Mining Consumption Intent from Social Data: A Survey

Faizan Khan
Sikkim Manipal University
Sikkim Manipal Institute of Technology
Dept. of CSE

Samarjeet Borah
Sikkim Manipal University
Sikkim Manipal Institute of Technology
Dept. of CSE

Ashis Pradhan
Sikkim Manipal University
Sikkim Manipal Institute of Technology
Dept. of CSE

ABSTRACT
Social Media is a rich source of information about the desires and needs of users to buy a product or service. There lies a huge opportunity in mining the Intent of users which can be applicable to the field of marketing, ecommerce, recommender systems, etc. This survey focuses on analyzing the techniques that can be used to mine the Intent of users from social data. Notable works that have contributed towards determining the Intent of users from social data has been highlighted.

General Terms
Intent Mining, Social Data Mining, Natural Language Processing, Text Classification

Keywords
Consumption Intent, Intent Mining, Feature Extraction, Text Classification, Machine Learning.

1. INTRODUCTION
People express their needs and desires on social media websites on a daily basis. Sharing views on social media, searching the World Wide Web for information, browsing products/services in ecommerce websites, participating in QA forums have become integral activities in the lives of a modern human being. Huge amount of data is shared everyday by users, which can be analyzed to determine Intent of users. Consumption Intent (CI) may be defined as the desire or need to purchase a service or product in future. E.g. "I want to buy iPhone6" exhibits CI, whereas "My iPhone is broken, feeling sad" does not exhibit CI.

Determining CI of users presents a tremendous scope in targeted marketing, in improving recommender systems and suggesting devices and products tailored to users' need. It can also be used to test a policy to be implemented by the government, to detect criminal intent and improve community observation systems.

There are many sources of Intent of users, significant ones being:

- Mobile data consumers generate on smartphones and apps.
- Web Search Data.
- Sensor Data from remote sensors and meters in homes, vehicles, etc.
- Social data where customer intent is revealed through tweets, comments, likes, connections, plans and thoughts shared by users on social media websites.

Our focus is to study the detection and classification of social posts as CI or nonCI. A general approach begins with collection of social data, cleaning of data, extracting relevant features and training a classifier. Testing data is then applied to the classifier in order to classify the posts as CI or nonCI. This has been explained in detail in Section IV.

Section II defines Intent Mining, its relation with text classification, and similarities and differences with Sentiment Analysis. Section III explains about social data, its characteristics and its tremendous growth in recent years. Section IV describes a general approach, along with steps and tools to perform Intent Mining. Section V highlights some of the notable works in this field. Section VI outlines the areas in which Intent Mining may be applicable. Section VII contains conclusion, shortcomings of current work, direction for future work.

2. INTENT MINING
User generated Data from which information about the intent of user/users can be drawn is known as Intent Data. The process detecting and extracting the Intent of users is known as Intent Mining. A post with CI is defined as a post with an obvious indication to make an immediate or future purchase of some product [6]. Intent may be exhibited in an explicit or implicit manner in social posts [4]. An explicit intent post directly and clearly indicates CI. An implicit intent post indirectly indicates CI. "I will buy a Desktop if my Synopsis is accepted" is an explicit Intent post, whereas "I have been saving for the Diwali offer iPhone6 here I come?" is an implicit Intent post. Some examples can be found in Table 1. Mining implicit intent posts is more challenging and also necessary since its common knowledge that most intent posts in social media are implicit in nature.

Table 1. Posts labelled as CI and nonCI

Posts	Category
I am thinking of buying a DSLR. Any suggestions?	CI(explicit)
I would like to visit Sikkim in March. Anyone knows a good tour operator?	CI(explicit)
Excited about the launch of new Playstation.	CI(implicit)
I like Hero mountain bikes. Sadly, there is no dealer in my town.	CI(implicit)
The battery life of HTC One is disappointing.	nonCI
I visited Sikkim this Spring, lovely weather, breathtaking views and friendly people. I would recommend it to every traveller and nature lover.	nonCI

One of the early contributions in detecting online commercial intent [6] aimed at capturing Intent by analyzing search

queries and web pages. When the data source is social data, which means data is essentially present in the form of text, Intent Mining can be viewed as a specialized form of text classification [11] with some overheads. Overheads include dealing with informal language of social posts, mis-spellings, emoticons, hashtags and limited contextual information.

