Global Journal of Computer Science and Technology
Network, Web & Security
Volume 13 Issue 13 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Emotion Detection and Event Prediction System

By Jahidul Arafat, Chongomweru Halimu, Mohammad Ahsan Habib & Rajib Hossain

Islamic University, Bangladesh

Abstract - The focus of this research is to build a cloud based architecture to analyze the correlation between social media data and events predictions. From analytical point of view this study refurbishes the viability of models that treat public mode and emotion as a unitary phenomenon and suggest the needs to analyze those in predicting the market event status of the respective companies. The major significance of this research is the normalization and the conversion process that has utilized vector array list which thereby strengthen the conversion process and make the cloud storing an easy process. Furthermore, the experimental results demonstrate its improved performance over the factor of emotion analysis and synthesizing in the process of prediction to extract patterns in the way events behave and respond to external stimuli and vice versa.

Keywords : public emotion, JSON, CSV, hashtag.

GJCST-E Classification : I.2.0

EMOTION DETECTION AND EVENT PREDICTION SYSTEM

Strictly as per the compliance and regulations of:

OPEN ASSOCIATION OF RESEARCH SOCIETY, USA
RESEARCH | DIVERSITY | ETHICS

Emotion Detection and Event Prediction System

Jahidul Arafat [α], Chongomweru Halimu [σ], Mohammad Ahsan Habib [ρ] & Rajib Hossain [ω]

Abstract - The focus of this research is to build a cloud based architecture to analyze the correlation between social media data and events predictions. From analytical point of view this study refurbishes the viability of models that treat public mode and emotion as a unitary phenomenon and suggest the needs to analyze those in predicting the market event status of the respective companies. The major significance of this research is the normalization and the conversion process that has utilized vector array list which thereby strengthen the conversion process and make the cloud storing an easy process. Furthermore, the experimental results demonstrate its improved performance over the factor of emotion analysis and synthesizing in the process of prediction to extract patterns in the way events behave and respond to external stimuli and vice versa.

Keywords : public emotion, JSON, CSV, hashtag.

I. Introduction

Micro blogging is an increasingly popular form of communication on the web. It allows users to broadcast brief text updates to the public or to a limited group of contacts. In this micro blogging, event prediction has attracted much attention from academia as well as business. but can an event really be predicted? Early research on event prediction [1], [6], [34]-[36] was based on random walk theory and the Efficient Market Hypothesis (EMH) [4]. According to the EMH events are largely driven by new information, i.e. news, rather than present and past prices. Since news is unpredictable, event forecast will follow a random walk pattern and cannot be predicted with more than 50% accuracy [2], [34].

However this research focused on to fetch the public emotions associating the recents events of world and classified them into distict format. It enables the researcher to do further analysis on public sentiments to depict how the emotions change over time. It thereby assists the event management to decide on their operational pattern.

This paper is organized as follows: section 2 presents literature reviews on the public emotional context while section 2.1 defines the public emotion and characteristics of tweets are described in section 2.2. Furthermore, section 3 and its subsequent sections presets technologies incorporated for sentiment management such as Query Generation and Tweet Management Technique and section 3.2 presents the Opinion Detection system that can be applied on tweets. Later section 3.3 portraits several existing software and technologies for analyzing those tweets and depicting the real cause of sentiments. Next to it is section 4 which presents the research's aim with that of the design, implementation, evaluation strategies and outcomes and both the limitations and advantages of it. Recommendation on further development is depicted in section 5. At the end, section 6 presents the degree upon which the research outcomes meet the objectives and section 7 comes out with some concluding remarks.

Author α : Researcher, HT Research and Consultancy, UK.
E-mail : jahidularafat@yahoo.com
Author σ : Depatment of Computer Science and Technology, Islamic University, Uganda. E-mail : halim@iuiu.ac.ug
Author ρ : Department of Information and Communication Technology, Mawlana Bhashani Science and Technology University, Bangladesh. E-mail : tareqiut@yahoo.com
Author ω : Freelancer, oDesk, Bangladesh.

II. Related Literature on Sentiment Analysis of Social Medias

An increasing number of empirical analyses of sentiment and mood are based on textual collections of public user generated data on the web. Sentiment analysis and opinion mining are now a research domain in their own which sometimes referred to as \subjectivity analysis" | whose methods and applications were extensively surveyed in much details in [10].

