A novel approach for predicting the outcome of request in RAOP dataset

Conference Paper · September 2018

1 author:

Amreen Ahmad
Jamia Millia Islamia

A novel approach for predicting the outcome of request in RAOP dataset

Amreen Ahmad, Tanvir Ahmad, and Abhishek Bhatt

Department of Computer Engineering,Jamia Millia Islamia,
New Delhi, 110025, India
{amreen.ahmad10@gmail.com,tahmad2@jmi.ac.in,abhishek1995bhatt@gmail.com}

Abstract. In today's era online social communities such as Q & A sites are widely used for asking favours, so it would be beneficial to formulate a technique that would help in predicting the success of response. The objective of the paper is to enhance the accuracy of prediction of success of altruistic request that follows the same approach as used by ADJ [1]. Three more features are proposed i.e. topic, role and centrality in addition to the features proposed by ADJ [1] to capture user's interaction in the past and topic effect on the prediction of response. We also propose a graph based success prediction model(GSP) that uses feature weights and uses underlying graph structure for propagation to predict the outcome of request. Experiments were conducted on the RAOP dataset which belongs to sub-community of Reddit.com using GSP and it outperformed ADJ and other baseline methods using limited training data.

Keywords: Success prediction, altruistic request, social interactions

1 Introduction

In today's era of World Wide Web, online social communities are being widely used for noble cause. These social communities are often used by needy people for donations or help. Recently researchers have focussed on studying the underlying hidden factors that increase the probability of a request being responded [4, 5] and help in promotion of such request. Some key factors were identified by existing researchers that included social interaction between the giver and user, scale of the request, and whether the giver reciprocates. ADJ [1] work focussed on uncovering the effect of linguistic factor on success prediction of request given a situation in which the giver is not receiving any reward. It was also proved that linguistic factor that included politeness, evidentiality and narrative structure had a great effect on decision making and is strongly correlated with the success of requests.

The paper aims to improve the prediction accuracy of altruistic request and it follows the same approach as proposed by ADJ [1]. A series of textual features were developed by him and exploited linguistic aspect of altruistic request to analyse the correlation of these features with request success. ADJ work suffered from three limitations and they are i) Accuracy was not satisfying ii) it required a large dataset for prediction iii) it did not concentrate on different features importance and lastly iv) it did not take into consideration users interaction in the history that could be one of the important reason for the success of request. The central assumption of the paper is that inclusion of three extra features i.e. topic, centrality and role along with temporal, social and textual features would prove to be effective in boosting the accuracy of success prediction of request.

For the extraction and detection of subjective information from the text, opinion mining techniques have been widely used. The opinions can be represented as feature,document and sentence and the objective is to assign them to class. The proposed work uses opinion mining techniques on request dataset collected from RAOP which is a subcommunity of Reddit.com and eleven features are selected. Then a graph based success prediction model(GSP) is proposed to predict outcome of request. Finally for evaluation of GSP model, ADJ,LR and SVM is used.

The remaining part is structured as follows: Related work is discussed in section 2. The dataset and features selected for the proposed work is explained in detail in section 3. Next we proceed with explanation of proposed graph based success prediction model explained in detail in section 4. We proceed with an analysis of results obtained from evaluation described by section 5 followed by conclusion discussed in Section 6.

2 Related Work

In this section different factors of success such as status, user similarity, narrative,temporal dynamics, content,topic, role and centrality are reviewed. Online social networks provide provision to users to share other users content and interact with them. In the context of online popularity such type of interactions have been extensively studied [19, 32].The goal of performing such study is to understand the reason behind marketing campaigns and the drivers behind users consumption. Users interaction is one of the dimension that has not been studied till date. When studying different factors that are responsible for fulfillment of request in online community ”what is being asked for” is very similar to content. In the proposed work the aim is to understand how the user should ask for a request so that it gets fulfilled.

Previously researchers have laid emphasis more on temporal dynamics ahead of the content. It was studied by Ceyhan et.al [9] that loans easily get approved if it had some earlier funding. The objective of funded matters played a major role since it was observed that art and films less funding as compared to design and technology projects.

