

TRCM: a methodology for temporal analysis of evolving concepts in Twitter

Article

Accepted Version

Adedoyin-Olowe, M., Gaber, M. M. and Stahl, F. ORCID: <https://orcid.org/0000-0002-4860-0203> (2013) TRCM: a methodology for temporal analysis of evolving concepts in Twitter. Lecture Notes in Computer Science, 7895. pp. 135-145. ISSN 0302-9743 doi: https://doi.org/10.1007/978-3-642-38610-7_13 (Proceedings, Part II. 12th International Conference on Artificial Intelligence and Soft Computing) Available at <https://centaur.reading.ac.uk/32163/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: http://dx.doi.org/10.1007/978-3-642-38610-7_13

Publisher: Springer

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

TRCM: A Methodology for Temporal Analysis of Evolving Concepts in Twitter

Mariam Adedoyin-Olowe¹, Mohamed Medhat Gaber¹, and Frederic Stahl²

¹ School of Computing, University of Portsmouth
Hampshire, England, PO1 3HE, UK

² School of Systems Engineering, University of Reading
PO Box 225, Whiteknights, Reading, RG6 6AY, UK

Abstract. The Twitter network has been labelled the most commonly used microblogging application around today. With about 500 million estimated registered users as of June, 2012, Twitter has become a credible medium of sentiment/opinion expression. It is also a notable medium for information dissemination; including breaking news on diverse issues since it was launched in 2007. Many organisations, individuals and even government bodies follow activities on the network in order to obtain knowledge on how their audience reacts to tweets that affect them. We can use postings on Twitter (known as tweets) to analyse patterns associated with events by detecting the dynamics of the tweets. A common way of labelling a tweet is by including a number of hashtags that describe its contents. Association Rule Mining can find the likelihood of co-occurrence of hashtags. In this paper, we propose the use of temporal Association Rule Mining to detect rule dynamics, and consequently dynamics of tweets. We coined our methodology Transaction-based Rule Change Mining (TRCM). A number of patterns are identifiable in these rule dynamics including, new rules, emerging rules, unexpected rules and ‘dead’ rules. Also the linkage between the different types of rule dynamics is investigated experimentally in this paper.

Keywords: Twitter, Hashtags, Association Rule Mining, Association Rules, New Rules, Emerging Rules, Unexpected Rules, Dead Rules, Rule Matching

1 Introduction

The surge in the acceptability of Twitter since its launch in 2007 has made it the most commonly used microblogging application [8, 16] (in this paper we used the terms *Twitter*, *Twitter network* and *network* interchangeably). The network permits the effective collection of large data which gives rise to major computational challenges. More people are becoming interested in and are relying on Twitter for information and news on diverse topics. Twitter is mainly known for short instant messaging that allows a maximum of 140 characters per message (tweet). Users follow other users’ comments or contributions on events taking place globally in real time [5]. The network is labelled the most commonly used

microblogging application around today with about 500 million estimated registered users as of June, 2012 [11]. Twitter has become a strong medium of opinion expression and information dissemination on diverse issues [20]. It is also a remarkable source for breaking news broadcasting [3, 14]. Considering the enormous volume of tweets generated on a daily basis, users have invented a common way of labelling tweets. This is often attained by including a number of hashtags as prefix to keywords in tweets to describe the tweet’s contents. The use of hashtags makes it easy to search for and read tweets of interest.

Tweets posted online include news, major events and topics which could be of local, national or global interest. Different events/occurrences are tweeted in real time all around the world making the network generate data rapidly. The network reports useful information from different perspectives for better understanding [22]. Twitter as a social network and hashtags as tweet labels can be analysed in order to detect changes in event patterns using Association Rules (ARs).

In this paper we use Association Rule Mining (ARM) to analyse tweets on the same topic over consecutive time periods t and $t + 1$. We also use Rule Matching (RM) to detected changes in patterns such as ‘emerging’, ‘unexpected’, ‘new’ and ‘dead’ rules in tweets. This is obtained by setting a user-defined Rule Matching Threshold (RMT) to match rules in tweets at time t with those in tweets at $t + 1$ in order to ascertain rules that fall into the different patterns. We coined this proposed method Transaction-based Rule Change Mining (*TRCM*). Finally, we linked all the detected rules to real life situations such as events and news reports.

