International Journal of Information & Computation Technology.
ISSN 0974-2239 Volume 4, Number 12 (2014), pp. 1145-1153

http://www. irphouse.com

Emotion Mining Techniques in Social Networking Sites

[1]Sanjeev Dhawan, [2]Kulvinder Singh, [3]Deepika Sehrawat

[1,2]Faculty of Computer Science & Engineering,
University Institute of Engineering and Technology,
Kurukshetra University, Kurukshetra-136119, Haryana, India
[3]M.Tech. (Computer Engineering) Research Scholar
University Institute of Engineering and Technology,
Kurukshetra University, Kurukshetra-136119, Haryana, India.

Abstract

In today's scenario social networking sites are a part and parcel of our lives that are hard to be avoided by anyone. These SNS are contributing in bringing people together from various communities of similar interest at a single collaborated destination. This not only helps people in sharing their views, experiences, likes and dislikes regarding a common topic of interest. Moreover most of the active and vibrant users have a great impact on influencing others views and thoughts. Such an impact can be negative as well as positive depending upon the conveyer and receptor's ability to think and understand and these factors effect on how a person react to a particular situation and there by influence others.

Keywords: Emotion mining, SNS, mining techniques, PIM, influential text

Introduction

Emotions are a mandatory part of human nature that can be considered as hereditary. Also it has been found that expression of a particular emotion by different human being is identical. Some persistent emotions that last much longer result in mood. Mood can be a result of a combination of certain emotions of a person. On the whole emotions can be categorised into two: basic and complex. Basic emotions are joy, sadness, anger, fear, disgust and surprise as defined by Ekman [10], [11], [12]. The complex emotions are a combination of two or more basic emotions that are experienced by a person at an instance. Even though emotions have no clear boundaries but Plutchik [26], [27] proposes a theory with eight basic emotions which include Ekman's six along with trust and anticipation. Plutchik has organized the emotions in a wheel (Figure 1). The radius indicates intensity, closer to the centre, the higher the intensity. Plutchik [27] also proposed that the eight basic emotions form four contrasting pairs like joy & sadness, anger & fear, trust & disgust, and anticipation & surprise.

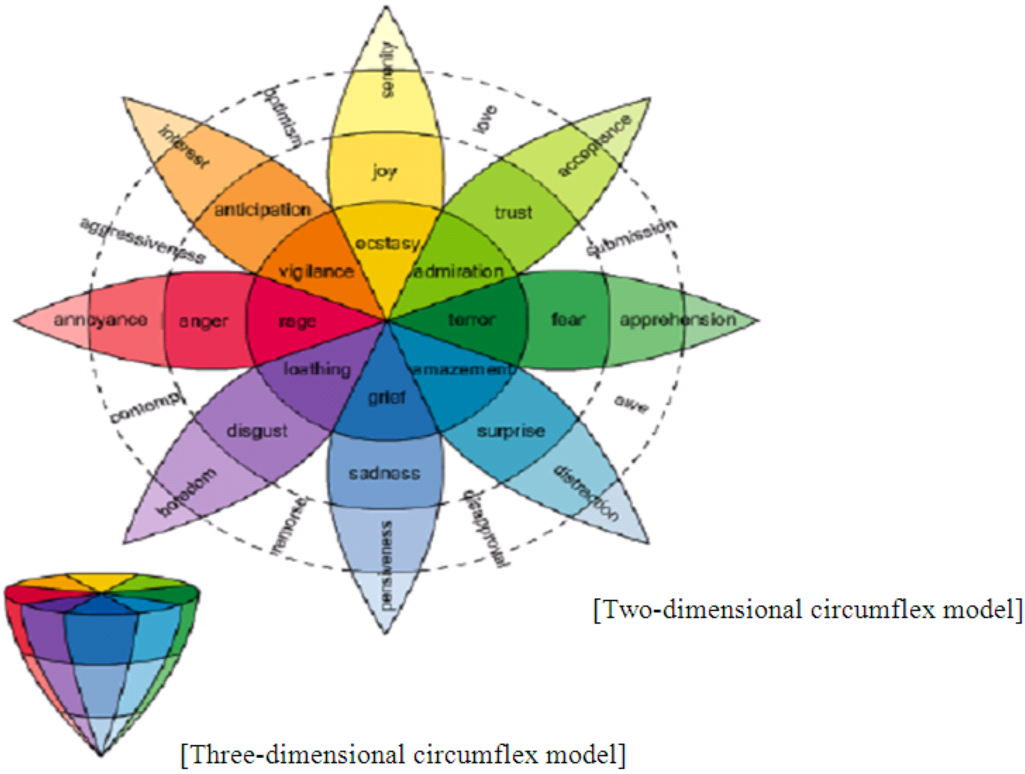

Fig. 1 Plutchik wheel of emotions [26]