Different methodological approaches have been used to extract sentiment from text. Some methods are grounded in natural language processing (NLP) and rely on word constructs (n-grams) found in text to extract sentiment towards a subject (favorable or unfavorable). NLP methods have been used to extract sentiment and opinion from texts such as camera [12] and pharmaceutical reviews [11]. Other techniques of sentiment analysis, rooted in machine learning, use support vector machines (SVM) to classify text in optimistic or pessimistic mood classes based on pre-classified training sets. SVM has been used to classify noisy customer feedback data [9] and movie reviews using a 5-point scale [8]. A number of hybrid methods that blend NLP and machine learning techniques have also appeared in the literature [15].

Besides the textual content discussed thus far (movie reviews, camera reviews, customer feedback), sentiment analyses have touched on many different kinds of personal online content. Personal websites such as blogs and online journals are often awash with emotive information and have been extensively used to deduct everyday happiness [10], explore trends and seasonality [14], forecast mood [17] and predict sales of books [18] and movies [20]. Some similar analytical tools operate entirely on the web: We Feel Fine5 constantly harvests blog posts for occurrences of the

Global Journal of Computer Science and Technology (E) Volume XIII Issue I Version I Year 2013

phrases "I feel" and "I am feeling" and provides statistics and visualizations of past and current geo-tagged mood states. A similar online site, Mood views, constantly tracks a stream of Live journal weblogs that are user-annotated with a set of pre-defined moods.

The results generated via the analysis of such collective mood aggregators are compelling and indicate that accurate public mood indicators can be extracted from online materials. Using publicly available online data to perform sentiment analyses reduces enormously the costs, efforts and time needed to administer large-scale public surveys and questionnaires. These data and results present great opportunities for psychologists and social scientists. Yet, while blogs have largely been analyzed for mood patterns, not much research has yet addressed social networking sites and micro blogging platforms. Recently, emotion has been extracted from public communication on Myspace [22], [28] and status updates on Facebook, but this study could not find any large scale sentiment analysis of Twitter, other than a study focused on micro bloggers' response to the death of Michael Jackson [13], [15]. This may be due to the fact that micro blogging and social networking sites are fairly recent forms of online communication (at least when compared to blogs).

Scale may be an issue as well. Sentiment analysis techniques rooted in machine learning yield accurate classification results when sufficiently large data is available for testing and training. Minute texts such as micro blogs may however pose particular challenges for this approach. In fact, the Twitter analysis of Jackson's death mentioned above was performed using a term-based matching technique based on the Affective Norms for English Words (ANEW). ANEW provides pre-existing, normed emotional ratings for nearly 3,000 terms along three dimensions (pleasure, arousal, dominance) [21] and has been recently employed to measure mood of song lyrics, blog posts and U.S. Presidents speeches [26]. Since it doesn't require training and testing, this syntactical approach may enable sentiment analysis for very small text data where machine learning techniques may not be appropriate. Choi and Varian [27] shows that using emoticons as labels for optimistic and sentiment is effective for reducing dependencies in machine learning techniques.

a) Defining Public Emotion

For the purposes of this research, we define sentiment or emotion to be a personal optimistic or pessimistic feeling. Table 1 shows some examples.

For example, the following tweet is considered neutral because it could have appeared as a news research headline, even though it projects an overall pessimistic feeling about General Motors:

RT @Finance Info Bankruptcy filing could put GMon road to profits (AP) http://cli.gs/9ua6Sb #Finance.

In this research, we do not consider neutral tweets for further analysis. we only use optimistic or pessimistic tweets. Many tweets do not have sentiment, so it is a current.

Emotion	Query	Tweet
Optimistic	jquery	dcostalis: Jquery is my new best friend.
Neutral	San Francisco	schuyler: just landed at San Francisco
Pessimistic	exam	jvici0us: History exam studying ugh.

Table 1 : Example Tweets

b) Characteristics of Tweets

Twitter messages have many unique attributes, which differentiates this research from previous research [14], [29], [30]:

1. Length: The maximum length of a Twitter message is 140 characters.
2. Data availability: Another difference is the magnitude of data available. With the Twitter API, it is very easy to collect millions of tweets.
3. Language model: Twitter users post messages from many different media, including their cell phones. The frequency of misspellings and slang in tweets is much higher than in other domains.
4. Domain: Twitter users post short messages about a variety of topics unlike other sites which are tailored to a specific topic. This differs from a large percentage of past research, which focused on specific domains such as movie reviews.

III. Opinion Detection in Twitter

a) Emotion Corpus based Method

Emotion Corpus Based Method is based on vector space model for calculating document similarity. For the emotion detection in tweets, an emotion corpus that is based on 8 basic classes can be used, E={Anger, Sadness, Love, Fear, Disgust, Shame, Joy, Surprise}, from [3]. Each class represents a dimension in the Boolean emotion vector of a tweet. Look for emotion words in a tweet, and if found, set the corresponding class dimension in the emotion vector to 1, otherwise it remains 0.