This paper is organised as follows. Section 2 covers related work to our research. Methods for rule dynamics discovery including a background of ARM is provided in Section 3. Our proposed method *TRCM* is provided in Section 4. Section 5 provides an experimental case study. Potential applications of the proposed method are given in Section 6. The paper is concluded in Section 7 with a summary.

2 Related Work

The authors of [17, 18, 21] refer to Twitter as a medium of expressing sentiment and opinion on diverse issues. In the research of [4] the chatter from Twitter.com was used to forecast box-office revenue for movies. The work of [17] reveals that postings on Twitter can influence decision making of different entities. Twitter is also known as a strong medium for the broadcast of breaking news [15]. There may not be news agents around on the scene of incident but there will always be tweeters to broadcast the event live on Twitter even before traditional news agents appear on the scene. Twitter is as well a network for information dissemination and awareness creation on diverse topics [6, 22].

Related work on Twitter has shown that the network is a viable platform for evaluating people’s opinion. The work of [18] created classification models from sentiment expressed on a common disease in Brazil. On the other hand,

[22] applied the empirically authenticated MEDIC approach to identify interesting hashtags. This approach was applied to draw the attention of targeted users on the network based on follow links, tagged contacts and topographical /geographical fusion of users [20]. The research of [7] was based on influencer on the network. They employed Spearman's rank correlation coefficient to determine the strength of association between two rank sets. However, the experiment of [21] was based on a graph-based hashtag sentiment classification approach. Their major study was on sentiment polarity of collective hashtags within a specified time frame rather than individual hashtags polarity. The research of [20] was based on the formation of Twitter community and the aims of the different users that constitute the network. They used the HITS algorithm to propose a two-level structure for the user aim discovery and detection of core influencers in the network. Work on Twitter as word of mouth medium of advertisement was carried out by [10]. They used a case study approach on some branded products. The investigation of [4] was centred on box office revenues based on film reviews through Twitter using the LingPipe linguistic analysis package they called DynamicLMClassifier. The authors of [5] offered multinomial Naive Bayes, stochastic gradient descent and Hoeffding tree which are efficient in dealing with data streams to discover sentiment knowledge in Twitter. They also proposed sliding window Kappa statistic for assessment in time limited data streams. Furthermore, the work of [6] investigated the effect of public mood portrayed in tweets on the stock market. We identified the work of [12] as closely related to our work. However, they proposed Event Change Detection (ECD) to show how changes in event trends can be used for decision support in environment scanning. Our work used ARs to detect changes in hashtag trends based on time drift and event trends on Twitter. They discussed emerging event patterns which signify rules with momentous increase pattern. We applied ARs to discover hashtags that became popular as a result of sudden occurrence (for example breaking news), thereby shifting the attention of tweeters to another event altogether (emerging rules).

3 Rule Dynamics of Association Rule Mining

Association Rule Mining (ARM) is a data mining technique made popular by Agrawal et al [1], when they came up with the Apriori algorithm for the discovery of frequent itemsets and strong association rules. ARM is an act of extracting interesting recurrent representation, associations or links, between different arrays of items within transactional databases (Market Basket), relational databases (Personal Details), or any other information warehouses [12, 13] in the form of rules. It discovers and reveals remarkable associations embedded in huge data sets which may include hidden information that can be useful for decision making [9]. This technique tends to reveal every probable association that satisfies definite boundaries using the lowest support and confidence [2]. An association rule is in the form $X \Rightarrow Y$, where X and Y are disjoint sets of items. In this paper, we applied ARM to hashtags in tweets to discover trends in dynamic

Table 1. The Tweet Matrix

Tweet 1	#datamining	#bigdata	#sql	#KDD
Tweet 2	#ecommerce	#ISMB	#datamining	
Tweet 3	#bigdata	#facebook	#data mining	#analytics
Tweet 4	#analytics	#privacy	#datamining	
Tweet 5	#datamining	#KDD	#bigdata	

rules and identify when different rules appear and disappear on Twitter based on event changes.