Sentiment Analysis[9][10] is as the process of computationally identifying and categorizing opinions in text to determine the attitude of the writer, which may positive, negative, or neutral towards a particular topic, product, etc. Intent Mining is not an alternative to Sentiment Analysis; it rather represents an orthogonal, intentional perspective. There are some notable differences between Intent Mining and Sentiment Analysis. Firstly, Sentiment Analysis is concerned with determining the opinion of people in a topic of interest. Intent Mining deals with detecting and extracting the intention of users to perform an activity (eg. purchase, criminal activity). Secondly, Sentiment Analysis is limited to assigning positive, negative and neutral polarity to documents. Intent Mining on the other hand is capable of categorizing documents/posts into numerous classes.

3. SOCIAL DATA

While there are different sources of information regarding intent of users, our focus in this paper is on social data due to several reasons. Firstly, it is common knowledge that users express their needs, desires and views on social media websites. Such posts represent users' real time and direct consumption needs. Secondly, there is tremendous amount of data being generated from social media websites, most of which is user-generated. This has also been termed as social data explosion. According to a study by wearesocial.net [12], there were 3.010 billion active internet users, with 2.078 billion active social media accounts, 3.649 billion unique mobile users, and 1.685 billion active mobile social accounts in 2014. 72% of the web traffic in 2014 was from mobile phones. Research conducted by the same internet firm suggests that the average social media user spends 2 hours and 25 minutes per day using social networks and microblogs, in 2015. This enormous growth in data in the past few years has also been termed as Social Data Explosion and has contributed immensely to Big Data. We live in an age where there is abundance of data, but scarcity of information. Processing these valuable data resources and obtaining useful information has become the need of the hour.

- However, working with this rich source of information has some disadvantages. Some of the issues and challenges in mining social data are as follows:
- Informal Language: Users often use internet slangs, short forms, emoticons in social media. Posts often contain misspellings and lack grammatical structure.
- Limited Post Length: There is a restriction on post length on most social media websites ((e.g. 140 characters limit on Twitter). This makes it difficult for algorithms to effectively extract information.
- Dissimilarity in actions and Intent: For instance, if a Facebook user likes the picture of a Porsche car, it does not suggest that he wants or is capable of purchasing one. Intent Mining systems may wrongly consider such information to be useful.
- Multiple Types of Data: Data in the form of text, audio, video, images, etc. makes the task of mining social data challenging.

In this paper we focus on tweets and posts from Question Answers forums which both fall under the social data. It is important to note some differences between the two. Firstly, for a particular product there are several intent posts in a discussion on a digital forum. The number of tweets in a discussion (in the form of replies) is relatively low for a particular product. Secondly, intent posts are longer than tweets and hence more information present in the former. We shall be using tweets, posts and data interchangeably.

4. METHODS AND TOOLS

Though there are several approaches towards mining intent from social data and each may differ from the other. However, a general methodology involves collecting social data; cleaning data to filter out irrelevant posts, extracting features, applying a machine learning algorithm in order to classify posts and categorizing the Intent posts. The problem may be approached as a binary classification problem in which posts are classified into CI and nonCI classes. Alternatively it may be approached as a multiclass classification in which posts are classifies into different categories.