Tweet: I was on Main Street in Norfolk when I heard about tiger woods updates and it made me feel angry, on 2009-12-11. Emotion vector: (1, 0, 0, 0, 0, 0, 0, 0).

For all the tweets in a chosen time interval, a centroid of all corresponding emotion vector dimensions can be calculated, and this centroid is considered a document for each interval. For a given time interval T that contains N tweets, let V ={v1, v2, . . ., vN} be a set

Global Journal of Computer Science and Technology (E) Volume XIII Issue I Version I Year 2013

of vectors (with l = 8 dimensions each) generated from these tweets. Define centroid $\bar{v}$ for period T as [16]:

$$\bar{v} = \left(\frac{\sum_{k=1}^{k=N} v_k^1}{N}, \frac{\sum_{k=1}^{k=N} v_k^2}{N} \dots, \frac{\sum_{k=1}^{k=N} v_k^t}{N}\right) \quad (1)$$

After finding centroid vector for each interval, define the opinion similarity between two intervals T1 and T2 by calculating cosine similarity between their centroid vectors as suggested by [24]:

$$Sim(T_1, T_2) = \frac{\overline{v_1}.\overline{v_2}}{|\overline{v_1}||\overline{v_2}|} \quad (2)$$

b) Set Space Model

Set Space Model prescribes representing each interval by a single document which is the union of the tweets posted in that particular time interval. After removing the stop words and stemming the terms using Porter stemmer 3, collect all terms in a hash set for each interval as suggested by [24]. Define the similarity between two intervals T1 and T2 by calculating Jaccard Similarity [2]:

$$Sim(T_1, T_2) = \frac{|(Set)T_1 \cap (Set)T_2|}{|(Set)T_1 \cup (Set)T_2|} \quad (3)$$

To find the changes, neither corpus based method nor the set space model alone is suitable. For the corpus based method, a change in the centroid can be misleading when the interval has very few emotion words compared to its neighbors. For the set space model, a change in similarity does not by itself imply an opinion change, because not all of the words are emotion words. In this method, first analyze vector space similarity as suggested by [33]. If detect a possible change, validate it by analyzing the Jaccard Similarity.

Tn is a time break, if the followings are satisfied in both corpus based method and set space model:

$$Sim(T_{n-1}, T_n) < Sim(T_{n-2}, T_{n-1}) \quad (4)$$

$$Sim(T_{n-1}, T_n) < Sim(T_n, T_{n+1}) \quad (5)$$

c) Public emotion and opinion analyses of online corporate in Twitter: Software and Techniques

i. *Opinion Finder (OF)*

Opinion Finder (OF) is a publicly available software package for sentiment analysis that can be applied to determine sentence-level subjectivity [38], i.e. to identify the emotional polarity (optimistic or pessimistic) of sentences. It has been successfully used to analyze the emotional content of large collections of tweets [19] using the OF lexicon to determine the ratio of optimistic versus pessimistic tweets on a given day. The resulting time series were shown to correlate with the Consumer Confidence Index from Gallup4 and the Reuters/University of Michigan Surveys of Consumers5 over a given period of time. Weadopt OF's subjective lexicon that has been established upon previous work [3], [6], [20].

Like many sentiment analysis tools OF adheres to a uni-dimensional model of mood, making binary distinctions between optimistic and pessimistic sentiment [25]. This may however ignore the rich, multi-dimensional structure of human mood. To capture additional dimensions of public mood a second mood analysis tools, labeled GPOMS, that can measure human mood states in terms of 6 different mood dimensions, namely Calm, Alert, Sure, Vital, Kind and Happy can further be used [12], [17], [29].

ii. *GPOMS*

GPOMS' mood dimensions and lexicon are derived from an existing and well-vetted psychometric instrument, namely the Profile of Mood States (POMS-bi) [4], [22]. To make it applicable to Twitter mood analysis it can expanded the original 72 terms of the POMS questionnaire to a lexicon of 964 associated terms by analyzing word co-occurrences in a collection of 2.5 billion 4- and 5-grams6 computed by Google in 2006 from approximately 1 trillion word tokens observed in publicly accessible Web pages [6]. The enlarged lexicon of 964 terms thus allows GPOMS to capture a much wider variety of naturally occurring mood terms in Tweets and map them to their respective POMS mood dimensions. Then match the terms used in each tweet against this lexicon. Each tweet term that matches an n-gram term is mapped back to its original POMS terms (in accordance with its co-occurrence weight) and via the POMS scoring table to its respective POMS dimension. The score of each POMS mood dimension is thus determined as the weighted sum of the co-occurence weights of each tweet term that matched the GPOMS lexicon [32]-[34].