We used the left hand side (*lhs*)/conditional and the right hand side (*rhs*)/consequent parts of rules in Apriori principle to analyse hashtags as conveyed in tweets over a period of time. The analysis of these hashtags is used to detect Association Rules (ARs) present in the tweets at different points in time. The similarities and differences in the ARs in the tweets at time t and time $t + 1$ are measured to reveal the ‘emerging’, ‘unexpected’, ‘new’, and ‘dead’ rules in tweets as they evolve from being strong rules to weak and then ‘dead’.

The capability of the ARM technique enables it to uncover different patterns in both transactional and relational datasets. Changes in rules dynamics patterns generated using the Apriori principle can be used as a decision support tool. An example of this is breaking news of a disaster, say an earthquake in Japan. The news will generate high rules in tweets at the early stage. This is referred to as speedy rule emergence. The emergence of this rule can result in rapid broadcast of breaking news by news agencies all around the world. It can also help other organisations like the ‘Red Cross’ to respond swiftly and dispatch aids to the affected areas more quickly.

3.1 Notation

Both [12] and [19] came up with methods for calculating similarities and differences between two rules at different points in time. Given that the methods adopted in [12] and [19] were designed for association rules discovered from relational datasets (general association rules). Our proposed methodology in this paper defines the similarity based on the concepts of degree of similarity proposed in [12] and [19]. Details of the calculations and notation used are given in the following.

n	Number of hashtags
i	an association rule in dataset 1 presented in binary vector
j	an association rule in dataset 2 presented in binary vector
lh_i/lh_j	number of hashtags with value 1 in conditional part of rule i/j
rh_i/rh_j	number of hashtags with value 1 in consequent part of rule i/j
lh_{ij}/rh_{ij}	number of same hashtags in conditional/consequent part of rules i and j
p_{ij}/q_{ij}	degree of similarity of features in conditional/consequent part of rules i and j
r_j^t	Rule present at time t
r_j^{t+1}	Rule present at time $t + 1$

3.2 Measuring Similarity

$$p_{ij} = \frac{lh_{ij}}{\max(lh_i, lh_j)} \quad (1)$$

$$q_{ij} = \frac{rh_{ij}}{\max(rh_i, rh_j)} \quad (2)$$

Equations 1 and 2 p_{ij} and q_{ij} were adopted from [12]. They show the similarity in the conditional and consequent parts of rule i and rule j at different points in time t and $t + 1$ in that order. Change in rules can be discovered by matching every rule in ruleset t with those in ruleset $t + 1$ using the similarity and difference comparison.

3.3 Rule Matching

The authors of [19] used ARM to mine the change in customer behaviour in an Internet shopping mall. They proposed RM as the most common approach to discovering rule changes in two datasets. In this paper, we used RM to detect change in patterns of rules in r_j^t and r_j^{t+1} , where t is the time and j is the rules present in the tweets. Rules in r_j^{t+1} are matched against rules in r_j^t to detect patterns of rule changes.

The RM threshold in our experiment is set in the interval $[0, 1]$, where 0 indicates dissimilarity and 1 indicates similarity. Please note that all rules are considered as ‘new’ until there is a matching rule found.

3.4 Patterns in Tweet Change Discovery

Having explained RM, we will now discuss Tweet Change Discovery (TCD) which is an integral part of our research. The TCD is used to develop our Transactional-based Rule Change Mining (TRCM) framework. In TRCM hashtags in the conditional and consequent parts of the rules are observed, as presented in Fig. 1, to detect the exact rule change pattern that has occurred. A framework showing the process of Tweet Change Discovery and our TRCM framework are presented in Fig. 1 and Fig. 2 respectively.

