy set. This fuzzy set releases a fuzzy number, and this is transformed into a crisp value, through different techniques.

The simplest form of data created by consumers consists of text and emoji comments in the social media, typically Facebook posts or Twitter tweets. Software “bots”, or social bots, can be used to gather information from social media networks. Data can be gathered manually, through search, copy and paste, but it would be very time consuming. Users of social media networks make data easier to find with the use of hashtags and topic names, so that data could be organised through hashtag use, even across social media platforms. Whether manually, or through the use of social bots, the user created data will be saved to a database, creating a data warehouse of information, to be analysed through data mining methods.

Currently social bots have recently gathered a reputation as being menacing (Ferrara et al., 2016), as they can be used to modify user data and to deceive social media network users, posing as real life people with fake accounts.

2.1. Text analysis

A human reader can quickly read and analyse thousands of comments by social media network users, and classify them into a sentiment category. Due to the extremely high number of posts or tweets in the social medi networks, this is not practical. However, for a computer to process the same data, it is not straight forward. Perticone et al. (2015) suggest only using six sentiments, the ones that are universally recognised and used as Facebook actions, but as mentioned above the classic emoji set of emoticons is 722. Bing and Chan (2014) highlight the difficulties of parsing the information from social media, and suggest a matrix based fuzzy algorithm. Most of the 722 emojis can be directly parsed into some meaningful feeling. Hashtags and text in a tweet can be parsed as natural language. However, the dictionary with which the text will be compared will be continually evolving to include new hashtags,

and forms of speech.

A number of ready made applications are available in the public domain that will assist a marketer in collecting specific information from social media networks.

2.2. Existing Open Source Intelligence Collection Software

Open Source Intelligence (OSINT) is a form of intelligence collection management that finds, selects, and acquires information from public sources such as Twitter and Facebook, and analyses it to produce intelligence that can be acted upon. There are free available tools such as Maltego. Creepy, Case File, and FoxOne Scanner.

Developed by Paterva, Maltego is an open source intelligence and forensics application, focusing on transforming data from open sources and visualizing it in a graph format, and is particularly suited to data mining and link analysis (Paterva, 2012). Its main focus is to map the real-world relationships between people, groups, websites , domains, networks, internet infrastructure and connections with social media. Java based, it can run on Windows as well as Linux. Kakavas created Creepy, a geolocation information aggregator, which will gather information on an individual, through their use of social networks, or from image hosting services to pinpoint their location [13]. The resulting analysis is presented as a map inside the application, showing where all the retrieved information stems from.

2.3. The Model

However, these available software applications stop short of the sentiment analysis needed for the professional marketer. Using a set of key words for searching, the marketer can reduce the dataset to a manageable size. The data would then need to be parsed semantically, and analysed in terms of sentiment. Let us assume that the marketer wishes to use data from Twitter, to look at consumer tweets about a particular brand of product or service. The model would need to find the tweets and parse the tweet to separate out the emojis and hashtags. The emojis can be classified through the use of a lookup table, while the hashtags can be used to create a classification. The natural text of the tweet would need to be lexically parsed, with the ability to process abbreviations, colloquialisms, slang and other linguistic features. The combination of emoji classification, hashtag classification, and textual meaning would be inputted to fuzzy logic. The output would reveal a classification for the tweet, to be stored in a database. On completing the analysis of a number of tweets, the model would calculate mean average of the tweet classifications.

This model builds on the project posited by Haque and Rahman (2014), in that in this model, both the parsed data structure and the output data structure have more elements.

In this model the tweets, as output from the fuzzy logic module, would be categorised into 5 distinct classifications:

- strongly positive tweets
- positive tweets
- neutral tweets
- negative tweets
- strongly negative tweets

Passing through the fuzzy logic module, the tweets would be scored from 0 to 1, and then assigned to the above classifications. The resulting descriptive statistics could then be represented graphically.

The steps of the proposed model are shown in Figure 2 below.

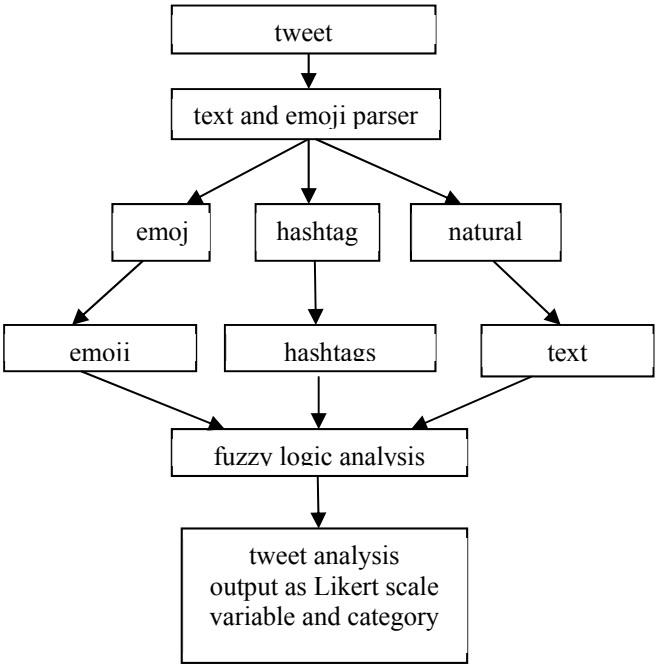

Fig. 2. Tweet Analysis Model

3. Results

Initally the tweets would be extracted and parsed. After this, through the application of fuzzy logic, the sentiment values of the tweets would be calculated. By having the values of the tweets, we can calculate the arithmetic mean. This would give us the percentage of postive sentiments, negative sentiments, and neutral sentiments. Currently, this method of applying fuzzy logic to sentiment analysis of social media network data is a model. The model as a fully functional software application has yet to be built.