There is scientific proof that at least five different emotions (fear, disgust, anger, happiness, sadness) are demonstrably different in the logic of triggering different combinations of brain parts adding surprise results in Ekman's main list of six basic emotions. Ekman's evidence found in support of emotions being basic is a set of six general characteristics common to all basic emotions. The above list ignores some emotions that are considered important by other researchers, like anxiety, guilt, shame, envy, jealousy, compassion and love. From the perspective of human felt experiences rather than at the neurological levels, it seems that there are two fundamental dimensions rather than a collection of differing kinds of emotions. First, the valence of an experienced emotion is the degree to which it is strongly positive or negative. Second level is level of arousal felt that is the amount of energy perceived. A consequence of this is that identifying valence and arousal linked to a particular word is likely to be far easier and more reliable than other types of emotion detection. Almost contracting the valence-arousal model of emotion perception, there is evidence that levels of positive and negative emotion are not correlated: a person can simultaneously experience varying levels of both, although they may be perceived as separate simultaneous emotions for instance, enjoying the fear in bungee-jumping or missing a loved one.

Related Work

The world has basically changed as the Internet has become a universal means of communication. Social Media are at the heart of our communications and are among the most visited places on the Web. SNS like Facebook, Friendster, MySpace and Orkut have established themselves as very popular and powerful platforms for finding friends and identifying other people who have similar interest. Search behaviour of Web users reflects the interest of Web users and this also leads to similar profiles. Some research has also being carried out so as to identify people who are highly associated with similar interest of Web search as stated by Gun Woo Park *et al*. [18]. Sentiment analysis (or opinion mining) is defined as the task of finding the opinions of authors about specific entities/events. The decision-making process of people is hindered by the opinions made by thought leaders and perceptions made by ordinary people [32]. Baumer *et al*. [4] also mentioned that social networking Websites provide a platform for engaging people belonging to different communities and regions. Social networks facilitate their users to communicate with people exhibiting different morals and social cultures. These Websites provide a very powerful medium for interaction among individuals that leads to mutual learning and sharing of valuable knowledge as studied by Sorensen [30]. The most popular social networking Websites are those that allow people to communicate with each other by joining different communities/groups. Social networking can solve coordination problems among people that arises due to geographical distance mentioned by Evans *et al*. [14], Li *et al*. [22] in their literature and can increase the success of social campaigns as said by Baumer *et al*. [4], Li & Khan [21] by disseminating the required information irrespective of time and place/location. However, in social networking Websites, people commonly use unstructured or semi-structured language for communication. In everyday life conversation (as noticed on social networking sites), people do not care about the spellings and accurate grammatical structure of a sentence that may leads to different types of ambiguities/ doubts, like lexical, syntactic, and semantic as mentioned by Sorensen [30]. Therefore, extracting logical patterns with correct information from such unstructured form is a critical task to perform. Their user-generated content allows the gathering of immensely many information that cannot be processed entirely by the users. Thus it is of interest to understand how these users are deciding, what pieces of information they are trusting and what are the reasons that influenced them as found out by Martin [20].

Text mining can be a solution of above mentioned problems. Sorensen defines text mining as a knowledge discovery technique that provides computational intelligence [30]. The technique consist of multidisciplinary fields, such as text exploration, feature retrieval, natural language processing, and mined data classification based on logical and non-trivial patterns from large data sets. The text mining techniques are turned out to be more complex in contrast to data mining due to the unstructured and fuzzy nature of text [19]. Unique challenges exist when we try to apply text mining to social media data. The data generated by social networking sites and blogs is big data. A large amount of unstructured and semi-structured data is generated on a daily basis, and old-style relational databases cannot efficiently support real-time analytics based on the text. Big data and NoSQL database provide a solution to this bulky data. Social media data, if not collected and adequately stored at regular intervals, is essentially delicate. Most open source social listening tools only store a few days' worth of social media data history. Twitter is the only SNS that recently announced that an entire history of data will be available, but that data will be restricted to comments posted specifically by the account holder [9]. This data is available from some of the large social data providers, such as Gnip and DataSift [9].