Politeness played a major role for the success of request and a computational model was built by Danescu-Niculescu-Mizil et.al [11] to analyse the relationship between them. It was proved by them that politeness is negatively correlated with power of the request which implies that needy people are more needy.

The significance of narratives was studied by Herzentein [18] for the success of loan funding. The loan borrowers ususally try to gain the trust of their lenders by placing themselves in a favourable manner in these narratives. The authors have analysed the importance of narratives for the success of fulfillment of request and have reached to the conclusion that narratives are more complex than any other identity.

Individuals on social media are often involved in a system of evaluation where they up-vote other users content. Researchers have studied that users similaty in characteristics [3] and staus of users [2] largely influences these evaluations. Similarity was measured as a combination of two factors: similarity of users [2] and similarity of social connections [3]. The results obtained by them proved that highly similar users are more likely to respond to each other irrespective of the difference in status.

3 Dataset and features

A brief description about the dataset and features selection is given below:

3.1 Request Dataset

The Pizza Request dataset is used which is a compiled version of RAOP dataset done by ADJ [1] and it comprises of entire collection from 8th December 2010 to 29th September 2013. In total it consists of 5728 pizza requests and 1.87 million relevant posts that are used for computing feature values.

3.2 Features

The features proposed by ADJ [1] are expanded to improve the prediction accuracy of requests. The features proposed by ADJ modelled request time /date, social reputation that are responsible for affecting the success of requests. A detailed description about the three proposed features i.e. topic, role and centrality are discussed. These three features are modelled to capture the users past interactions in Reddit.com that has taken place through comments and and analyse the effect of topic usage for the success prediction of request.

Textual Features ADJ [1] proposed six types of textual features that proved to be helpful for the fulfilment of request of needy people in RAOP dataset.

1. Narrative: Success of a request is significantly determined by its narrative [16, 24]. Each narrative word count is computed according to “Linguistic Inquiry and Word Count” and is considered as feature values [26].
2. Politeness: Researchers have established the fact that polite requests have higher potential of being fulfilled [7]. The computational politeness model is used to measure politeness of a request [12] and has developed 20 politeness strategies. These strategies are considered as a binary feature which indicates their presence or absence in the request text. The features are extracted from these politeness strategies using Stanford Dependency Parser [14].
3. Evidentiality: The request can be made more convincing by showing some evidence which proves its genuiness. The proof of evidence can be shown in the form of images such as proof of unemployment, passbook screenshot etc. Regular expression can be used to count the number of images and can be used as feature values if the presence or absence of image is known.
4. Reciprocity: It is generally observed that help is provided by only those people who have been previously helped [22]. Previous posted request can be observed for phrases like “return the favour” or “pay it back” and can be expressed in binary form that can be considered as feature values.

