Implementing Feature Selection Technique to estimate influence from voting records

Sumit Laxman Nikalje, Pravin Rathod

Department of Computer Science And Engineering.

DIEMS College of Engineering Aurangabad

Abstract

Social platforms and network sites are increasingly used by people to express their opinions. Users like to spend their free time discussing the latest news, political issues, sporting events and new products. As a result, there is a growing interest in the use of social networks and social networking sites to recognize and predict opinions, as well as to understand the dynamics of opinions. For example, political parties regularly use social networks to recognize people's opinions about political discourse. Quantitative investment firms measure investor confidence and commerce using social networks and companies. This implementation will be tested using numerical simulations with measures of similarity and kernel methods in the voting records of the United States Congress dated 1984. The system will obtain a dispersed and dense matrix based on the attributes of the train and test models. The precision, AUROC and NMSE will be the key parameters of the index on which the system will be thoroughly tested.

Introduction

In recent years, a lot of literature has been developed about online voting. While online voting is an important part of research in recent years, efforts to develop solutions in the real world have just begun to create new challenges. Good records of misuse and recent safety infractions have been recorded. These challenges and concerns should be This was resolved to generate public confidence in online voting.

Therefore, properties are propagated through social relationships over time. Due to the importance of understanding evolution, many works have been done to analyze and simulate the network dynamics [citation], but most of the factors in these tasks can not be explained according to the theory.Social or empirical truth. And they always treat everyone without differences.

Distinguished from previous work, the objective is to understand the dynamics of social networks from the perspective of ideology and psychology.For this reason, therefore examining the mechanism of dissemination of opinions and features of opinions to create relationships at a personal level.The new loop pattern has been meticulously invented, where the network topology and personal ideology evolve together.

Literature Survey

A large body of previous art (see for example [13] - [15]) about digging social influences from information

related to finding features in a given social graph.For example, in [16] - [18] the author assesses the social influence matrix under previous knowledge in the social graph. The proposed method is different because evaluating both the social graph and the strength of influence by using votes from representatives in social groups only.In fact, this work is in the broader scope of the inverse problem in graph signal processing [19] which includes how to specify the network parameters that adjust the graph signal set. In this case, the signal is a vote in which the previous distribution is the result of changes in the social graph.The work involved in the context of social network inference is in [3], [20] - [22] different from this article [3], [20] - [22] assuming that a dynamic belief agent can directly observed. But in the form of thediscussions and votes: (i) Observation is a random action based on the belief of the agent and (ii) when the vote is believed to be in a stable state This method is more suitable to process the voting record, which is usually the culmination of many discussions.For this reason, the method is similar to [12], [23], [24] that takes advantage of constant data.In particular, creating work earlier in [12] here offers a systematic approach to processing vote data to extract information about the relative influence that varies according to the time of each affiliate representative

Challenges with existing system

1. They will not distinguish between passive and confident opinions. (Or comment) which is a noisy observation of opinions (Eg thumbs up / down confidence in the message)
2. They consider the opinions of users that will be updated at the same time, not continuous
3. The model parameters are difficult to learn from real detailed data and instead are arbitrarily set, so they give incorrectly accurate predictions.
4. They focus on the specific analysis of the stable status of user feedback by ignoring the temporary

behavior of real feedback changes, which helps forecasting methods, comments.

Proposed System

Proposed hybrid model that uses linear classifiers and kernel methods to get the results of each vote, with non-void actions that the agent can do. The model consists of two steps: conversation and voting. First of all, it is explained the conversation process. Suppose there is a $B\tau$ round of voting in the period τ and let $B\tau$: $= \{1, ...,B\tau\}$ be the set of all voting rounds. For voting in each round of $b \in B$ period during the agent period, the agent I hold the initial comment shows that the probability mass function is dimension M (PMF) xi (0, τ, b) $\in$ [0, 1] m associated with The agent's inclination towards voting for one of the possible decisions.

- To represent the underlying opinions of the user as a multi-dimensional random process xt (t), in which the item u-th, $x\leftarrow u$ (t) 2 R, represents the opinion of the user u at t and $\leftarrow$ means It may be based on the history of H (t). Then, every time a user posts a message at t, to draw m confidence from the distribution of confidence. P (m | $x\leftarrow$ u (t))
- Goal here is to develop an effective method that takes advantage of this model to anticipate the opinions of users. You xu (t) at t let history H (t0) reach time t0.
- To achieve this, it is used the LDA for modeling topics and NLP techniques for estimating influence.
- Obtain upper and lower boundaries to identify their own influence and classify according to the class value criteria as follows "Marginally Pass ", "Marginally-Not-Pass", "Pass", "Strongly Pass".
- Analytical Forecasting: In order to perform analytical forecasting, will use the least reliable method and Pearson's correlation

method and calculate conditional expectations for text intensity.