Since the last decade as the e-commerce has flourished so is the significance of knowing the views of users is also becoming important to companies who have adopted e-trading. Internet contains a treasure of data that can be mined to detect valuable opinions, with inferences even in the political field. The snippets of text are a gold mine for companies that want to observe their reputation and get timely feedback about their products and actions. It's not just in the field of online selling sites but also the SNS that have become a hub for people of similar interest a lot of valuable information can be mined from online social media. The process of mining personal information from social networking sites is known as PIM (Personal Information Mining) [3]. Personal information mining can find out the hidden relationship and features of the target people which can be used for active post operation. The original attributes which are included in the personal information usually have high dimension and redundancy which often lower down data mining efficiency.

Fig. 2 Flowchart of PIM [3]

The mining of emotions can be done at two levels, binary level and discrete level. Binary level defines a particular text with positive or negative polarity. But as the research was further carried out by several other researchers it was found out that binary level of sentiment analysis is not sufficient to analyse all the emotions related to a particular comment. After that basic emotions were mined as per defined by Ekman and Plutchik. Analysis of text having basic emotions is easier but text having sarcasm; condition and question need unique strategies. Sarcasm is extremely difficult to detect and it exists mainly in political perspective. One solution for identifying sarcastic sentences is described by Tsur *et al* [32].

A feature optimization method was proposed to resolve the problem by GuilanHu & Xiaochun Cai [17]. The method with the purpose of data dimensionality deduction is based on the association of rough set theory with PCA approach. The classification features are derived by following two steps of optimization and deduction operation.

Minas *et al*. [24] stated that the State-of-the-art techniques for probability sampling of users of online social networks (OSNs) are based on random walks on a single social relation that defines friendship between two nodes, and these methods rely on the fully connected graphs". Other than these relations other relations also exist between OSN users, such as involvement in the same group or participation in the same event. So as to exploit such graphs we perform a random walk on the union of multigraphs. A graph is randomly selected and walked through in a way to perform multigraph sampling after each iteration. He also concluded that multigraph sampling is a novel technique for random walk sampling of OSNs using multiple underlying relations because it is more robust to poor connectivity and clustering for node to node relations. Multigraph sampling generates probability samples in the same manner as conventional random walk methods [24]. Recently, researchers have used decision trees and hierarchical clustering as text mining techniques for group recommendation in Facebook where user can join the group based on similar patterns in user profiles and interests as per Baatarjav [5]. In research by Durga & Govardhan [8], the authors proposed a new model for textual categorization to capture the relations between words by using WordNet ontology. This has also been found out by Xu *et al*. [33]. The proposed approach maps the words comprise of same concepts into one dimension and present better efficiency for text classification. The authors indicated a best practice in information extraction process based on semantic reasoning capabilities and highlighted various advantages in terms of intelligent information extraction [33] as well as elaborated the suggested methods, like query expansion and extraction for semantic based document retrieval.

Most of the existing research papers cover the text mining techniques without mentioning the pre-processing phase as done by Yin *et al*. [34], Xu *et al*. [33] that is an important phase for the simplification of text mining process. Several techniques have been used to automate emotion mining. These can be generally, with some exceptions, classified into three categories: (i) Keyword Spotting: The first category employs Keyword Spotting as stated by Mohamed and Hazem [23]; it is based on a lexicon or a dictionary grouping words that have emotional implications. These techniques predict the emotions of the writer by identifying these affective words from the text. These techniques are popular because of their simplicity and economical advantage. However, they rely on individual word that is why they perform poorly when the sentence structure is more intricate, (ii) Lexical Affinity: The second category employs Lexical Affinity measures. These techniques are a bit more refined than keyword spotting where they assign for each word a probabilistic affinity for a certain emotion. Similar to keyword spotting, lexical affinity techniques perform poorly when facing intricate sentence structures, and (iii) Statistical Natural Language Processing: The third category uses Statistical Natural Language Processing techniques. These techniques employ machine learning algorithms to learn words' lexical affinities and words' co-occurrence frequencies [31, 7]. Unfortunately, the findings have no predictive value unless a large text corpus is used for training, especially in the social networks domain where the used language lacks proper structure and statistical rules are harder to learn; thus it might not be feasible to use these techniques simply because openly available training data are hard to find.

Ronen Feldman [29] has explained five problems within the field of sentiment analysis Document-level sentiment analysis; Sentence-level sentiment analysis; Aspect-based sentiment analysis; Comparative sentiment analysis; and, Sentiment lexicon acquisition.

Document-level sentiment analysis is assumed that the document contains an opinion on one main object expressed by the author of the document. There are two main approaches to document-level sentiment analysis: supervised learning and unsupervised learning. The supervised approach assumes that there is a finite set of classes into which the document should be classified and training data is available for each class. Unsupervised approach to document-level sentiment analysis are based on determining the semantic orientation (SO) of specific phrases within the document [29]. If the average SO of these phrases is above some predefined threshold the document is classified as positive and otherwise it is deemed negative.