To enable the comparison of OF and GPOMS time series it can be normalized to z-scores on the basis of a local mean and standard deviation within a sliding window of k days before and after the particular date. For example, the z-score of time series Xt, denoted ZXt, is defined as [38]:

$$\mathbb{Z}_{X_t} = \frac{X_t - \bar{x}(X_{t\pm k})}{\sigma(X_{t\pm k})} \quad (6)$$

Where, (Xt±k) and (Dt±k) represent the mean and standard deviation of the time series within the period [t−k, t + k]. This normalization causes all time series to fluctuate around a zero mean and be expressed on a scale of 1 standard deviation.

The mentioned z-score normalization is intended to provide a common scale for comparisons of the OF and GPOMS time series [39]. However, to avoid so-called "in-sample" bias, Bayir [4] suggests not to apply z-score normalization to the mood and DJIA time series that are used to test the prediction accuracy of the Self-Organizing Fuzzy Neural Network.

iii. *SOFNN*

SOFNN (self-organizing fuzzy neural network) has been developed specifically for regressions, function approximation and time series analysis problems [29]. Compared with some notable fuzzy nerural network models, such as the adaptive-network-based fuzzy inference systems (ANFIS), self-organizing dynamic fuzzy neural network (DFNN) and GDFNN, SOFNN provides a more efficient algorithm for online learning due to its simple and effective parameter and structure learning algorithm . In some researches work, SOFNN has proven its value in electrical load forecasting, exchange rate forecasting and other applications [31]-[32].

IV. Research Design

a) Research Aim

The aim of this research is to build a cloud based architecture to further analyze the correlation between social media data and the financial markets.

b) Emotion Detection and Event prediction Approach-EDEPA

Roadmap 01: Emotion Detection and Event Prediction Approach

Precondition: Public emotions to be correctly classified
Post condition: Event has been predicted

1. Define the source of public emotion
   a. Develop a framework on fact and people's perspective of certain event
   b. Identify emotion categories: Optimistic, Pessimistic, Question Optimistic form and Question Pessimistic form
2. Defining emotion extraction and culturing scale
   a. Optimistic- Keywords: Love, affection, yes, agree
   b. Pessimistic- Keywords: No, Pessimistic, hate, disagree
   c. Question Optimistic form- Keywords: are, is, have, had, has, was, were, Love, affection, yes, agree
   d. Question Pessimistic form- Keyword: are, is, have, had, has, was, were, not, Pessimistic, hate, disagree
3. Document the people's sentiments or tweets into mentioned cluster
4. Store the tweets or feed into Array.
5. Do lexical analysis on the stored feeds to retrive the meaning of it.
6. To aid to this analysis process all the englih words associating the meaning to Obtimistic, pessimistic or neutral are recommended to be gatherd and stored into four different text file such as optimistic.csv, optimisticQuestion.csv, pessimistic Question.csv and pessimistic.txt.
7. Read those files using java File Reader function and bring the associating words into Array.
8. Compare those words with the each word in the comment field.
9. Finally, categorinzing the feeds into associating field of emotion.
10. Use Self-Organizing Fuzzy Neural Network to train the program for self learning using past event data

c) Technologies and methods used for implementation

i. *JSON*

This research acknowledges the use of JSON ((JavaScript Object Notation)) data structures that generates the outputs based on the inputs received from queries and display and store the resultant outcomes into cloud.

JSON is built on two structures: (a) A collection of name/value pairs. In various languages, this is realized as an object, record, struct, dictionary, hash table, keyed list, or associative array [11, 24]. (b) An ordered list of values. In most languages, this is realized as an array, vector, list, or sequence [25]-[27].

ii. *Hashtag*

A hashtag is simply a relevant word or series of characters preceded by the # symbol [11]. Hashtags help to categorize messages and can make it easier for other Twitter users to search for tweets [15].