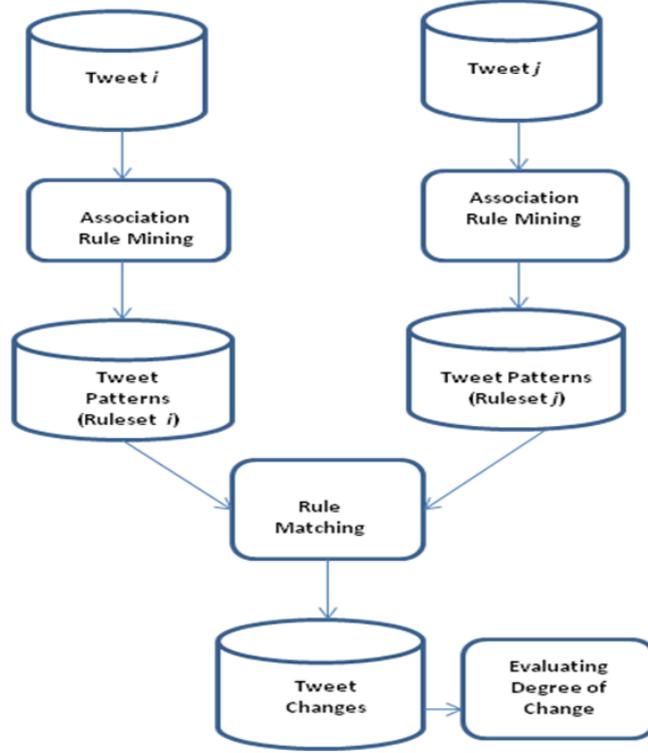


Fig. 1. The Process of Tweet Change Discovery.

The five patterns in TCD identified in this research are defined and explained as follows:

Unexpected Consequent Change in Tweet. Where a rule in r_j^t and r_j^{t+1} has similar conditional part but different consequent part ($p_{ij} \geq thp_{ij}$ and $q_{ij} < thq_{ij}$), then there is an unexpected consequent change provided the difference measure between r_j^t and rules r_j^{t+1} is greater than 0 (say 0.62). However, the status of this rule changes when there is a similar event pattern in $t+1$ that has a higher similarity measure (for example 0.90) than the initial difference detection. In this case an emerging change is said to have occurred. The implication of these two scenarios is that emerging change overwrites unexpected change.

Unexpected Conditional Change in Tweet. An unexpected conditional change is detected when the consequent parts of rule r_j^t at and r_j^{t+1} are similar, but the conditional parts are different ($p_{ij} < thp_{ij}$ and $q_{ij} \geq thq_{ij}$). If the absolute difference measure is less than 0, then the consequent part is similar and the conditional part is different. On the other hand, if the absolute value of the difference measure is greater than 0, then an unexpected conditional change is detected in the tweet.

Emerging Change in Tweet. Emerging rules are discovered when two hashtags at time t and $t + 1$ have similarities greater than the user-defined threshold ($p_{ij} \geq thp_{ij}$ and $q_{ij} > thq_{ij}$). The similarities take place in both the conditional and consequent part of the rule.

New Rules. All rules are said to be ‘new’ until there is a matching rule found. Every hashtag at time $t + 1$ is completely different from all the hashtags in time t ($p_{ij} < thp_{ij}$ and $q_{ij} < thq_{ij}$).

‘Dead’ Rules. A ‘Dead’ rule occurrence is the opposite of new rules detection. A rule in t is labeled ‘dead’ if its maximum similarity measure with all the rules in $t + 1$ is less than the user-defined threshold ($p_{ij} < thp_{ij}$ and $q_{ij} < thq_{ij}$). Fig.2 (rulesets t and $t + 1$) shows the different rules patterns already discussed.

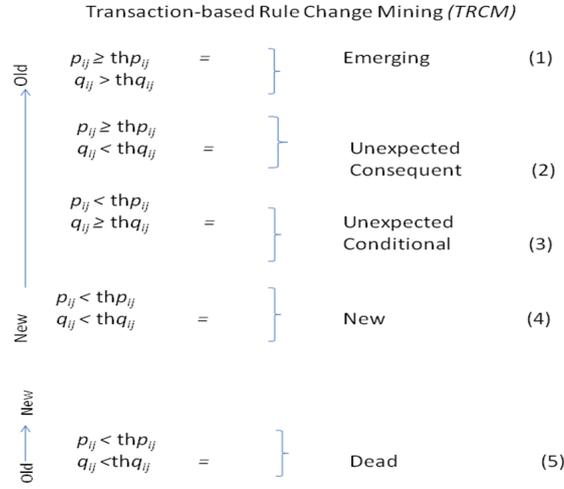


Fig. 2. Transaction-base Rule Change Mining (TRCM).