In *Sentence-Level Sentiment Analysis* a single document may contain multiple opinions even about the same entities so for simplification we further assume there is a single opinion in each sentence. So as to implement this assumption we split the sentence into phrases where each phrase contains just one opinion. But this technique is applicable to only subjective data.

Aspect-based sentiment analysis (also called feature-based sentiment analysis) is the research problem that focuses on the recognition of all sentiment expressions within a given document and the aspects to which they refer. The entities that have many aspects or attributes and people have a different opinion about each of the aspects.

In some cases we need to identify the sentences that contain comparative opinions and to extract the preferred entities in each opinion. In most of the cases it is found that adjectives and adverbs are used such as most, more, least, superior, prefer are encountered in *comparative sentiment analysis*.

The *sentiment lexicon* is the most crucial resource for most sentiment analysis algorithms. There are three options for acquiring the sentiment lexicon: manual approaches in which people code the lexicon manually, dictionary-based approaches in which a set of seed words is expanded by utilizing resources like WordNet and corpus-based approaches in which a set of seed words is expanded by using a large corpus of documents from a single domain.

Automatic systems for analysing emotional content of text follow many different approaches: a number of these systems look for specific emotion denoting words [13], some determine the tendency of terms to co-occur with seed words whose emotions are known [28], some use hand-coded rules [25] and some use machine learning and a number of emotion features, including emotion denoting words [2], [1]. Recent work by Bellegarda [6] uses sophisticated dimension reduction techniques to automatically identify emotion terms, and obtains marked improvements in classifying newspaper headlines into different emotion categories. Goyal *et al.* [16] move away from classifying sentences from the writer's perspective, towards attributing mental states linked to entities mentioned in the text. Federico *et al.* [15] also stated that despite much progress in Natural Language Processing (NLP), this field is still a long way from a full Natural Language Understanding (NLU). Basically NLU requires processing and knowledge that goes beyond parsing and lexical lookup and that is not explicitly conveyed by linguistic elements.

Conclusions

Emotion generation and analysis have a number of practical applications including managing customer dealings, human machine interaction, information retrieval, natural text-to-speech systems, and in social and literary analysis. However, only a limited-coverage on emotion resources exist, and that too only for English language. Recent research has shown that it is advisable to handle different types of sentences by different strategies. Some specialized tools must be devised to mine particular emotions from variable sources of data which provide precise results. Also such tools need to be developed that can access and process multilingual data sets. We all are just near the basics; it's a long way to go.

References

[1] Aman, S. and Szpakowicz, S. "*Identifying expressions of emotion in text*", In V. Matou_sek and P. Mautner, editors, Text, Speech and Dialogue, volume 4629 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, pp. 196-205, 2007.

[2] Alm, C. O., Roth, D., and Sproat, R. "*Emotions from text: Machine learning for text-based emotion prediction*", In Proceedings of the Joint Conference on Human Language Technology / Empirical Methods in Natural Language Processing (HLT/EMNLP-2005), Vancouver, Canada, pp. 579-586, 2005.

[3] Aggarwal C., "*Text mining in social network"*, In Social Network Data Analytics, 2nd edition. Springer, pp. 353-374, 2011.

[4] Baumer E. P. S., Sinclair J. & Tomlinson B., "*America is like Metamucil: Fostering critical and creative thinking about metaphor in political blogs*", In Proceedings of 28th International Conference on Human Factor in Computing Systems ACM, Atlanta, GA, USA, pp. 34-45, 2010

[5] Baatarjav E., Phithakkitnukoon S. & Dantu R, "*Group Recommendation System for Facebook*", 2nd edition Springer, 2008.

[6] Bellegarda, J. "*Emotion analysis using latent affective folding and embedding"*, In Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, Los Angeles, California, 2010

[7] C. O. Alm, D. Roth, and R. Sproat, "*Emotions from text: machine learning for text-based emotion prediction*", In Proceedings of HLT/EMNLP"05, pp. 579–586, 2005.

[8] Durga A. K. & Govardhan A., "*Ontology based text categorization-telugu document*", International Journal of Scientific and Engineering Research, 2(9), pp. 1-4, 2011.

[9] Kimberly Chulis, "*Build social media datamarts using SPSS text mining tools*", 02 October 2012. http://www.ibm.com/developerworks/library/ba-social-media-spss-text-mining/?cmp=tumblr&cm=h&csr=tumblr_20130109&cr=tumblr&ct=tumblr&cn=20x20_SPSS

[10] Ekman, P. "*An argument for basic emotions*", Cognition and Emotion, 6(3), pp. 169-200, 1992.