When one search for or click on a hashtag he/she will see all other tweets that use the same hashtag (see Twitter Advanced search option) [21]. Only others who are interested in the same topic thread will likely be using the same hashtag. For example, if one search for "Apple company" then "#Apple" will assit in most for having that company oriented information instead of using "Apple" [12].

iii. *The tweet fetching and classification*

To normalize all the finding and store into a cloud this program has used a conversion technology to store the outcomes into respective .csv files as depicted in step 6. However, before that when fetching down the tweets or feeds, the available hangtag technology of tweeter is also used. Moreover to identify the emotional polarity (optimistic or pessimistic) of sentences lexical analysis was done using the keywords where Love, affection, yes, agree for optimistic feeds, No, Pessimistic, hate, disagree for pessimistic feeds and are, is, have, had, has, was, were, Love, affection, yes, agree for question form optimisitc feeds and finally, are, is, have, had, has, was, were, not, Pessimistic, hate, disagree for question form pessimistic feeds. Finally the public modes of the corresponding companies were converted and stored into four different csv file such as optimistic.csv, optimisticQuestion.csv, pessimistic Question.csv and pessimistic.txt.

This CSV file conversion process utilized vector array list for getting data and then put that data into JTable as depicted in step 4. Thus the desined process enable the reader to search for any event i.e IPLT20 using #IPLT20 and categorize it into the mentioned clusters and correlate those to predict the event, people's needs and outcomes.

V. Program Evaluation

a) Survey Results

The entire survey results have been clustered around three sessions as tabulated below to show how the public emotion on a particular company of share market changes over time.

Table 2 : Session Durations

Session Duration	1/04/2012 to 10/04/2012	11/04/2012 to 20/04/2012	21/04/2012 to 30/04/2012

i. *Result on IPLT20*

Hypothesis: IPLT20 will attain a success in 2012.

SEE FIGURE 1

Table 3 : Pattern of Public Emotion Changes: Result on IPLT20-Session Wise breakdown

	Session 01	Session 02	Session 03
Number of Tweets	1000	1672	936
The oldest one	#IPLT20 offerd spectator with surprise and provides them with pure entertainments.		

ii. *Result on London Olympic 2012*

Hypothesis: London Olympic 2012 will have a significant impact of regions economic development

SEE FIGURE 2

Table 4 : Pattern of Public Emotion Changes: Result on London Olympic 2012-Session Wise breakdown

	Session 01	Session 02	Session 03
Number of Tweets	1042	1673	976
The oldest one	**The rural emonomy has shown much promises**		

b) Summary of the Results: Impact of the results on predicting events causes

All these results tabulated above depict how the public emotions had changed over time on a particular event. The emotion carves and charts present the time varient fluctuations. This found to be high on optimistic response in some sessions but high on pessimistic responses in other sessions. Thereby these illustrate the public's changing nature and thought regarding a particular event. The oldest tweets on each event as depicted above with that of other fetched tweets portraits somehow the reasons behind all these changing emotion dimension. For IPLT20, figure 1 shown the optimistic though is on raise, while pessimistic downed to certain degree. This resembles the present success of the event and predict the future success forcasts. Though a marginal level on optimistic and pessimistic tweets results on the impact of London olympic on local economy, which thereby left the hypothesis 2 in neutral state to be predicted, while forecast, this event could produces a mixed impact on London's local economy.

c) Main Findings

Fetching the public emotion, analyzing them and enabling those to predict the share market pattern was the prime concern of this research to undertake. The findings of this study illustrates that the mathematical prediction morphology is of significant interest for the present event prediction system. The experiments indicate that the proposed emotion fetching and analysing solution enables event marketer to figure out the social impact of vulnerabilities. These further assist the share market leaders to do justification on their market setting while associating the public concern. These would also help the regulatory body of event management to have a holdings on the event setting by configuring event's social responses and activities.

VI. Conclusion

In this research we only focus on the events of India and UK to analyze the public mode and emotion associating it. Information fetching is done over an one month period on those share market companies to identify the factors and correlate them into event prediction process.

Event prediction has been an active area of research for a long time. The Efficient Market Hypothesis (EMH) states that event are largely driven by new information and follow a random walk pattern. Though this hypothesis is widely accepted by the research community as a central paradigm governing the markets in general, several people have attempted to extract patterns in the way events behave and respond to external stimuli. Concerning these issues, this paper acts as a guideline for the present research to design a model using JSON which will not only categorize the public emotions into different dimensions but will also act as a predictor for the future event.

This approach achieved maximum accuracy with a time lapse lower than other existing methods. The implimented solution efficiently fetch the relevent tweets, decode them into different dimensions of emotions and further associate them into the prediction process of event management.