Intent classification may be represented by a simple equation:

$g:T \rightarrow C$

where $C=\{c1,c2,\ldots,cn\}$ i.e. the set of intent classes

$T=\{t1,t2,\ldots,tn\}$ i.e. the set of tweets

In simple words given a set of tweets, and intent classes, the work is to assign individual tweets to these intent classes. General steps involved in mining intent from social data can be seen in Figure 1 and may be listed as follows:

4.1 Data Collection

It involves collecting data from social media websites. Some commonly used data sources are Twitter, Sina Weibo, Quora, Yahoo! Answers, Amazon and reviews. Twitter's Streaming API makes it convenient to collect data. Tweets can be collected by using NodeXl [13] a Excel plugin, Tweepy [14] a Python API, TwitteR [15] a package in the R library can be used to access data from twitter. Alternatively custom crawlers may be built to access data from social media websites.

4.2 Data Preparation

From the data collected in the above step, only those posts are kept which are relevant to the problem domain. This step is n known as filtering. Stemming is the process of limiting words to their base/root form. Eg running becomes run. Stopword removal involves removing common words like the, in, on, etc. Tokenizing involves reducing the words in posts to individual tokens. However, phrase and dependency based features are used it may be disadvantageous to use stemming and stop word removal. One of the most commonly used tools is the NLTK toolkit [16]. The training data now needs to be either manually or automatically labelled.

4.3 Feature Extraction

Most classifications algorithms do not work with text data in raw form. The posts need to be converted to a form that can be taken as input, by classification algorithm. This involves dimensionality reduction. Also it is important to select the most relevant features that would prove to be optimal for classification task. Feature selection/extraction is one of the most important steps and needs to be carefully designed. Some of the techniques are parts of speech tagging, n grams

and bag of words. Renowned toolkits include StanfordNLP [17] and OpenNLP [18].

4.4 Classification

This step classifies a given set of posts into predefined categories. It consists of a training step and a testing step. In the training step, a set of labelled posts (features) are fed to a machine learning algorithm. A classifier model which is capable of classifying unlabeled posts is developed. In the training step, a set of unlabeled posts is fed to the classifier model which labels the posts into respective categories. Open source machine learning tools like Weka[19] and RapidMiner[20] may be used.

4.5 Result Analysis

The performance and accuracy of the applied approach measured in a quantified manner. Commonly use metrics are precision, recall and f-measure. The result is visual manner in the form of graphs, pie charts, bar graphs, etc.

