The Road to Emotion Mining in Social Network

Hany Mohamed
Department of Computer Science
Helwan Unviersity

Ayman Ezzat
Department of Computer Science
Helwan Unviersity

Mostafa Sami
Department of Computer Science
Helwan Unviersity

ABSTRACT
Affective computing is a highly computer research trend in last years, which related to human emotions and how computer interacts. While emotion is a fundamental in human experience, it becomes an ideal resource for servicing business or decision making. In ancient times, natural interfaces are likely to be used to provide ubiquitous computing. Although great achievement done, there still exist three challenges which are Cheap, Low power and software system. With the explosive growth of social media, people are using it to express their emotion or opinion. Currently, there are large amount of user generated data in different format (i.e. Blog, Tweets, Posts, discussion forums) that represent individual expression feelings towards daily life activates whether it is product, topic, event, news, or personal life. As a result, a lot of researchers are done for detecting what humans feel now in social network; they fall under the scope of topic called emotion mining, opinion mining, or sentiment analysis. In this paper, we will survey the development done for emotion mining with a comparative study for different approaches. In addition, an investigation for technology used in this area and how it is applied, will be presented.

Keywords
Natural Language processing, Machine Learning, Emotion mining, sentiment analysis, social network.

1. INTRODUCTION
Affective computing gets its name from the field of Psychology, in which "affect" is, basically, a synonym for "emotion". It targets offering a lot of benefits or services based on individual emotion. For the past decade, application driven research in ubiquitous computing has pushed three interaction themes: natural interfaces, context-aware applications, and automated capture and access [1]. Natural interfaces are likely to be used instead of GUI interactions. Speech, handwriting, facial expressions, keystroke, and gesture recognition are samples. Five Ws is a sample of context of ubiquitous computing, which is shortcut to Who, What, Where, When, and Why. Who focus on the identity of user, what relates to the action itself, where the physical place the user exists, when represents the time of action, and why represents an effective reason for the human doing. Three requirements are declared for ubiquitous computers systems: Cheap, Low power, and software systems. Ubiquitous does not offer anything new, but it makes everything easier and exploits the high power of computing services.

Several issues have to be considered like Display, Storage, and Data transfer issues. A computing device has the ability to detect human emotion and respond or extract info to serve business. Affective computing could offer benefits in an almost limitless range of applications. For example, in e-learning situations, the computer could detect from available cues when the user was having difficulty and offer expanded explanations or additional information. Other applications include e-therapy: psychological health services, such as counseling, delivered online. Internet-based therapy, although increasingly common, does not give a therapist as many cues to the client's emotional state as are available in a real-world session [3]. Despite that, a lot of limitation exists in above approaches: High computing devices required, existence for individuals within the context, and minimal amount of information extracted compared to cost. \\

The World Wide Web has brought change to a point where it would be difficult to imagine a world not connected through online networks. In 2010, over 58\% of European's population (813 million) were internet users (Internet Worlds Stats, 2010), and online social network sites has emerged as one of the latest innovative applications from the web [4]. The idea starts with connecting people and evolved to content or opinion sharing. Social network categorized based on users' communication, whether is bi-directional like Facebook or Unidirectional like twitter or Google+. Another classification is based on their function, whether it is general like Facebook, Twitter, Business use like LinkedIn, video sharing like YouTube, photo sharing like Flicker or Instagram. Moreover, they can be categorized based on domain whether it is global or specific (like organization or specific location).

Currently, people can declare their opinion or express what they are feeling via multi social channels. Social media has different types and properties of communication. Blogs, review sites, and micro blogging are common examples for such channel types. Blogs can be considered as an article written by people for specific topic and can be edited and published freely by common systems. Review sites are like questionnaire for a given topic to collect a feedback from individuals. Micro blogging is the most interesting channels for individuals to share their emotion among their friends, Twitter and Facebook is an example, although it is not easy task for researchers to work on such channel due to limitation and challenges that will be stated later.

Online Social Networking(OSN) has been designed for two primary purposes; (i) to enable the sharing and interaction of data, (ii) to support the social activities of users, Meanwhile, Trends in social networking has be seen to have been driven by the advancements in technology, with many of today's popular sites developed after the emergence of web2.0, it provided users with generation of sites which facilitated interactive information sharing, enabling users to become active authors and contributors of content[4].