5. Sentiment: Sentiment analysis of the request is performed to analyse the mood of the person asking for help since it is considered that positive mood promotes help from others [15].
6. Length: Length of the request can be considered an important feature since longer request reflect the determined attempt of the needy person. The number of words that comprises a request can be taken into consideration as one of the feature.
7. Social Features: Recent studies have proved that interaction between people is highly correlated with the request's success especially in case of similarities between users and their status.
   (a) Karma: The Karma point captures user activity in Reddit.com i.e. up votes and down votes and can be considered as one of the feature.
   (b) User Similarity: It is believed that one is often eager to help similar thinking person [2]. The subreddit of the user is extracted and then Jacquard similarity is used for the similarity computation that exists between other users on RAOP and the requester.
8. Temporal Features: The temporal features are captured by measuring the specific weekday, hour of the day, day of the month and number of months.
9. Centrality features: We capture user interaction that takes place in Reddit.com through comments so that their effect can be analysed which would affect the success of a request. Based on the interaction of users a directed weighted graph is constructed where each node represents a user and the directed edge represents the comment from the user to the recipient. Let r_q represent the count of directed edges from one node to another, then weight on the edge is calculated as the reciprocal of r_q. Stronger interaction is represented by lower weight edge values. Centrality measure is used to calculate the importance of a user in the underlying graph such as in degree and out degree of a node and is considered as one of the feature value.
10. Role Features: the importance or role of a user can be measured in terms of the network communities in which he is active. It is believed that the lesser a person interacts in network communities, less is his need whereas if a user is active in more no of communities then his reachability is increased to more number of people and has high chances of receiving help from others [15]. Hence we aim to estimate the impact of role of a user on the success of a request. Louvain's algorithm is to detect communities in the underlying graph [26]. Role of a user in communities can be measured in following ways and is taken into account as one of the feature value.
    (a) Structural hole: It was suggested by structural hole theory[8] that intermediary nodes have higher chances of earning the favor since they have access to more information. The number of overlapping communities [30] is used to measure the the extent to which a node can be a structural node. The nodes that belong to more number of overlapping communities are considered as stuctural hole.
    (b) Core Node: Individuals who are highly active in a network community implies they have high TPR i.e triangle participation ratio(TPR is no. of triangles involved)and have high probability of receiving help from community members(since they are close to the core of a community) [23].
    (c) Bridging Effect: Some individual involved in the network community play the role of transmitting information and have proved to be effective in information diffusion. Some researchers have used rawComm to measure the bridging effect [27] of a node.
11. Topic features: It is assumed that the success of altruistic request is highly dependent on the topic to which it belongs. Topics such as loss or deprivation is likely to elicit public sympathy as compared to other topics like unemployment. Hence the objective is to select latent topic of request as features.For modelling latent topics based on the textual content of request,following ways are used:
    (a) Bag-of-Word: It is one of the widely used method for document classification where frequency of the occurrence of the word is considered as the feature. The words with AD(adverb), VA(adjective), VV(verb), NR(proper noun), and NN(noun) are considered as BOW features.
    (b) N-gram: In this case conditional probability is used to estimate the likelihood of the occurrence of a sentence. It is used to determine some important terms that are one of the factors behind the success of the request. Bi-gram and tri-gram are considered by eliminating stop words. Recursive Feature Elimination techinique [30]is used for the identification of discriminative and important features of bi-gram and tri-gram.

4 Graph based Success Prediction Model(GSP)

Using Graph based Success Prediction (GSP) model our aim is to predict the outcome of an altruistic request. GSP model contains two phases: Graph construction based on request text and learning based propagational optimization. A detail explanation of two phases of GSP is given below:

4.1 Graph Construction

In the underlying graph request text are represented as nodes and feature correlation between nodes is computed using probabilistic approach that is used to predict the outcome of unseen request. The graph is used to map the interaction between request nodes denoted by $G = (V, E)$. Let $V = U_{train} \cup U_{test}$, where U_{train} is a set of training nodes(t_r) and U_{test} is a set of testing nodes(t_{st}).Let node u represent testing request node i.e $u \in U_{test}$ and v represent training request node i.e $v \in U_{train}$. Now there are two steps: (a) Connect every node $u \in U_{test}$ to the top-$\lambda1$ similar node $v \in U_{train}$ and (b) Connect every $u \in U_{test}$ to the top-$\lambda2$ similar nodes u_w such that $u \neq u_w$ and $u_w \in U_{test}$, where $\lambda1 = k \times |U_{train}|$ and $\lambda2 = k \times |U_{test}|$, k is a parameter $\in [0, 1]$. In the proposed approach the value of k is varied to determine its efficiency and effectiveness. Probability of each request node for each label $sc(x)$ (for node x) can either be 1 (successful) or 0 (unsuccessful) denoted by $Psc(x)$. Firstly, the probabilistic scores for trained request nodes is set. For $v \in U_{train}$, let $P_1(v) = 1$ and $P_0(v) = 0$ if $sc(v) = 1$ else $P_1(v) = 0$ and $P_0(v) = 1$. Also for $u \in U_{test}$, let $P_1(u) = 0$ and $P_0(u) = 0$.