4 Transaction-base Rule Change Mining (TRCM)

TRCM is our research framework that defines rule change patterns in tweets at different period of time. In 1 - 4 hashtags in new rules (r_j^{t+1}) are matched with those in old rules (r_j^t).

On the other hand, in 5 hashtags in old rules (r_j^t) are matched with hashtags in rules r_j^{t+1} as shown in Fig.2. In order to measure the TCD we propose the following steps.

Step 1: For each rule in the conditional part of the new ruleset r_j^{t+1} , match with all rules in the conditional part of the old ruleset r_j^t . Compute the number

of hashtags that appear in the conditional parts of both rulesets. In this case we have #KDD, #excel appearing in r_j^t and r_j^{t+1} .

Step 2: Divide the computed figure in step 1 (2, #KDD, #excel) by the maximum number of hashtags with value 1 in the conditional parts of either the old or new ruleset. For example in the conditional part of ruleset r_j^{t+1} (new rules), 5 hashtags (#KNN, #sqlserver, #analytic, #Facebook and #Privacy) have value of 1 each. However in ruleset r_j^t (old rules) have 2 hashtags with value 1 (#DT and #CART). Since r_j^{t+1} have a higher number of 5, then:

$$p_{ij} = \frac{lh_{ij}}{\max(lh_i, lh_j)} = \frac{2}{5} = 0.4$$

Step 3: Apply the same method in step 2 to the consequent parts of the two rulesets to detect the q_{ij} .

$$q_{ij} = \frac{rh_{ij}}{\max(rh_i, rh_j)} = \frac{2}{3} = 0.6$$

Step 4: Identify the degree of similarity of rules in the old and new rules. However, for two rules to be similar, their degree of similarity must be greater than the pre-defined RMT which is 1. Where the degree of similarity is less than the RMT, the rules are said to be different. With the foregoing, the computation of the p_{ij} and q_{ij} in our sample datasets shows that neither the conditional nor the consequent parts of the rules are similar. This is because the p_{ij} (0.4) and the q_{ij} (0.66) are lower than the pre-defined RMT of 1.

5 Experimental Case Study

We applied the Apriori method on hashtags in rules r_j^t and r_j^{t+1} to mine the rule changes and detect the change pattern over time. This was obtained by declaring all the rules in the two datasets (Fig.3 and 4) as new rules. Rules in r_j^{t+1} were then matched one after the other with rules r_j^t . If there is a matching in rules, then the RMT is checked to determine the type of rule change found. All the rules in r_j^{t+1} were matched against those in r_j^t . Finally any rule in r_j^t that was still left unmatched is labeled 'dead' as the rule no longer exist.

6 Applications of TRCM to the Real World

Twitter is a dynamic network that produces several hundreds of millions of tweets every day. These data are often related to different events and event changes in the real world. Our proposed TRCM methodology is applied to tweets and these facts were discovered:

- Different entities such as business organisations, personalities and governments can use the TRCM architecture to discover change trends in tweets. The management of business organisations can adjust any aspect of their

lhs	rhs	supp	conf	lift
1{DT}	=> {#datamining}	1.0	1.000	1
2{CART}	=> {#DT}	1.0	1.000	1
3{#excel}	=> {#datamining}	0.4	1.000	1
4{#excel}	=> {#KDD}	0.4	0.625	1
5{#sql}	=> {#datamining}	0.4	0.435	1
6{#sql}	=> {#KDD}	0.4	0.555	1
7{#datamining}	=> {#KDD}	1.0	0.422	1
8{#KDD}	=> {#datamining}	1.0	1.000	1
9 {#datamining, #excel}	=> {#KDD}	0.4	1.000	1
10 {#excel, #KDD}	=> #datamining}	0.4	1.000	1
11 {#datamining, #sql}	=> {#Datamining}	0.4	0.112	1
12 {#KDD,#sql}	=> #datamining}	0.4	0.122	1