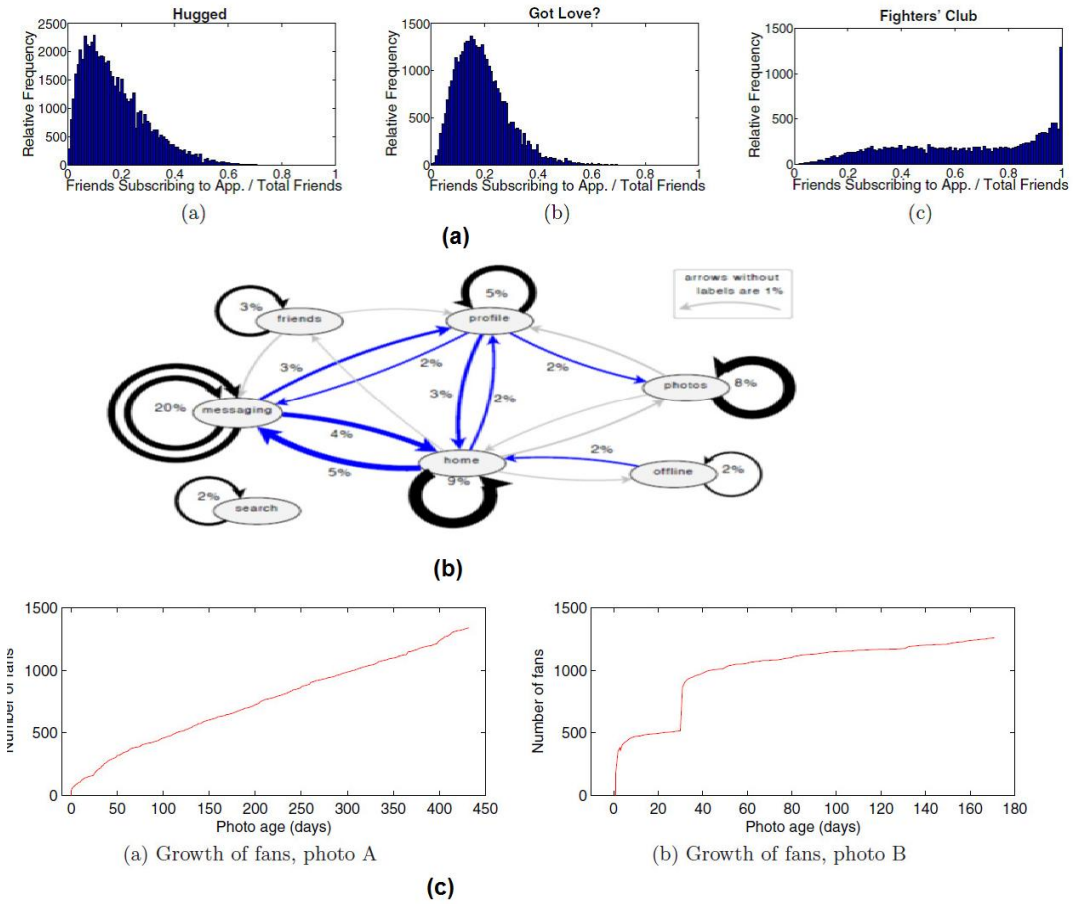

Figure 1 (a) static property (b) dynamic property (c) social property of OSN

With the appearance of "Facebook Developer Platform", the number of applications uses Social Network horribly rises. It can be reasonably stated that the data on OSNs, if properly collected and analyzed, could contribute to the development of various fields of behavioral research, marketing, political science, education, etc. On the top of this, tremendous amounts of research are required merely to maintain the efficient growth and development of these instruments and their successors [6]. Analyzing online social network can be applied in different business areas due to vast amount of data published there such marketing, political, science, and education. Social Network Analysis (SNA) leads us to understand different properties of OSN that are categorized in three types: Static, Dynamic, and Social. Static, as from its name, is changed slowly over the life of the network; Degree distribution is an example which characterizes the distribution of user connection to others. Dynamic are changed rapidly, for example the interactions of users and propagation of a content. Social properties discuss user habits and demographics analysis. Fig 1 shows an example for above three properties.