The edge of graph G denotes the correlation between requests texts based on their features. The process of edge creation is explained further.Let us assume that there is a certain feature f_t where f_t represents one of the feature of nodes,then feature based request correlation factor $F_c(n_1, n_2)$ is computed for nodes $(n1, n2) \in E$ that can be derived from difference in features $F_c(n_1, n2) = \|f_t(n1) - f_t(n2)\|$. For a given set of features denoted by $F = f_t$(where $t = 1, ..., z$) and z is the count of features, feature-aware correlation factor can be computed using feature based request correlation factor and is given as follows:

$$F_arc(n1, n2) = exp(-\sum_{t=1}^{z} w_t^2 \times F_c(n_1, n_2)) \tag{1}$$

where w_t denotes the weight of respective feature f_t. The $F_arc(n1, n2)$ is edge weight which can be denoted as $w_c(n1, n2)$ for edge$(n1, n2) \in E$ in the graph G.

4.2 Learning based propagational optimization

It works in two steps:

1. The probabilistic values of $P_1(u) and P_0(u)$ are infered from $P_1(v)$ and $P_0(v)$.
2. The probabilistic values of $P_1(u) and P_0(u)$ can also be derived from $u's$ neighboring nodes. Let degree of node u be denoted as d_u, then $P_1(u) and P_0(u)$ can be derived using given approach that enables us to get optimal set of edge weights in the graph:

$$P_{sc}(u) = \frac{1}{d_u} \sum\nolimits_{(u,y) \in E} w_c(u, y) \times P_{sc}(y) \tag{2}$$

where y denotes other nodes(both training and testing dataset) present in the graph G.

Furthermore, learning and optimizing w_t can enable us to compute similar probabilities for nodes in close vicinity of the testing node. In this model, a heuristic approach is proposed to learn the feature weights w_t based on entropy of the probabilities and the probabilities are fine tuned towards more accurate results. E_p denotes the propbability of success that is given as:

$$E_p(u) = \frac{1}{U_{test}} \times -\sum\nolimits_{sc \in (0,1)} P_{sc}(u)(1 - P_{sc}(u)) \log(1 - P_{sc}(u)) \tag{3}$$

where U_{test} denotes the count of requests in testing dataset of graph G and sc denotes success label.

Learning Optimization It is assumed that success probabilities computed would be of higher confidence if they are achieving lower entropy values. Gradient descent algorithm is used to minimize the entropy w.r.t w_t given by:

$$\delta E_p(u) = \frac{1}{U_{test}} \times \sum\nolimits_{y \in (U_test)} \frac{(\log(1 - P(y))}{P(y)} \times \delta P(y) \tag{4}$$

It is further differentiated using chain rule and the output is:

$$\delta w_t = 2 \times w_c(n1, n2) \times F_c(n_1, n_2) \times w_t \tag{5}$$

Prediction Likewise the feature weights w_t are updated with each iteration where $w_t = w_t - \delta w_t$. These updated feature weights w_t are used to derive new edge weights $w_c(n1, n2)$ using equation 1 and can be further used for the derivation of success probabilities of testing request $P_1(y)$ and $P_0(y)$ from the underlying graph as follows:

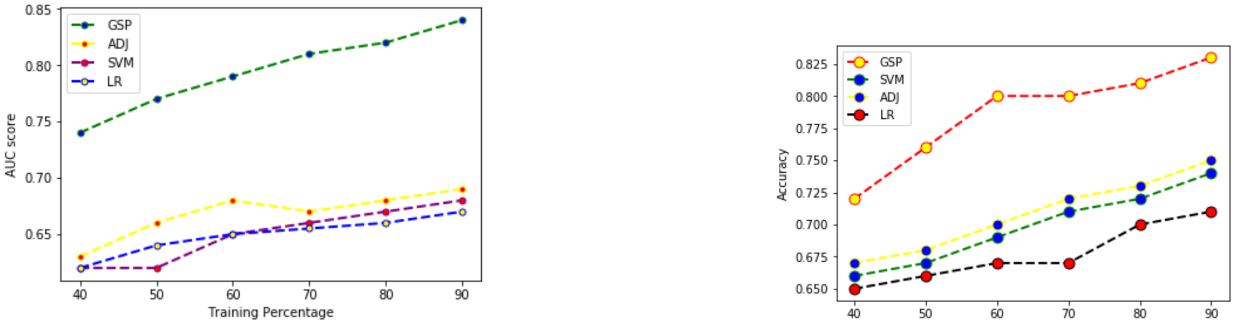
$$max_{P_{sc}u} = 2 \times w_c(n1, n2) \times F_c(n_1, n_2) \times w_t \tag{6}$$

The next step proceeds with the updation of E_p using equation 3 and the procedure is repeated iteratively till E_p converges. The success status of y nodes where $y \in U_{test}$ can be derived using success probability obtained from equation 5.