[11] Ekman, P. "*Emotion in the Human Face*", Oxford University Press, 2005.

[12] Ekman, P. and Friesen, W. V. "*Unmasking the Face: A Guide to Recognizing Emotions From Facial Expressions*", Malor Books, 2003.

[13] Elliott, C. "*The affective reasoner: A process model of emotions in a multi-agent system*", Ph.D. thesis, Institute for the Learning Sciences, Northwestern University, 1992.

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

[15] Federico Neri, Paolo Geraci & Furio Camillo, "*Monitoring the Web Sentiment, the Italian Prime Minister's case*", International Conference on Advances in Social Networks Analysis and Mining, 2010 IEEE.

[16] Goyal, A., Rilo_, E., Daume III, H., and Gilbert, N. , "*Toward plot units: Automatic affect state analysis*", In Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, Los Angeles, California, 2010.

[17] GuilanHu & Xiaochun Cai, "*Research on Feature Optimization Method in Personal Information Mining*", WASE International Conference on Information Engineering, 2009.

[18] GunWoo Park, SooJin Lee, SangHoon Lee, "*To Enhance Web Search Based on Topic Sensitive_Social Relationships Ranking Algorithm in Social Networks*", IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology – Workshops, 2009.

[19] Kano Y., Baumgartner W. A., McCrohon L., Ananiadou S., Cohen K. B., Hunter L. & Tsujii T., "*Data Mining: Concept and Techniques*", Oxford Journal of Bioinformatics, 25(15), pp. 1997-1998. 2009

[20] Lionel Martin, "*User Behaviour under the Influence of Groups in Social Media*", EDIC Research Proposal, 2009.

[21] Li J. & Khan S. U., "*On how to construct a social network from a mobile ad hoc network*", Technical Report, NDSU-CS-TR-09-009, 2009.

[22] Li J., Li Q., Khan S. U. & Ghani N., "*Community-based cloud for emergency management*", In Proceedings of the 6th IEEE International Conference on System of Systems Engineering (SoSE), Albuquerque, USA, pp. 55-60, 2011.

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

[24] Minas Gjoka, Carter T. Butts, Maciej Kurant, and Athina Markopoulou, "*Multigraph Sampling of Online Social Networks*", IEEE Journal on selected areas in communications, VOL. 29, NO. 9, OCTOBER 2011.

[25] Neviarouskaya A., Prendinger H., and Ishizuka M., "*Compositionality principle in recognition of fine-grained emotions from text*" In Proceedings of the Proceedings of the Third International Conference on Weblogs and Social Media (ICWSM-09), San Jose, California, pp. 278-281, 2009.

[26] Plutchik R. "*A general psych evolutionary theory of emotion*". Emotion: Theory, research, and experience, 1(3), pp. 3-33, 1980.

[27] Plutchik R. "*The psychology and biology of emotion*", New York: Harper Collins, 1994.

[28] Read J. "*Recognising affect in text using point wise-mutual information*", Ph.D. thesis, Department of Informatics, University of Sussex, 2004.

[29] Ronen Feldman, "Techniques and Applications for Sentiment Analysis", Communications of the ACM, Vol. 56 No. 4, Pages 82-89, 2013.

[30] Sorensen L., "*User managed trust in social networking comparing Facebook, MySpace and LinkedIn*", In Proceedings of 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic System Technology, (Wireless VITAE 09), Denmark, pp. 427-431, 2009.

[31] T. Wilson, J. Wiebe, and R. Hwa., "*Just how mad are you? Finding strong and weak opinion clauses*", In Proceedings of AAAI-04, 21st Conference of the American Association for Artificial Intelligence, pp. 761–769, San Jose, US, 2004.

[32] Tsur, O., Davidov, D. and Rappoport, A. "*A great catchy name: semi-supervised recognition of sarcastic sentences in online product reviews*", In Fourth International AAAI Conference on Weblogs and Social Media, 2010.

[33] Xu X., Zhang F. & Niu Z., "*An ontology-based query system for digital libraries*", In Proceedings of IEEE, Pacific-Asia Workshop on Computational Intelligence and Industrial Application, Wuhan, pp. 222-226, 2008.

[34] Yin S.,Wang G., Qiu Y. & Zhang W., "*Research and implement of classification algorithm on Web text mining*", In Proceedings of 3rd International Conference on Semantics, Knowledge and Grid, China, pp. 446-449, 2007.