References Références Referencias

1. Cheng-Te Li , Chien-Yuan Wang , Chien-Lin Tseng , Shou-De Lin, MemeTube: a sentiment-based audiovisual system for analyzing and displaying microblog messages, Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Systems Demonstrations, p.32-37, June 21-21, 2011, Portland, Oregon

2. Pedro Henrique Calais Guerra , Adriano Veloso , Wagner Meira, Jr. , Virgílio Almeida, From bias to opinion: a transfer-learning approach to real-time sentiment analysis, Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, August 21-24, 2011, San Diego, California, USA
3. ApoorvAgarwal , BoyiXie , Ilia Vovsha , Owen Rambow , Rebecca Passonneau, Sentiment analysis of Twitter data, Proceedings of the Workshop on Languages in Social Media, p.30-38, June 23-23, 2011, Portland, Oregon
4. M. A. Bayir, I. H. Toroslu, A. Cosar, and G. Fidan. Smart miner: a new framework for mining large scale web usage data. In *WWW*, pages 161–170, 2009.
5. H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li. Context-aware query suggestion by mining click-through and session data. In *KDD*, pages 875–883, 2008.
6. N. Diakopoulos and D. A. Shamma. Characterizing debate performance via aggregated twitter sentiment. In *Conference on Human Factors in Computing Systems (CHI)*, April 2010.
7. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pages 56–65. ACM, 2007.
8. W. Jin, H. H. Ho, and R. K. Srihari. Opinionminer: a novel machine learning system for web opinion mining and extraction. In *KDD*, pages 1195–1204, 2009.
9. H. Kwak, C. Lee, H. Park, and S. B. Moon. What is twitter, a social network or a news media? In *WWW*, pages 591–600, 2010.
10. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, 2007.
11. C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1–27:27, 2011.
12. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury. Micro-blogging as online word of mouth branding. In *CHI EA '09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, pages 3859{3864, New York, NY, USA, 2009. ACM.
13. B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1{135, 2008.
14. Dodds, P. S., and Danforth, C. M. Measuring the happiness of Large-Scale written expression: Songs, blogs, and presidents. Journal of Happiness Studies 116, 2009.
15. Gilbert, E., and Karahalios, K. 2010. Widespread worry and the event. In Proceedings of the International Conference on Weblogs and Social Media.
16. Hopkins, D., and King, G. A method of automated nonparametric content analysis for social science. American Journal of Political Science 54(1):229–247,2010.
17. Lindsay, R. Predicting polls with Lexicon,2008. Available Online (accessed on: 22 January, 2012) http://languagewrong.tumblr.com/post/55722687/predicting-polls-with-lexicon.
18. Manning, C. D.; Raghavan, P.; and Sch¨utze, H. Introduction to Information Retrieval. Cambridge University Press, 1st edition,2008.
19. Mei, Q.; Ling, X.; Wondra, M.; Su, H.; and Zhai, C. X. Topic sentiment mixture: modeling facets and opinions in weblogs. In Proceedings of the 16th International conference on World Wide Web,2007.
20. Ounis, I.; MacDonald, C.; and Soboroff, I. On the TREC blog track. In Proceedings of the International Conference on Weblogs and Social Media,2008.
21. Pang, B., and Lee, L. Opinion Mining and Sentiment Analysis. Now Publishers Inc,2008.
22. Velikovich, L.; Blair-Goldensohn, S.; Hannan, K.; and Mc- Donald, R. The viability of web-dervied polarity lexicons. In Proceedings of Human Language Technologies: The 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics,2010.
23. Wilcox, J. Forecasting components of consumption with components of consumer sentiment. Business Economics 42(4):2232, 2007.
24. Johan Bollena, Huina Maoa, Xiaojun Zengb Twitter mood predicts the event. Journal of Computational Science 2 (2011) p.1–8
25. S. Asur, B.A. Huberman, Predicting the Future with Social Media, 2010, http://www.arxiv.org arXiv:1003.5699v1.
26. S. Bergsma, L. Dekang, R. Goebel, Web-scale N-gram models for lexical disambiguation, in: Proceedings of the Twenty-first International Joint Conference on Artificial Intelligence (IJCAI-09), Pasadena, CA, 2009, pp. 1507–1512.
27. H. Choi, H. Varian, Predicting the Present with Google Trends, Tech. rep., Google, 2009.
28. P.S. Dodds, C.M. Danforth, Measuring the happiness of large-scale written expression: songs, blogs, and presidents, Journal of Happiness (July) (2009).
29. Edmans, D. García, O.Y. Norli, Sports sentiment and stock returns, Journal of Finance 62 (4) (2007) 1967–1998. Available Online (accessed on: 12 January, 2012) http://econresearchs.repec.org/RePEc:bla:jfinan:v:62:y:2007:i:4:p:1967-1998.
30. B.S. Frey, Happiness: A Revolution in Economics, MIT Press Books, The MIT Press, June 2008.