- **Various Stages in opinion forecasting**

An influential process: the process that influences the opinions described by the following:

Please note that since W (τ) is a random matrix of improved opinions, xi (t + 1, τ, b) is the correct pmf, stacking comments into the matrix X (t, τ, b): = (x1 (t, τ, b), ...,xN (t, τ, b)) T ∈ [0, 1] N × m,

xi(t + 1, τ, b) = PN j=1 Wij (τ)xj (t, τ, b), t ≥ 0

Generating Influence Matrix

In this model there are two types of agents: stubborn and stubborn. In particular, there are rebellious agents that have constant opinions throughout the debate, while the remaining unruly representatives are influenced by DeGroot's comments which have less influence on themselves. While the opinion of the rebellious agent cannot be dominated by others But they always try to convince others.

Voting Stage

At the 'voting' stage, the agents cast their votes according to their opinions. It is possible for an agent to abstain from voting and regard this as a null action, different from the others, because it provides no evidence of how the decision maker may have exerted his/her influence on his/her peers. As such, it is modeled the voting outcomes using two discrete random variables (r.v.s)2 — firstly, the absent indicator Ai(τ, b) ∈ {0, 1} is a Bernoulli r.v. with:
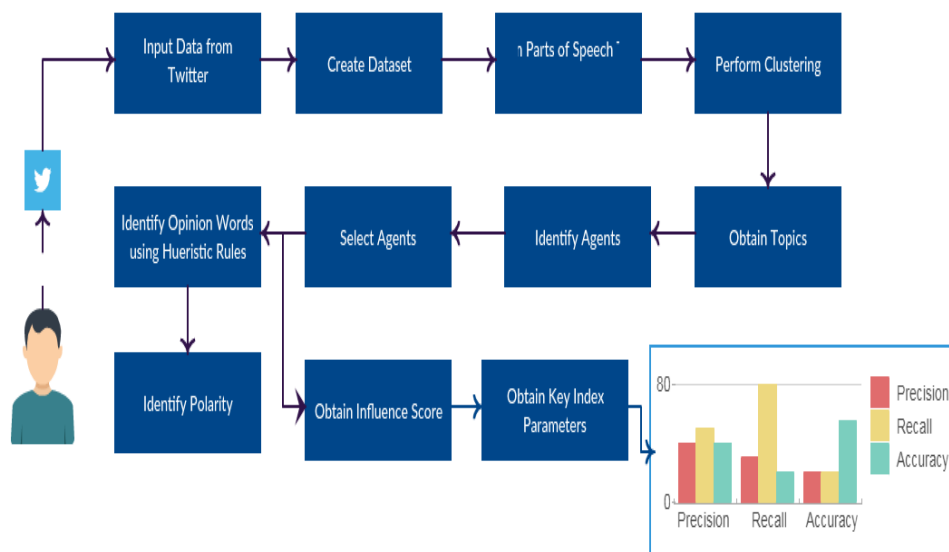
Pr(Ai(τ, b) = 1) = ai(τ, b) .

Figure 1 Proposed Architecture

The lower boundaries of model inference: one will be tempted to use the lower boundary statistics which result from the inequality of Van-Trees to assess the efficiency of inference. However, say that Bayesian Cramer-Rao bound (CRB) [43] cannot be used for problems that occur. In particular, the scope must use the previous distribution in order to have small support that is missing in the scope of support.

Obtain Gaussian And Logistic Distribution

The specific choice of confidence distribution p (m | x↩ u (t)) depends on the mark recorded. For example, one might consider: I. Gaussian distribution The confidence is considered to be a real random variable m 2 R, such as, p (m | xu (t)) = N (xu (t), u). The situation in which confidence is drawn from the message using confidence analysis. [13] The second

time, conviction logistics are considered binary random variables m 2 {1, 1}, such as, p (m | xu (t)) = 1 / (1 + exp (m • xu (t))) This is suitable for situations that measure confidence by voting up- Vote or like.

The goal here is to develop an effective method that elevates the model to predict user feedback. Xu (t) at the specified time. History H (t0) is time t0 <t.