(a) AUC score

(b) Accuracy score

Fig. 1: AUC and Accuracy score when training data percentage is varied from 5%-30%



(a) AUC score

(b) Accuracy score

Fig. 2: AUC and Accuracy score when training data percentage is varied from 40%-90%

(a) AUC score of every feature category

(b) Accuracy score of every feature category

Fig. 3: AUC and Accuracy score of every feature category when Training:Testing=30%:70%

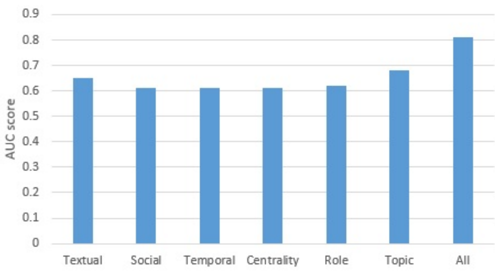
5 Evaluation

The evaluation procedure comprises of four steps: (i) First, the objective is to prove that the proposed GSP model can achieve high accuracy in comparison to ADJ for the same percentage of training data(t_r data). (ii) Second, the efficiency of GSP model with limited training data is proved. (iii) Third, the efficiency of the feature is proposed (iv) Lastly, the time efficiency of GSP is analysed for different values of k($0<k<1$).

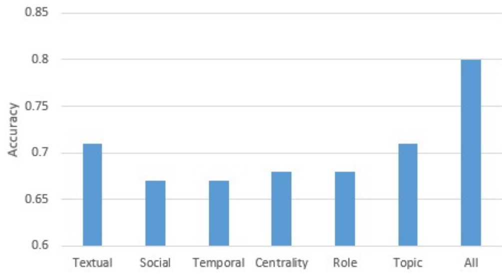
The t_r data is varied in the ratio 30%:70% and 70%:30% for analysis. Initially k is set to 0.1 and is varied to assess the efficiency of GSP model using different percentages of t_r data. The performance of GSP is evaluated using AUC score and accuracy where $accuracy = \frac{CPR}{t_s}$ where CPR is the count of requests that are correctly predicted and t_s is the sum total of testing data. The percentage of t_s and t_r data is selected and the experiment is repeated upto 200 times. Finally, the results are output as average of AUC and accuracy score.

5.1 Competitive methods

The proposed GSP model is compared with ADJ work[1] and other competitive classifiers such as SVM[31] and LR[13]. Fig. 1-2 presents the results for different percentages of t_r data. It can be seen from the obtained

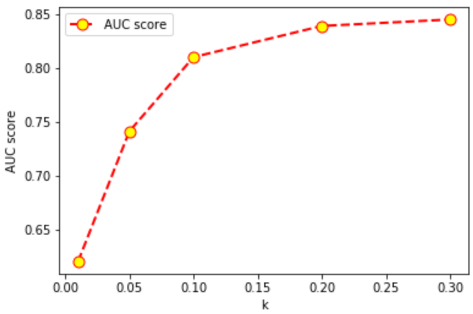


(a) AUC score of every feature category

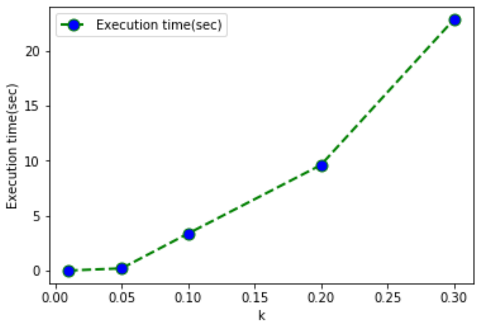


(b) Accuracy score of every feature category

Fig. 4: AUC and Accuracy score of every feature category when Training:Testing=70%:30%