Available Online (accessed on: 22 January, 2012) http://ideas.repec.org/b/mtp/titles/0262062771.html.

31. Gilbert, K. Karahalios, Widespread worry and the event, in: Fourth International AAAI Conference on Weblogs and Social Media, Washington, DC, 2010, pp. 58–65. Available Online (accessed on: 9 April, 2012) http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/research/download/1513/1833.
32. H. Mao, X.-J. Zeng, G. Leng, Y. Zhai, J.A. Keane, Short and mid-term load forecasting using a bilevel optimization model, IEEE Transactions On Power Systems 24 (2) (2009) 1080–1090.
33. O'Connor, R. Balasubramanyan, B.R. Routledge, N.A. Smith, From tweets to polls: linking text sentiment to public opinion time series, in: Proceedings of the International AAAI Conference on Weblogs and Social Media, Washington, DC, May, 2010. Available Online (accessed on: 10 April, 2012) http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/research/view/1536.
34. Pak, P. Paroubek, Twitter as a corpus for sentiment analysis and opinion mining, in: N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, D. Tapias (Eds.), Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), European Language Resources Association (ELRA), Valletta, Malta, May 2010, pp. 19–21.
35. Pang, L. Lee, Opinion mining and sentiment analysis, Foundations and Trends in Information Retrieval 2 (1–2) (2008) 1–135.
36. R.R. Prechter, W.D. Parker, The financial/economic dichotomy in social behavioral dynamics: the socionomic perspective, Journal of Behavioral Finance 8 (2) (2007) 84–108. Available Online (accessed on: 3 June, 2012) http://dx.doi.org/10.1080/15427560701381028.
37. R.P. Schumaker, H. Chen, Textual analysis of event prediction using breaking financial news, ACM Transactions on Information Systems 27 (February (2)) (2009) 1–19. Available Online (accessed on: 11February, 2012) http://portal.acm.org/citation.cfm?doid=1462198.1462204.
38. X. Zhang, H. Fuehres, P.A. Gloor, Predicting Event Indicators Through Twitter I Hope It is Not as Bad as I Fear, Collaborative Innovation Networks (COINs), Savannah, GA, 2010. Available on (accessed on: 11 January, 2012) http:\\www.elseviercom/locate/procedia.
39. X. Zhu, H. Wang, L. Xu, H. Li, Predicting stock index increments by neural networks: the role of trading volume under different horizons, Expert Systems with Applications 34 (4) (2008) 3043–3054.