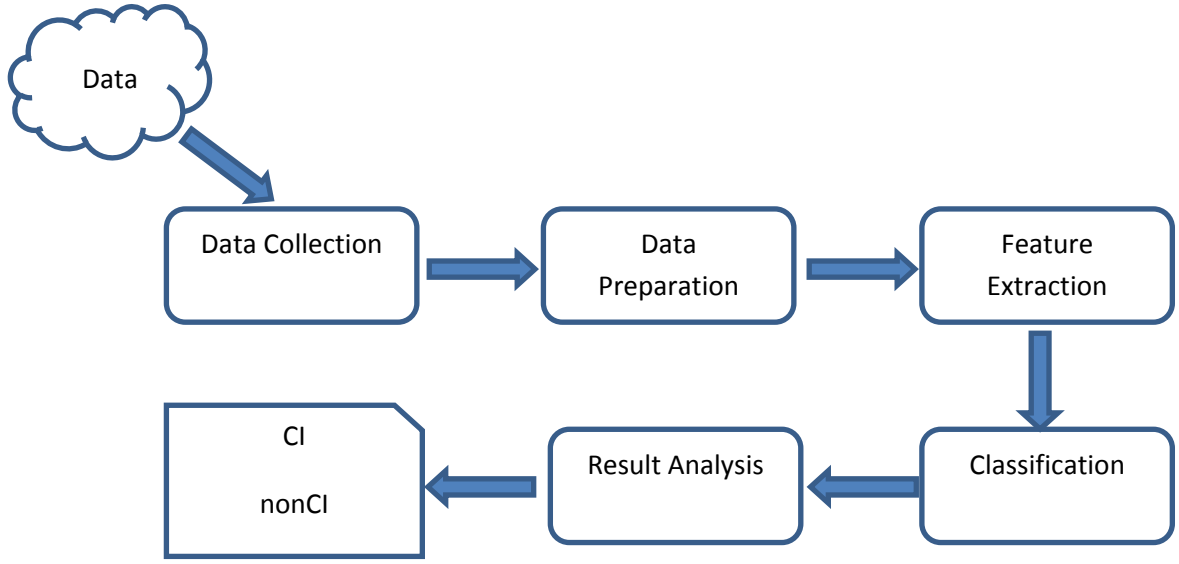

Fig 1: Block Diagram of Intent Detection

5. RELATED WORK

A Consumption Intention Mining Model (CIMM) has been proposed by **Xaio Ding** et al. [1]. This paper focuses on mining implicit intent from social data. Identification of consumption intent is based on convolution neural network (CNN). Only some domains have a large number of annotated instances, whereas some it is very difficult to find annotated instances for other domains. Hence, an adaptation layer is added in order to transfer mid-level sentence representations from one domain to another. 1,000 tweets from Sina Weibo that contain user consumption intention were manually labelled, of which 625 tweets represented implicit consumption intention and 375 tweets represented explicit consumption intention. The corpus was split into training, development and test dataset, 4/5 of which are used as the training data, 1/10 for development and 1/10 for test. Baby and Kids domain was used as source domain and movie domain was used as target domain. A Word representation Layer applied n-grams and hinge loss to produce a feature vector containing every word in the dictionary. Contextual features of words were captured based on sliding window in the Convolution layer. The most useful local contextual features were selected to form global contextual features using a max pooling layer. Sigmoid layer was used to extract non-linear vectors. For the target domain, a Domain Adaptation layer was added. The training algorithm calculates margin loss based on sentence label pairs. It repeats for several iterations over the training data. If the loss is zero, the algorithm continues to the next unlabeled sentence. Else, the parameters are updated using back-propagation. CIMM has been compared with two baseline methods, Words of Bag+ SVM and Word Embedding + SVM. It has been observed word embedding gives better results than bag of words approach. One shortcoming of this work is the lack of categorization of products into their respective domains.. The researchers have left recommending products to users for future work.

The problem of identifying and classifying tweets into intent categories was studied by Jinpeng, Wang et al. [2]. Tweets were classified into six categories namely Food & Drink, Travel, Career & Education, Goods & Services, Event & Activities and Trifle. The tweet dataset is prepared using a bootstrap method. Parts of Speech tagging has been performed on the collected tweets. A weakly supervised optimization model built on the intent graph with tweets and intent-keywords as nodes was proposed. An edge in the intent graph can be established to model the association between two tweets, two intent keywords, or a tweet and an intent-keyword. The researchers report this study as the first in inferring intent categories for tweets in the context of commerce marketing. However, this study does not deal with implicit intent tweets.

A weakly-supervised approach to detect user posts containing consumption intent in microblogs was proposed by Bo FU et al.[3]. The task of assigning classes to posts has been approached as a binary classification problem. A large number of posts containing a few designated hashtags are collected and are regarded as containing CI (positive instances). Posts without such hashtags are considered as negative instances. Textual content features, post specific features, and trigger words related features have been considered. Support Vector Machine has been used for classification. Alternatively, labels (CI, nonCI) assign to training data with the help of manual annotation. 10-fold cross validation was performed with the 7,902 human annotated data, 7,110 posts were used for training and 792 for testing in each run. Manually annotated methods (supervised approach)

were compared with the proposed method (weakly-supervised approach). It was found that the weakly supervised approach outperforms supervised approach. AUC improved from 0.9 to 0.95. One major drawback is that only hashtag information is used for filtering of tweets. In this approach, a large number of tweets containing Consumption Intent but without hashtags may be skipped during filtering.