(a) AUC score for different values of k



(b) Execution time for different values of k

Fig. 5: AUC score and execution time values obtained under the GSP approach

results shown in fig. 1-2 that GSP model outperforms ADJ and other competitive rivals for every case. As can be seen from fig.2(a) that the ADJ approach achieved 0.67 AUC score with 70% t_r data while 0.81 AUC score and accuracy was achieved by the proposed GSP with same amount of t_r data. This is mainly because of the incorporation of three extra featuresi.e centrality,role and topic that played a major role in success prediction. As seen from Fig. 3-4 that the three proposed features have performed competitively in comparison to social and textual features proposed by ADJ [1]. Among the three proposed features, topic feature has performed best in terms of generating a more accurate result. The results obtained prove that evidence and language usage(e.g.,narrative and politeness) features are less deterministic than topic fetaure. It can be noted that if the size of training data is increased, then there is a possibilty of improvement in performance of topic feature as compared to textual features. The reason behind it is that a large training data can uncover more number of hidden topics. Finally, a combination of all the above mentioned features helps in generation of good AUC score.

Fig. 5(a) and 5(b) reports the AUC score and execution time for different values of k. As can be seen from fig.5(a) when the value of k increases, AUC score also increases since it increases the size of the graph G. In fig. 5(b),although good results are achieved by increasing k value but alongside execution time also increases that adds to its disadvantage. But for k = 0.10 ,GSP takes 4 seconds to execute and its feasible for practical use so we set k = 0.10.

6 Conclusion

The proposed work focusses on collaborative knowledge obtained from sites Reddit.com and evaluates the efficiency of proposed GSP model for the prediction of real world events. Considering the work done by ADJ as base, a new methodology is proposed to predict the outcome of request. By considering three more features i.e topic,role and centrality apart from those proposed by ADJ and using graph based success prediction model (GSP), 0.81 AUC score was obtained. It was observed that proposed GSP model has outperformed other classifiers and has shown an improvement over ADJ in terms of AUC score with limited training data.