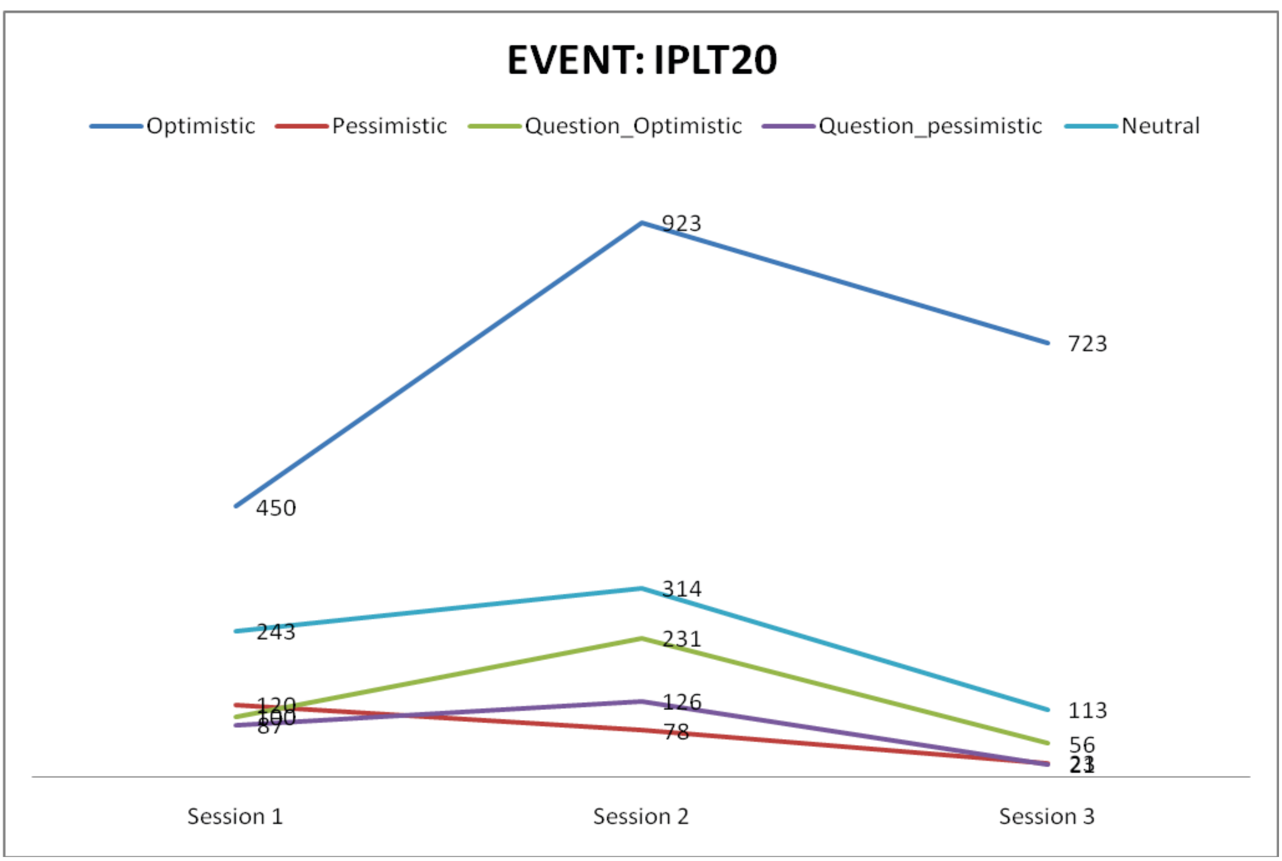

Figure 1 : Pattern of public emotion chnages: IPLT20

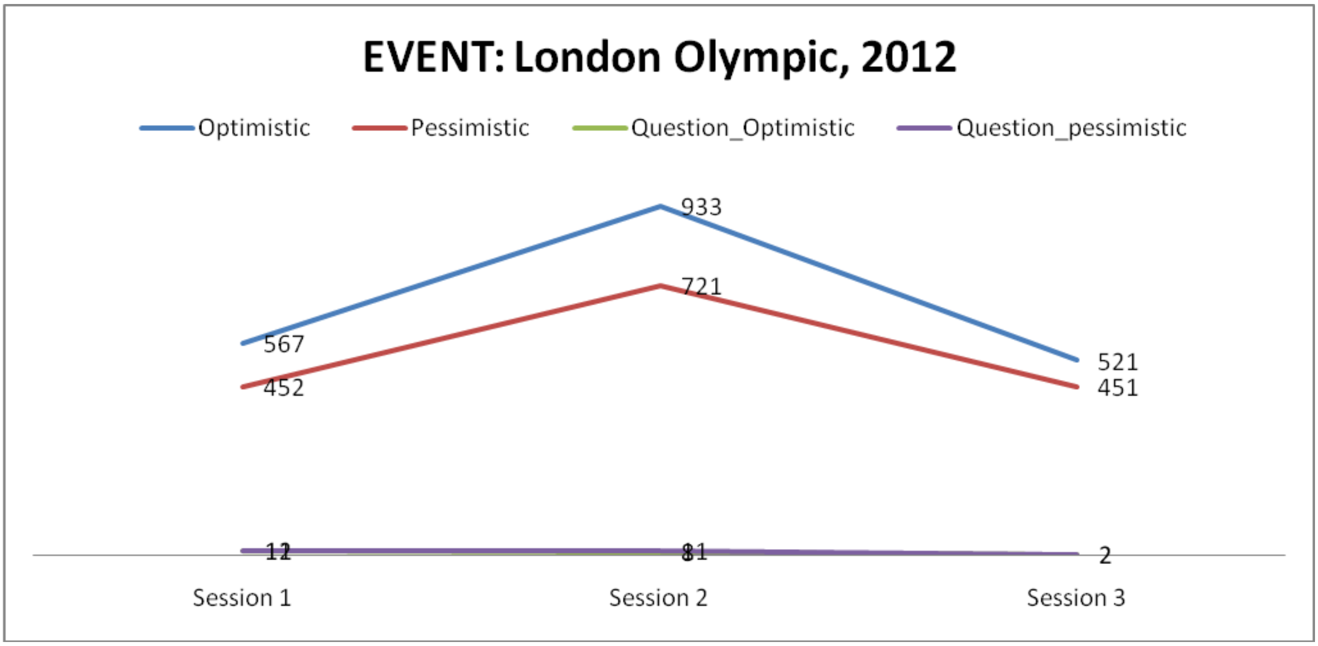

Figure 2 : Pattern of public emotion chnages: London olypmic, 2012