A novel unified graph based ranking algorithm which jointly models relevance and associativity for identifying trend-driven ecommerce-products was proposed by Jinpeng Wang et al. [7]. The major novelty of this work is in automatically learning commercial intents from microblogs. Hot trends in microblogs are correlated to the sale of products in ecommerce platforms. Sina Weibo is the microblogging platform and Taobao is the e-commerce platform that has been considered since these were the biggest microblogging platform and the largest C2C Company in China respectively at the time. Identification of 'trends' is not a part of the work and trends are derived 'trending topics' of microblogging platform. For each trend, keywords are used to retrieve related products from the e-commerce website. A pattern-based product keyword extraction method has been used, and it tends to incorporate some irrelevant words. A novel algorithm JMRA(Jointly Modeling Relevance and Associativity) is applied.

A novel framework to automatically mine intention-related products was proposed by Junwen Duan et al.[8]. The domain considered is Baby and Child Care domain.. Question-answer pairs from the three websites (Taobao Wenda, BabyTree and Sina Baby & Child Care) were crawled and 700 thousand pairs were extracted. Intention keywords and intention-related products were manually annotated. A pattern-based method with dependency parser was used to automatically extract the candidate products from the answers. In order to identify the intention-related products, a novel collocation extraction model was proposed. It was assumed that the intention keyword set was available. Detecting Intent was not a part of the work. The approach has been proved to perform better than two baseline methods Co-occurrence and Jaccard Coefficient. It has been mentioned by the researchers that adding a filter process after candidate product extraction, the precision of intention-related product identification would be greatly improved.

An automatic method for detection of commercial intent has been presented by Bernd Hollerit et al.[4]. This work addresses the detection of commercial intent on Twitter. Twapperkeeper was used to collect tweets containing specific commercial keywords. 16 keywords were selected and 100 tweets per keyword were annotated. It was also considered (i) whether the intent was implicit or explicit and (ii) whether the commercial activity was buying or selling intention. Filtering, stemming and stop-word removal was applied on the tweets. The attributes used were word features and part-of-speech n-grams (2 to 5). For the classification task, WEKA was used. Recall score of 77.4% was achieved using a Bayes Complement Naïve Bayes classifier. A precision score of 57.1% were achieved by using a linear logistic regression classifier. It has been mentioned that it is important to better understand the language used on twitter. It was also observed that POS tagging led to many false negatives.

The task of detecting Purchase Intent from social posts and classifying them has been studied in this work by Gupta et al. [5]. The corpus is collected from Quora and Yahoo! Answers using a custom web crawler. Purchase Intent has been defined as "A text expression containing one or more Consumption Indicative words(e.g. cell phone, lunch) which provide the object of intent along with one or more Action Indicative words(e.g. buy, eat) which further indicate an intent of consumption." The posts were manually labelled as PI or nonPI. Features are extracted two different levels of text granularity i.e. word and phrase based features and grammatical dependency based features. Purchase Action (PA) words are extracted based on the verbs in a sentence, and Purchase Object Categories (PO) were extracted based on nouns in a sentence. It has been assumed that in a PI post, the consumable object is usually the directly dependent object of the purchase action verb. A Support Vector Machine has been used for classification. Experiments were performed using three different set of features viz. i) Delta TFDIF, ii)PA and PO and iii)PA and PO along with dependency based features. PA and PO along with dependency based features was found to exhibit the best AUC.

Table 2. Summary of Related Work

Study	Overview	Data Source	Features Extraction	Classification Technique	Results
1) Mining User Consumption Intention from Social Media Using Domain Adaptive Convolutional Neural Network. 2015	A model based on, convolution neural network (CNN), for identifying consumption intent in tweets has been proposed. CNN requires a large number of annotated instances for learning, which is present in only a handful of domains. Transferring mid-level sentence representation for one domain to another is an important part of this work.	Tweets from Saino Web	n-grams, Hinge Loss and word embedding	CIMM based on CNN	Works significantly better than SVM.