Fig. 3. Transaction-base Rule Change Mining (TRCM).

lhs	rhs	support	confidence	lift
1 {#KNN}	=> {#datamining}	1.0	0.111	1
2 {#sqlserver}	=> {#bigdata}	0.1	0.113	1
3 {#excel}	=> {#datamining}	1.1	0.110	1
4 {#KDD}	=> {#sqlserver}	0.1	1.000	1
5 {#excel}	=> {#datamining}	0.1	0.116	1
6 {#bigdata}	=> {#KDD}	0.1	0.111	1
7 {#analytic}	=> {#datamining}	0.1	0.110	1
8 {#KDD}	=> {#datamining}	1.1	1.000	1
9 {#bigdata}	=> {#excel}	0.1	1.000	1
10 {#Facebook}	=> {#datamining}	0.1	0.111	1
11 {#privacy}	=> {#datamining}	0.1	0.114	1
12 {#bigdata}	=> {#excel}	0.1	0.117	1

Fig. 4. Transaction-base Rule Change Mining (TRCM).

policy/decisions that does not receive a 'thump up' on the Twitter network. Real life examples are cases of the Apple iTunes icon change, GAP clothing company logo change and MySpace logo change (all within the year 2012). These changes are put forward by their organisations but received a high degree of criticism from the social media with Twitter network in the forefront. Using TRCM in future, these organisations will be able to detect early the (negative) reactions of their customers and make a prompt adjustment to their decision making. By doing so, detrimental changes are guarded. TRCM can therefore serve as a decision support tool for different entities that deal with people.

- News agents can also use TRCM to discover news trends and report news that arises from among other news as emerging news (news that the similarity degree is greater than the user-defined threshold). An example is the breaking news of an earthquake in Japan described earlier in this paper. Not

only will this news come across as an emerging rule, but also as a speedy rule emergence which will need to be reported immediately.

- Similarly TRCM is also useful to Twitter users, in that it will aid their activities on the network, which may be in the form of contributions, re-tweets and following topics on the network. Emerging trends can be captured and re-tweeted in order to increase its chances of being read and consequently draw the attention to the necessary audience of the topic.

The results of our research show the capability of TRCM of being a reliable tool in the effectiveness and beneficial use of Twitter network. More investigations will be conducted to discover how the degree of support and confidence can be used to reveal the rate of speed in rule change. Investigation will also be made to detect at what point different rules appear on the network and when they finally disappear. This experiment will propose a robust architectural framework that will be able to automatically detect all these changes.

7 Conclusion

Twitter data has become important to different entities in so many ways. To individuals, it has become a tool for awareness creation on diverse issues. It has also become a medium of information dissemination, which includes breaking news. Individuals sometimes rely on the Twitter network to obtain opinion of others while deciding on items to purchase, film to watch at the cinema or even political candidate to vote for during elections. The enormous data generated on the network need data mining techniques such as ARM to analyse and classify tweets, to detect their similarities and differences in relation to event patterns. The rule dynamics of ARM shows that tweets go through different phases. We proposed the Transaction-based Rule Change Mining (TRCM) that detects rules changes based on hashtags present in tweets and how the changes relates to events/occurrences in the real world. All the rules detected can be applied to the real world as a decision support tool for different entities, including individuals, organisations and government.

References

1. Agrawal, R., Imielinski., Swami, A.: Mining Association Rules between Sets of Items in Large Databases. In: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data (1993).
2. Ale, J., Rossi, G.: An Approach to Discovering Temporal Association Rules. Proc. of the 2000 ACM SAC , Vol. 1, 294-300. 2000.
3. Alfred, H.: "Twittering the news: The Emergence of Ambient Journalism." Social Science Research Network, Journalism Practice, Vol. 4, No.3, 297-308. 2010.
4. Asur, S., Huberman, B.: Predicting the future with social media. arXiv:1003.5699 [cs.CY]. 2010.
5. Bifet, A., Frank E.: Sentiment Knowledge Discovery in Twitter Streaming Data. Pfahringer, B., Holmes, G., and Mann, A H. (Eds.) DS 2010, 1 - 15, Springer-Verlag Brin Heidelberg, 2010.