The model is specifically aimed at applications in consumer relationship management, customer retention and other aspects of marketing. Currently social bots exist and interact with Twitter. They are not sophisticated, and are seen as puppets to be detected (Ferrara et al., 2016). In a project with computer science students at the University of Washington, Tacoma's Institute of Technology, McCarthy [11], social robots were constructed. The purpose was for these social robots to operate autonomously for 5 days, and gain as many Twitter followers as possible, and mentions, from the target population as possible

Fig. 3. Profile of one of the Socialbots (McCarthy, 2011)

The social bots of this experiment were very successful. Some 27% of the user population decided to follow one of the social bot accounts. 12% of the user population mentioned at least one of the social bots in a tweet. The social bots combined elements of Artificial Intelligence, along with social networking and network security, as well as aspects of technology, psychology, sociology, ethics and politics (McCarthy, 2011). The social bots were mainly written in Python 2.6, and Twitter API. Some even incorporated a Python bot based on the classic Eliza program.

4. Conclusion

Using social media and online social networking to gather intelligence information for customer relationship management is fairly new. It is, however, a very serious development in the use of such data in monitoring social media comments for business intelligence. Many organisations are now using social bots to trail for information on the Internet. The model proposed above would highlight the relevant information on customer perception of a brand. Further development on the model could include communication with the user, to elicit more data on a particular topic of interest to the marketer.

This objective of this paper is to suggest, that in combining social bots with data mining and fuzzy logic, it is possible to build a highly efficient tool that will not only find a target population, but will communicate and prompt the target for even more information. This would be an extremely useful tool in customer relationship management, as the marketer would be able to probe the user for more feedback on a particular subject.

References

Bing, L. & Chan, K.C.C. 2014. A Fuzzy Logic Approach for Opinion Mining on Large Scale Twitter Data, *2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing*, London, pp. 652-657.

Donzé, L., & Meier, A. 2012. Applying Fuzzy Logic and Fuzzy Methods to Marketing, in "Fuzzy Methods for Customer Relationship Management and Marketing: Applications and Classifications", In: Meier, A. (Ed)., IGI Global, USA, pp.1-15

Enache, I.C. 2015. Fuzzy Logic marketing models for sustainable development. Series V: Economic Sciences. Vol. 8. Iss. 57. No. 1-2015

Ferrara, E., Varol, O., Davis. C., Menczer, F., & Flammini, A. 2016. The rise of social bots. Communications of the ACM. Vol. 59. No. 7. pp. 96-104

Haque, A. & Rahman, T. 2014. Sentiment Analysis By Using Fuzzy Logic. International Journal of Computer Science, Engineering and Information Technology (IJCSEIT). Vol. 4. No. 1.

Internetlivestats, 2017.Twitter Usage Statistics, http://www.internetlivestats.com/twitter-statistics/, [accessed 13/06/2017].

Novac, P.K., Smailovic, J., Sluban. B., & Mozetic, I. 2015. Sentiment of Emojis, PLOS one 10 (12).

Kallas, P. 2017. Top 15 most popular social networking sites and apps [June 2017]. https://www.dreamgrow.com/top-15-most-popular-social-networking-sites/ . [accessed 15 June 2017]

McCarthy, J., (2011), Socialbots 2: Artifical Intelligence, social intelligence and Twitter, personal blog "Gumption", http://gumption.typepad.com/blog/2011/06/socialbots-2-artificial-intelligence-social-intelligence-and-twitter.html; [accessed 2012]

Paterva (2012), Maltego, http://www.paterva.com/web6/products/maltego.php [accessed: 2012]

Perticone, V., D'Aleo, F., Rizzo. G., & Elio Tabacchi, M. 2015. Towards a fuzzy-linguistic based social network sentiment-expression system, 16th World Congress of the International Fuzzy Systems Association, Proceedings of the 2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology Atlantis Press.

ter Veen, J., Sarkani, S., & Mazzuchi, T. 2011. Effects of Systems Engineering on Command and Control Systems Utilizing Innovations in Social Networking, Lockheed Martin Corporation, http://www.dtic.mil/ndia/2011system/12957_terVeenThursday.pdf [accessed: 2012]

Tuten, T., & Solomon, M., 2014. Social Media Marketing, 1st ed., Pearson, Harlow, Essex, UK.

Zadeh, L.A. 1965. Fuzzy Sets. Information and Control. Vol. 8. pp. 338-353