2) Mining User Intents in Twitter: A Semi-Supervised Approach to Inferring Intent Categories for Tweets. 2015	A weakly supervised optimization model built on the intent graph with tweets and intent-keywords as nodes. Six categories namely Food & Drink, Travel, Career & Education, Goods & Services, Event & Activities and Trifle.	Twitter data (Kwak et al. 2010)	Parts of Speech Tagging	Semi supervised Graph	Food Travel Self Goods Event Trifle Non-intent	54.63% 58.64% 45.73% 43.25% 27.13% 20.04% 35.56%
3) Towards Linking Buyers and Sellers. 2013	This work focuses on detecting Commercial Intent in tweets using an automatic method. The emphasis is on buying as well as selling Intent.	Twitter Data	Words, parts-of-speech, n-grams	Bayes Compliment Naïve Bayes Linear logistic regression classifier	Recall-77.4% Precision- 57.1%	
4) Mining New Business Opportunities: Identifying Trend related Products by Leveraging Commercial Intents from Microblogs. 2013	A novel unified graph based ranking algorithm which jointly models relevance and associativity has been proposed in order to automatically learn commercial intent from microblogs. A pattern-based product keyword extraction method has been used.	Sina Weibo	Pattern-based product keyword extraction	JMRA(Jointly Modeling Relevance and Associativity)	Precision- 0.734	
5)Weakly-supervised consumption intent detection in microblogs. 2013	A weakly-supervised approach to detect user posts containing consumption intent in microblogs in which hashtags are used for automatic collection of data. A comparision of supervised(human annotated) method with weakly supervised(proposed) method is carried out.	Sina Weibo	Bag of words, post specific features, trigger word	SVM	Supervised AUC Weakly Supervised AUC	0.9 0.95
6) Mining Intention-Related Products on Online Q&A Community." *Social Media Processing*. Springer Berlin Heidelberg, 2014	A novel framework to automatically mine intention-related products was proposed. A corpus of Question Answers pairs in Baby and Child Care domain was mined.	QA pairs from Taobao Wenda, BabyTree and Sina Baby & Child Care	Pattern based method, Dependency parser	MWA- A novel collocation extraction model	Precision MWA Jaccard Co-occurence	 0.80 0.60 0.60
7)Identifying Purchase Intent from Social Posts	The problem of detecting Purchase Intent from social posts has been studied. The characteristics of social posts have been considered, and by extracting features at two different levels of granularity, a classification model has been built.	Quora and Yahoo! Answers	Named Entity Recognition, PCFG parser and dependency parser	SVM	AUC Delta TFIDF Purchase Action(PA) and Purchase Object(PO) PA, PO and dependency based	 0.79 0.86 0.93

6. APPLICATIONS

There are many fields in which Intent Mining may find its use. It may be used in the field of Targeted Marketing and Advertisements. Intent Mining may be used in detection of fraud, plagiarism and criminal intent. Intent Mining may be used in analysis of the response of a new policy introduced by the government. Predicting the inflow of tourists and timely preparation in a tourist destination could be a potential use of Intent mining. Before the launch of a new product/technology, the adoption Intent of users can be mined. By finding the Consumption Intent of a large number of users in a city, warehouses can be prepared and delivery facilities can be arranged in a timely manner. In predicting the demand of a product/service before starting a business in a particular area, based on web search data of potential users.

7. CONCLUSION AND FUTURE WORK

It has been observed that there is a huge scope in mining Consumption Intent from social data. The used of word/phrase level features along with dependency based features yields good results. Naïve Bayes and Support Vector Machine was used either for classification or as a baseline method in most works. Only one work so far[1] has successfully mined implicit intent from social data. The use of neural networks has exhibited better results. Also, it has been observed it would be advisable to focus on tweets/posts that represent future state of affairs, for determining Intent. E.g. "I'm going to buy a new laptop next month".

Based on the literature survey carried out, following issues can be brought out, which are not being addressed partially/fully by the researchers. Firstly, most research works focus on posts with explicit user intent, and not implicit user intent. Secondly, there is a lack of complete solution right form Consumption Intent Detection to Categorization of Intent. Thirdly, a combined framework which takes into account web search, social data and enterprise data (purchase history, past trips/travels, etc.) is yet to be developed. Lastly, detecting the Consumption Intent of Users of an entire geographical region (city) based on social data is an emerging area of research.