OSN can be viewed as structure of individuals, groups or organizations and their connection. Groups or organization represented as Nodes, and connections represented as ties. Social Network's Models can be classified into three categories: network evolution Model (NEM), Network attribute models (NAM), Exponential Random Graph Model (ERGM)[4].

The location of a node defines its centrality which is a measure of importance in the network. To evaluate network, there are parameters detailed here in [7]:

1. Degree: number of direct connections that
2. Betweenness: the number of paths that connect.
3. Pairs of nodes that pass through a given node.
4. Prestige: a number of links to highly central nodes
5. Closeness: the number of other nodes that are linked to a given node \newline

In [5], researchers address several challenges that are signification in SNA area:

Social graph analysis, Social media search and management, exploiting social graphs, Identity algorithms, Mobile social networks given node, Social ranking and opinion sites, Business and social networking, Architectures for open and federated social network platforms , **Emotion mining** which is our interest here.

Here we are going to state number of definitions:

- **Defination 1(Emotion):** Emotion is a complex state of the mind influenced by external events, physiological changes, or relationships with others [8]. Paul Ekman, the author of popular book "Emotion Revealed" classifies human emotion in six categories, which are Love, Joy, Anger, Sadness, Fear and Surprise.

- **Definition 2 (Opinion from mathematical perceptive)**: In [9], researchers declare mathematical definition for opinion, they define an opinion as a quintuple (ei, aij, sijkl, hk, tl), where ei is the name of an entity, aij is an aspect of ei, sijkl is the sentiment on aspect aij of entity ei, hk is the opinion holder and tl is the time, when the opinion is expressed. An entity is the target object of an opinion; it is a product, service, topic, person, or event. The aspects represent parts or attributes of an entity (part-of-relation). The sentiment is positive, negative or neutral or can be expressed with intensity levels. The indices i, j, k, l indicate that the items in the definition must correspond to one another. Emotions may be expressed by a person's speech, face expression and written text known as speech, facial and text based emotion respectively. Sufficient amount of work has been done regarding to speech and facial emotion recognition but text based emotion recognition system still needs attraction of researchers. In computational linguistics, the detection of human emotions in text is becoming increasingly important from an applicative point of view [10].
- **Definition 3 (Sentiment analysis):** is the process of detecting meaning / emotion/ opinion of user's statement towards subject whether it is an event, topic, product, etc. Sentiment analysis has another name called emotion mining or opinion mining.

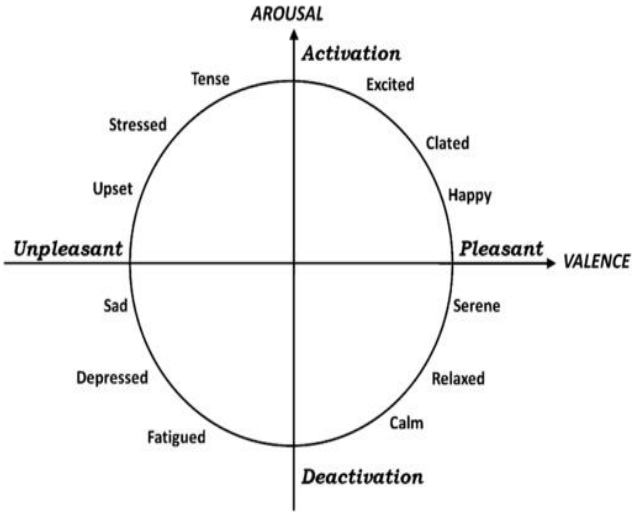

Figure 2 Emotion category

Table 1 show some examples of tweets, and its emotion based on manual annotation

Table 1 a sample of subjective tweets

Tweet	Subject	Emotion
Egypt, is a safe place to live in	Egypt	Positive
I face hard times in, mathematical exam	Exam	Negative
I just reach my home	Home	Natural

Emotion mining categorizes emotion whether subjective or objective, subjective means users feeling towards specific thing while objective is a fact. Another category is based on valance indicating whether it is positive or negative like above example. Tolerance based is the final category tells intensity of emotion. Figure 2 show sample of emotion category.