6. Bollen, J., Mao, H., Pepe, A.: Modelling public mood and emotion: twitter sentiment and socio-economic phenomena. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2009.
7. Cha, M., Haddadi, H., Benevenuto, F., Gummadi, K.: Measuring User Influence in Twitter: The Million Follower Fallacy. In: Proceedings of the 4th International AAAI Conference on Weblogs and Social Media, 10-17. 2010.
8. Grossecck, G., Holotescu, C.: Can We use Twitter for Educational Activities. 4th International Scientific Conference. eLearning and Software for Education, Bucharest, April 17 - 18, 2008.
9. Jain, D., Sinhal, A., Gupta, N., Narwariya, P., Saraswat, D, et al.: Hiding Sensitive Association Rules without Altering the Support of Sensitive Item(s). In: International journal of artificial & applications (IJAIA), Vol.3, No.2. 2012.
10. Jansen, B.J., Zhang, M., Sobel, K., Chowdury, A.: Micro-blogging as Online Word of Mouth Branding. In: CHI'09. pp. 3859-3864. 2009.
11. Koetsier, J.: Twitter Reaches 500M users, 140M in the U.S. VB News 30 July, 2012. <http://venturebeat.com/2012/07/30/twitter-reaches-500-million-users-140-million-in-the-u-s/>. Retrieved 22/9/12 at 19.05.
12. Liu, D-R., Shih, M-J., Liau, C-J., Lia, C-H.: Mining the Change of Event Trends for Decision Support in Environmental Scanning. Science Direct, Expert Systems with Application Vol.36, 972-984. 2009.
13. Moen, P.: Data Mining Methods: Some Properties of Association Rules. Lecture notes, University of Helsinki/Dept. of Computer Science. 2005. <http://www.cs.helsinki.fi/u/ronkaine/tilome/luentomateriaali/TiLoMe-140305.pdf>. Retrieved on 4/8/12; 18.55 CET.
14. Murthy, D.: Twitter: Microphone for the Masses? Media, Culture & Society, Vol.33 (5). 779-789, doi: 10.1177/0163443711404744. 2011.
15. Nobil, A.: Is Twitter a Useful Tool for Journalists? Journal for Media Practice, Vol 11, No. 1, 145-155(11). 2010.
16. Pak, A., Paroubek, P.: Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In: Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), European Language Resources Association (ELRA), Valletta, Malta, May 2010, 19-21. 2010.
17. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. Now Publishers Inc, Hanover, USA. 2008.
18. Silva, I. S., Gomide, J., Barbosa, G. A., Santos, W., Veloso, A., Meira Jr, W., and Ferreira, R.: Observatorio da Dengue: Surveillance Based on Twitter Sentiment Stream Analysis 2011.
19. Song, H., Kim, J., Kim, S.: Mining the Change of Customer Behaviour in an Internet Shopping Mall. Expert Systems with Applications. Vol.21, 157-168. 2001.
20. Java, A., Song, X., Finin, T., Tseng, B.: Why We Twitter: Understanding Microblogging Usage and Communities. In: 9th WEBKDD and 1st SNA-KDD Workshop '07, August 12, 2007, San Jose, California, USA. Copyright 2007 ACM 1-59593-444-8. 2007.
21. Wang, X., Wei, F., Liu, X., Zhou, M., Zhang, M.: Topic Sentiment Analysis in Twitter: A graph-based hashtag sentiment classification approach. In: Proceedings of 20th ACM international Conference on Information and Knowledge Management, CIKM, USA, 1031-1040, 2011.
22. Weng, J., Lim, E-P., He, Q., Wing-ki Leung, C.: What do People Want in Microblogs? Measuring Interestingness of Hashtags in Twitter. In: Proceedings of the 10th IEEE International Conference on Data Mining (ICDM), Sydney, 1121-1126.