References

1. Althoff,T., Danescu-Niculescu-Mizil,C.,Jurafsky,D., How to ask for a favor: a case study on the success of altruistic request. In Proceedings of AAAI International Conference on Web and Social Media, ICWSM(2014)
2. Anderson,A.,Huttenlocher,D.,Klevinberg,J.,Signed Networks in Social Media. In Proceedings of the 5th ACM International Conference on Web Search and Data Mining, pp. 703-712, ACM,2012.
3. Anderson,A.,Huttenlocher,D.,Klevinberg,J.,Leskovec,J. Effects of User Similarity in Socila Media, in Proceedings of the 5th ACM International Conference on Web Search and Data Mining,WSDM,pp. 712-716,ACM,2012.
4. Backstrom, L., Huttenlocher, D., Kleinberg, J., Lan, X.,Group formation in large social networks: membership, growth, and evolution. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD, pp. 44–54 (2006)
5. Blei, D.M., Ng, A.Y., Jordan, M.I.,Latent Dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022 (2003)
6. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.Fast unfolding of communities in large networks. J. Stat. Mech.: Theor. Exp. 2008(10), P1000 (2008)
7. Brown,P.,Levinson,S.C.,Universals in Language use:politeness phenomena.Questions and Politeness:Strategies in Social Interaction,pp.56-311,Cambridge University Press,Cambridge(1978)
8. Burt, R.S., Holes, S.The Social Structure of Competition. Harvard University Press,Cambridge
9. Ceyhan,S.,Shi,X.,Leskovec,J.,Dynamics of bidding in a P2P service: effects of predicting loan success,In Proceedings of the 20th International Conference on World Wide Web,WWW,pp.547-556(2011)
10. Chierco, S., Rosa, C., Kayson, W.A.Effects of location, appearance, and monetary value on altruistic behavior. Psychol. Rep. 51(1), 199–202 (1982)
11. Danescu-Niculescu-Mizil,C.,Thuvain,P.,Launch hard or go home!:predicting the success of Kickstarter campaigns,In Proceedings of the 1st ACM Conference on Online Social Networks,pp.177-182,ACM,(2013)
12. Danescu-Niculescu-Mizil,C.,Sudhof,M.,Jurafsky,D.,Leskovec,J.,Potts,C.,A computational approach to politeness with applications to social factors,In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics,ACL,pp.250-259,2013
13. David A. Freedman,: Statistical Models: Theory and Practice. Cambridge Univer- sity Press. pp.128,(2009).
14. de Manneffe,M.C.,MacCartney,B.,Manning,C.D.,Generating typed dependency parses from phrase structure parses,In Proceedings of the Language Resources and Evaluation Conference,LREC,pp.449-454,(2006)
15. Fagan, J., Meares, T.L.,:Punishment, deterrence and social control: the paradox of punishment in minority communities. Ohio State J. Crim. Law 6, 173 (2008)
16. Greenberg,M.D,Paro,B.,Hariharan,K.,Greber,E.,Crowdfunding support tools:predicting success and failure,In CHI 2013 Extended Abstracts on Human Factor in Computing Systems,pp. 1815-1820 (2013)
17. Guha,R.,Kumar,R.,Raghavan,P.,Tomkins,A.,Propagation of trust and distrust,In Proceedings of the 13th International Conference on World Wide Web,pp.403-412,ACM,2004
18. Herzentein,M.,Sonenstein,S.,Dholakia,U.M.,Tell me a good story and I may lend you money: The role of narrative in peer-to-peer lending decisions,Journal of Marketing Research,48(SPL):S138-S149,(2011)
19. Lakkaraju,H.,McAuley,J.,Leskovec,J.,What's in a name?understanding the interpla between titles,count and communities in social media,In 7th International AAAI Conference on Weblogs and Social Media(2013)
20. Leskovec,J.,HuttenLocher,D.,Kleinberg,J.,Signed Networks in Social Media,In Proceedings of the SIGCHI Conference on Human Factors in Computing systems,pp.1361-1370,ACM,2010
21. Lou, T., Tang, J.Mining structural hole spanners through information diffusion in social networks. In: Proceedings of the ACM International Conference on World Wide Web, WWW, pp. 825–836 (2013)
22. Mallick,E.,The dynamics of crowdfunding:an exploratory study,J.Bus.Ventur.29(1),pp.1-16, (2014)
23. Malliaros,F.D,Vazirgiannis,M.,To stay or not to stay:modelling engagement dynamics in social graphs,In Proceedings ACM Conference on Information and Knowledge Management,CIKM,pp.469-478(2013)
24. Mitra,T.,Gilber,E.,The Language that gets people to give peolple:phrases that predict success on Kickstarter,In Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing,CSCW,pp.49-61(2014)
25. Narayanam,R.,Narahari,Y.,A Shapley value approach to discover influential nodes, IEEE Transaction on Automation Science and Engineering,pp.130-147(2011)
26. Pearce, J., Witten, K., Hiscock, R., Blakely, T.,:Are socially disadvantaged neigh- bourhoods deprived of health-related community resources?.In Int. J. Epidemiol. 36,pp.348355 (2007)
27. Scripps, J., Tan, P.-N., Esfahanian, A.-H.,Exploration of link structure and community-based node roles in network analysis. In: Proceedings of the IEEE International Conference on Data Mining, ICDM, pp. 649–654 (2007)
28. Tausizik,Y.R.,Pennebaker,J.W.,The psychological meaning of words:LIWC and computerized text analysis methods,J.Lang.Soc.Psychol.29(1),pp.24-54,(2010)
29. Tin Kam,:Random Decision Forests (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC,pp. 278282, 1416 August (1995)
30. Ugander, J., Backstrom, L., Marlow, C., Kleinberg, J.Structural diversity in social contagion.In Proc. Nat. Acad. Sci. U.S.A. (PNAS) 109(16), 5962–5966 (2012)
31. Vapnik, V.,: Support-vector networks. Machine Learning. 20 (3): 273297, (1995)
32. Yano,T.,Smith,N.A., What's worthy of comment? context and comment volume in political blogs,In ICWSM,(2010)