Regardless research challenges exists in Natural Language Processing scientific area, there are specific for OSN, the list below are some of them:

- Text Informality: Traditionally, people are used to write in social media with ignoring to language structure or grammar.
- Language Acronyms: social media has a lot of abbreviations that are publicly and popular to be used by social users. Unfortunately, they are increased, changed dramatically which make their dedication by automated system hard.
- Languages Mixture: Franco-Arab is an example that Arab people is used to express their opinion or message their friends.
- Emotion icons: Most of social network offers a symbol that expresses specific feelings, ":)" or ":(" is a sample of such icons.
- Relevance: Researchers still face the problem of facing non related topic like advertisement \newline

The rest of the paper will be divided as follows: Sections 2 presents different approaches for emotion mining in online social networks. Section 3 provides the technology used and general steps for such systems. Finally, conclusion presented in section 4.

2. BACKGROUND AND RELATED WORK

Problem of emotion recognition from text can be formulated as follows: Let E be the set of all emotions, A be the set of all authors, and let T be the set of all possible representations of emotion-expressing texts. Let r be a function to reflect emotion e of author a from text t, i.e., r: A x T $\rightarrow$E, then the function r would be the answer to the problem [10].

There a lot of researches proposed and applied in the area of emotion mining. We categorize them in both two main branches: Lexicon based Approaches and ML based Approaches.

2. 1 Lexicon Based Approaches

2.1.1 Key word Spotting

It is the simplest technique in emotion mining area; it depends on extracting emotion words and classifies them into one of Ekman categories. Figure 3 shows the steps of algorithm. The source text will be input to the algorithm in order to identity its emotion. Algorithm starts by tokenizing sentence into words. Recognizing which words are emotionally annotated. Then, fetched words are intensity analyzed, and then negation check is analyzed, and tells us the final classification of emotions.

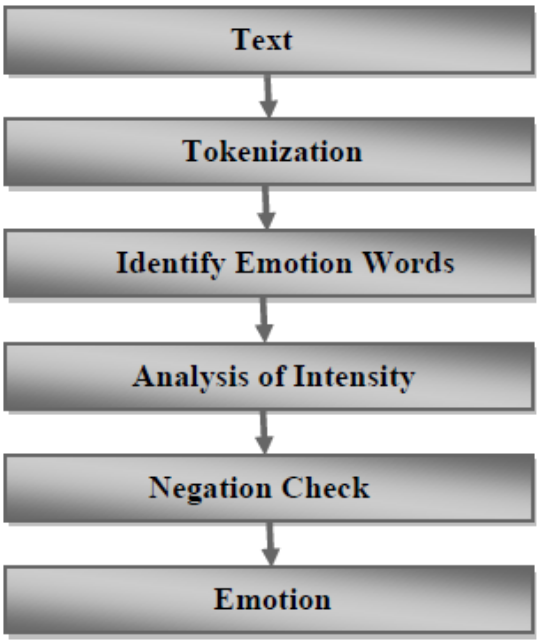

Figure 3 Keyword spotting technique

2.1.2 Lexical Affinity

It is an extension to spotting, by assigning probability for each word term. It assigns a probabilistic, affinity" for a particular emotion to arbitrary words apart from picking up emotional keywords [10].

It is considered to be more accurate than keyword spotting as it went to calculate probability for each token. Although that, it suffer from the same disadvantage like keyword spotting.

2.2 ML Based Approaches

Paper [11] presents a comparative study between three common machine learning algorithms for sentiment analysis, which are Naïve Bays, Support Vector Machine, and Maximum Entropy. The process of detecting emotion using ML is done by applying two steps: training for a given data and features. The output from training is model or classifier that is used for classification and run it on new data to extract final emotion. Effectiveness of any of above methods is determined by precision or recall values [12]. Where:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = TP\ /(TP + FN)$$

- True positives (TP) - number of reviews correctly labeled as belonging to particular class (positive/negative).
- False positives (FP) - number of reviews incorrectly labeled as belonging to particular class.
- False negatives (FN) - number of reviews were not labeled as belonging to the particular class but should be labeled.