With respect to the research gaps it has been observed that there have been very few efforts on detecting commercial intent from social posts. There is a huge scope in identifying implicit Intent with only one notable work [1] so far. User(s) express Intent on social media and microblogs mostly in an implicit manner. The application of CNN has shown good result in determine Consumption Intent of users [1]. There lies a huge opportunity towards detecting Consumption Intent of users, recommending products and services tailored to users' needs from vast amounts of social data generated by the users. This problem may be addressed by designing an approach that uses analyses both explicit and implicit intent data in order to identify Consumption Intent of users and categorize the detected Intent tweets into their respective domains. Also, the application of Deep Neural Networks to the problem of Intent Mining is in early stages and upcoming area of research.

8. REFERENCES

[1] Ding, Xiao, et al. "Mining User Consumption Intention from Social Media Using Domain Adaptive Convolutional Neural Network." Twenty-Ninth AAAI Conference on Artificial Intelligence. 2015.

[2] Wang, Jinpeng, et al. "Mining User Intents in Twitter: A Semi-Supervised Approach to Inferring Intent Categories for Tweets." Twenty-Ninth AAAI Conference on Artificial Intelligence. 2015.

[3] Fu, B., and T. Liu. "Weakly-supervised consumption intent detection in microblogs." Journal of Computational Information Systems 6.9 (2013): 2423-2431.

[4] Hollerit, Bernd, Mark Kröll, and Markus Strohmaier. "Towards linking buyers and sellers: detecting commercial intent on twitter." Proceedings of the 22nd international conference on World Wide Web companion. International World Wide Web Conferences Steering Committee, 2013.

[5] Gupta, Vineet, et al. "Identifying Purchase Intent from Social Posts."ICWSM. 2014.

[6] Dai, Honghua Kathy, et al. "Detecting online commercial intention (OCI)."Proceedings of the 15th international conference on World Wide Web. ACM, 2006.

[7] Wang, Jinpeng, et al. "Mining New Business Opportunities: Identifying Trend related Products by Leveraging Commercial Intents from Microblogs." EMNLP. 2013.

[8] Duan, Junwen, Xiao Ding, and Ting Liu. "Mining Intention-Related Products on Online Q&A Community." Social Media Processing. Springer Berlin Heidelberg, 2014. 13-24.

[9] Kouloumpis, Efthymios, Theresa Wilson, and Johanna D. Moore. "Twitter sentiment analysis: The good the bad and the omg!." Icwsm 11 (2011): 538-541.

[10] Pak, Alexander, and Patrick Paroubek. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." LREc. Vol. 10. 2010.

[11] A. Khan, B. Baharudin, L. H. Lee, K. Khan, "A Review of Machine Learning Algorithms for TextDocuments Classification", Journal of Advances Information Technology, vol. 1, 2010.

[12] Social Media Statistics, URL: http://wearesocial.net/, access date: 15/01/2016.

[13] Smith, Marc, et al. "NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007/2010." Social Media Research Foundation (2010).

[14] Tweepy, URL: http://www.tweepy.org/, access date: 15/02/2016.

[15] R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: http://www.R-project.org, access date: 15/02/2016.

[16] Bird, Steven. "NLTK: the natural language toolkit." Proceedings of the COLING/ACL on Interactive presentation sessions. Association for Computational Linguistics, 2006.

[17] Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60.

[18] Baldridge, Jason. "The opennlp project." URL: http://opennlp. apache. org/index. html,(accessed 2 February 2016) (2005).

[19] Holmes, Geoffrey, Andrew Donkin, and Ian H. Witten. "Weka: A machine learning workbench." Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on. IEEE, 1994.

[20] RapidMiner, URL: https://rapidminer.com,access date: 18/02/2016.