Researchers here [11], prove that for unigram model, Naïve Bayes is the most accurate, while for Trigram model, SVM perform the best

3. IMPLEMENTATION GUIDE

Figure 4 shows general process cycle for emotion mining system and below are details:

- **Step 1 (Data Collection):** System should start with exploring World Wide Web in addition to various social media, for selective topics or predefined in system DB and finally output list of URLs. Web content extraction is not easy task, it starts to scan provided these URLs and extract all text content in automatic way. It becomes high required to detect location of users, the owners of social content. Although privacy issue is a critical challenge, but still there is a lot of techniques can recover that, Technology used are the following:
    - **Web crawling:** it is used to browse the World Wide Web in a systematic way targeting web indexing. This type of tool is considered as input to web search engine to update their data or to index other website data. Most of these tools have another function called Web Scarping. Simply, it targets extracting information from websites simulating human exploration of Internet. Our idea is to build custom crawling tool to index world emotion using different type of channels whether Social Network (like popular ones: Facebook, Twitter, LinkedIn, Google+) or Social Blogs (Like Google Blogger, Word Press, or others), or through different web sites. Already existing tool can help; our scope is open source to build customization we need, while Google Web master tools are the most common. Scalability, Fast, and maintainability are three factors we are looking for to evaluate different tools. We are not going here to do a survey for existing tools, but we can list common ones like Nutch, Hetrix, Scrapy, Grub, Open Search Server, Crawler4j, mnoGoSearch, ASPseek, OpenSE,and TrendSniffer. Our recommendation is using Scrapy(Python Based), An open source and collaborative framework for extracting the data you need from websites in a fast, simple, and extensible way.

    - **Social Network API platforms:** The high revolution of social network is due to first appearance of social network development platform created by Facebook. Most of social network take the same vision and build their own Framework API (Application Programming Interface) to allow developers build their application and integrate with social media content. OAuth and openID are most common protocols used to allow access to social content for third party applications. OAuth provides a process for end-users to authorize third party access to their server resources without sharing their credentials (typically, a user name and password

pair), using user-agent re-directions [5]. While OpenID is represented as single sign on technique to allow users login to different social media with same user name/ password, Yahoo!, Facebook, MySpace already use it.

- **Step 2 (Data Processing):**
  Topic identification and categorization is used to detect the topic of fetched content, it may occur that content is not related to topic; Spam filtering is applied to drop such contents. Now, the role of sentiment analysis is coming, extract the opinion of author towards the topic whether it is objective or subjective.
  Here a list of content, emotion, author, and location that already prepared from previous step. User can view topics and overall opinion towards them, categorized by location, Technology used are the following:

  1. Word Net-AffectN:
  As described in emotion mining approaches section, Lexicon based approaches is still the preferred way to detect emotion in micro blogging social media like Facebook, Twitter, and so on. Wordnet – Affect is an extension for Wordnet, in which a number of wordnet synsets is assigned one or more affective labels(a-labels). Emotion is classified into one of four categories Positive, negative, ambiguous, and neutral. Two types of words are maintained here, direct affective words and the indirect affective words. For indirect affective words, a selection method named Affective-Weight, based on a semantic similarity mechanism to extract emotion from large corpus of texts (100 million of words).

  2. ML Tools:
  In recent researchers for opinion mining, Machine learning (ML) is tremendously used for the sentiment recognition to reach high level of accuracy especially for non-labeled text. Our proposed framework will utilize ML for multi-purpose like predictive analysis, spam filtering, and for text mining. Scikit-learn (python based), Shogun, Accord.Net, Appache Mahout, MLlib, H2O, Cloudera Oryx, GoLearn, Weka, CUDA-Convnet, and ConvNetJS are most common open source tools to implement different machine learning techniques. Accord.Net is our recommended one because of large number of functionalities presented beside ML techniques.

  3. Hadoop
  According to vast amount of data generated continuously by social media or through World Wide Web, thinking about data processing, storage, and computation becomes a critical matter. Apaches comes with open source framework solution for reliable, scalable, distributed computing. Hadoop is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of computers, each of which may be prone to failures [26].

- **Step 3 (Final output):**

Here a list of content, emotion, author, and location are already prepared from previous step. User can view topics and overall opinion towards them, categorized by location or by emotion, Technology used are the following:

- JSON (JavaScript object notation) is an alternative to XML for the purpose of storing and exchanging data in a lightweight data-interchange format. Most of social network API use JSON for representing its content, as it is language independent and easy to understand". JSON can be implemented with different technology like .Net, Java, and so on.

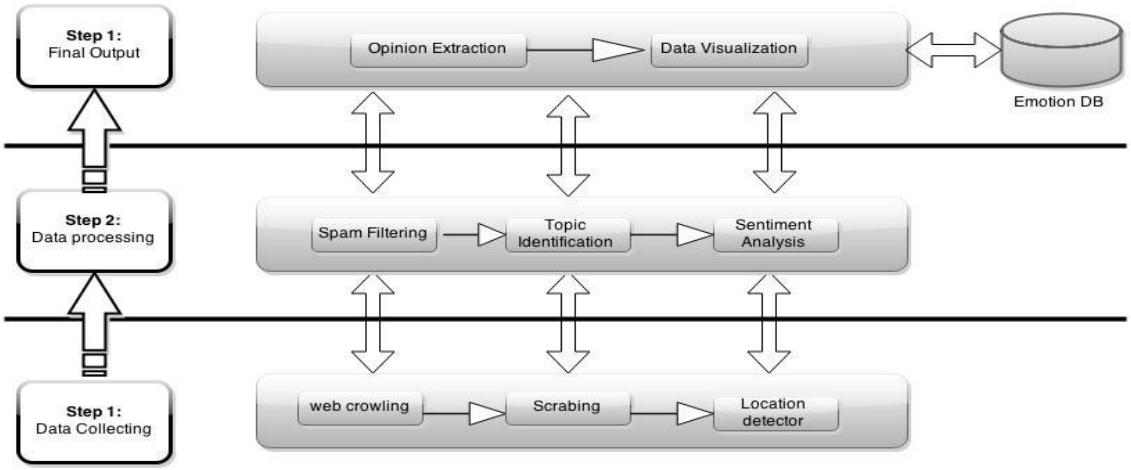

Figure 4: General process cycle for emotion mining system

4. CONCLUSION AND FUTURE WORK

During last decade, the interest for social networks is tremendously increased to the degree that people consider them as a mandatory for life. Vast amount of data represented as people daily activity, their feedback about life events or product or anything else, or whatever they feel now!. Information can be extracted from these data for servicing different area of business domains. Emotion mining, Opinion mining, or sentiment analysis is a technology aims to collect user's reactions towards trending topics or any life event.

In this paper, we present an exploration for online social network, and the evolution happen in emotion mining related to affective computing. We summarize most common techniques in this area, dividing into two categories which are lexicon based approaches or ML based approaches. A general process cycle for such systems is explained in details helping market to build such system, declaring various technologies used.

Although great achievement done as stated above, critical challenges still acquire attention for large number of researches. Big Data, Real Time, Accuracy, and visualization are list of challenges faced in area of emotion mining in online social network

5. REFERENCES

[1] GREGORY D. ABOWD and ELIZABETH D. MYNATT, "Charting Past, Present, and Future Research in Ubiquitous Computing", Journal of ACM transactions on Computer Human Interaction, Vol7 issue 1, March 2000, Pages 29-58.

[2] Mark WeiserPalo Alto Research Center, Xerox, CA, "The Computer for 21st century", Newsletter of ACM SIGMOBILE Mobile Computing and Communications Review, Volume 3 Issue 3, July 1999, Pages 3-11.

[3] Margaret Rouse, "Definition of effective computing", "http://whatis.techtarget.com/definition /affective-computing, September 2005.

[4] Darren Quinn, Liming Chen, Maurice Mulvenna, "Social Network Analysis: A survey", International Journal of Ambient Computing and Intelligence, Volume 4 Issue 3 , September 2012, Pages 46-58.

[5] "Social Networks Overview: Current Trends and Research Challenges", Coordinated by the "nextMEDIA" CSA, Supported by the Future Media Networks cluster, November 2010.

[6] Edward M. Lazzarin, Raj Jain, "An overview of the analysis of online social networks", 2011.

[7] Sudhir Saxena, K. Santhanam, and Aparna Basu , "Application of Social Network Analysis (SNA) to Terrorist Networks in Jammu Kashmir", Journal of STRATEGIC ANALYSIS ,Volume 28 Issue 1, January 2004.

[8] Hosub Lee, Young Sang Choi, Sunjae Lee, and I. P. Park, "Towards Unobtrustive Emotion Recognition", The 9th Annual IEEE Consumer Communications and Networking Conference - Special Session Affective Computing for Future Consumer Electronics, 2012.

[9] Gerald Petz , Michał Karpowicz , Harald Fürschuß, Andreas Auinger,Václav Stříteský, and Andreas Holzinger, "Opinion Mining on the Web 2.0", Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data, Springer, Lecture Notes in Computer Science Volume 7947, Pages 35-46, 2013.

[10] Shiv Naresh Shivhare, Saritha Khethawat, "Emotion Detection from text", 2010.

[11] Mohamed Yassine, Hazem Hajj , "A Framework for Emotion Mining from Text in Online Social Networks", IEEE International Conference on Data Mining Workshops, 2010.

[12] Anyim, Julianne A, "A comparative evaluation of sentiment analysis techniques on Facebook data using three machine learning algorithms".

[13] Vijay B. Raut , D.D. Londhe, "Survey on Opinion Mining and Summarization of User Reviews on Web (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2), Pages 1026-1030, 2014.

[14] Johan Bollen, Huina Mao, Alberto Pepe, "Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena", Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011.

[15] Valentina Sintsova, Claudiu Musat, Pearl Pu, Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 12–20, Atlanta, Georgia, 14 June 2013.

[16] Hosub Lee ; Intell. Comput. Lab., Samsung Electron. Co., Ltd., Yongin, South Korea ; Young Sang Choi ; Sunjae Lee ; Park, I.P., "Towards unobtrusive emotion recognition for affective social communication", The 9th Annual IEEE Consumer Communications and Networking Conference – Special Session Affective Computing For Future Consumer Electronics.

[17] Jahidul Arafat, Mohammad Ahsan Habib, and Rajib Hossain3, "Analyzing Public Emotion and Predicting Stock Market Using Social Media", American Journal of Engineering Research (AJER), Volume-02, Issue-09, pp-265-275, 2013.

[18] Daniel Preotiuc-Pietro, Sina Samangooei, Trevor Cohn, Nicholas Gibbins, Mahesan Niranjan. "Trendminer: An Architecture for Real Time Analysis of Social Media Text", Workshop on Real-Time Analysis and Mining of Social Streams (RAMSS), International AAAI Conference on Weblogs and Social Media (ICWSM), June 2012, Dublin.

[19] Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka, "Compositionality Principle in Recognition of Fine-Grained Emotions from Text", Proceedings of the Third International ICWSM Conference, 2009.

[20] M.Vasuki1, J.Arthi2, and K.Kayalvizhi, "Decision Making Using Sentiment Analysis from Twitter", International Journal of Innovative Research in

Computer and Communication Engineering, Vol. 2, Issue 12, December 2014.

[21] De Choudhury, M., and Counts, S. (2012). The Nature of Emotional Expression in Social Media: Measurement, Inference, and Utility. In 2012 Human Computer Interaction Consortium (HCIC) Workshop (Pacific Grove, California, Jun 24-29, 2012). HCIC 2012.

[22] Munmun De Choudhury, Scott Counts, and Michael Gamon, "Not All Moods are Created Equal! Exploring Human Emotional States in Social Media", Association for the Advancement of Artificial Intelligence, In proceedings of Association for the Advancement of Artificial Intelligence, June 2012.

[23] Chowdhury Mofizur Rahman, Ferdous Ahmed Sohel, Parvez Naushad, S. M. Kamruzzaman, "Text Classification using the Concept of Association Rule of Data Mining",Proc. International Conference on Information Technology, Kathmandu, Nepal, pp. 234-241, May. 2003.

[24] Keyur J. Patel, Ketan J Sarvakar, "Web Page Classification Using Data Mining", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 7, July 2013.

[25] Kirk Roberts, Michael A. Roach, Joseph Johnson, Josh Guthrie and Sanda M. Harabagiu: EmpaTweet: Annotating and Detecting Emotions on Twitter. In: LREC, p. 3806-3813(2012).

[26] The Apache Hadoop Homepage.[Online]. Available: https://hadoop.apache.org/