In the context of the probability model, will predict this opinion by calculating expectations according to EH conditions (t) H (t0) [x↞ u (t) | H (t0)] where H (t) H (t0) means the average throughout the history from t0 to t, while the historical conditions of H (t0).

Using Feature Selection to remove erroneous subset or attributes

Algorithm 1 Proposed Feature Selection
Input: The feature id idle f t, first objective ob j1, second objective ob j2, |ob j1| = |ob j2| = |idle f t|.
Output: Non-dominated feature id idns, the second objective ob j2ns of non-dominated features.
1: k = 1;
2: for i = 1 : |idle f t| do
3: t = 0;
4: for j = 1 : |idle f t| do
5: if then(i! = j)
6: if then(ob j1(i) ≤ob j1(j)&ob j2(i) ≤ob j2(j));
7: else if then(ob j1(i) <ob j1(j)&ob j2(i) >ob j2(j))||ob j1(i) >ob j1(j)&ob j2(i) <ob j2(j));
8: else
9: t = 1;
10: break;
11: end if
12: end if
13: end for
14: if then(t == 0&j == |idle f t|)
15: idns(k) = i;
16: obj2ns(k) = ob j2(i);
17: k = k + 1;
18: end if
19: end for

Experimental Setup

For experimental setup the Senate data set is used from the UCI repository. The Senate call collection data is collected from the 114th Congress during the period of January 1, 2015 to September 28, 2016 and labeled a total of 490 billboards at the time From V1 to V490, especially the first 374 bills (about 75% of all data) from 1 January 2015 to 14 March 2016, divided into 3 periods and used to evaluate the influence matrix Each of which consists of the 120 120 votes and the remaining 134 116 tickets to be used for the test, see Figure 5 for illustration. Stubborn agent is used selection criteria. Identify 15 stubborn Republicans, 13 stubborn Democrats, 39 stubborn Republicans, 31 nonstubborn Democrats and 2 Independent Senators. People in the network For the grouping process by recording the vote each year and following the call, collect the first bill and ideology sponsor committee to establish the group. If some billing or billing is not specified, there is no specific board identifier. (For example, as a nominee) group votes based on voting categories, such as "edit", "seal", "nominate". To avoid numerical problems, collect groups that have comments based on (18) For example, every senator has at least one correct vote in the cluster. It is found that all K = 20 groups in the data set.

Republican influential network - Democrats: The area in the red box means that the Democrats have to agree with the Republicans more to approve bills. This is the reason, because the Republicans hold a majority in Congress, set the agenda which makes them more influential than just having more members. It is extended the two influence matrix and show basic information by listing specific senators.

ROC curve
The ROC curve (function curve of the receiver) is a graph showing the performance of the classification model at all classification criteria. This curve plots two parameters:
• True positive rate
• False positive rate
True Positive Rate (TPR) is a synonym for the recall and is defined as follows:
TPR=TPTP+FN

False Positive Rate (FPR) is defined as follows:
FPR=FPFP+TN
ROC TPR curve compared to FPR at different classification criteria Reducing the classification

criteria will make the list more positive, which will add both false positive and real positive values. The following figure shows general ROC curves.

Classification: ROC Curve and AUC

ROC curve

An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters:

- True Positive Rate
- False Positive Rate

True Positive Rate (TPR) is a synonym for recall and is therefore defined as follows:
$TPR = \frac{TP}{TP+FN}$

False Positive Rate (FPR) is defined as follows:
$FPR = \frac{FP}{FP+TN}$

ROC TPR curve compared to FPR at different classification criteria Reducing the classification criteria will make the list more positive, which will add both false positive and real positive values. The following figure shows general ROC curves.

In calculating points in the ROC curve, can evaluate logistic regression models multiple times with different classification criteria. But this is not effective Fortunately, there is an efficient sorting algorithm that can give us this information, which is called AUC.

AUC: Area under the ROC curve

AUC stands for "Area under ROC Curve", that is, AUC measuring all two dimensions under the entire ROC curve (think one calculus) from (0,0) to (1,1).

NMSE

$$NMSE = \frac{1}{N} \sum_i \frac{(P_i - M_i)^2}{\bar{P}\bar{M}}$$

$$\bar{P} = \frac{1}{N} \sum_i P_i$$

$$\bar{M} = \frac{1}{N} \sum_i M_i$$

In contrast to the bias in NMSE, the deviation (absolute value) will be combined instead of the difference. For this reason, the NMSE shows the most outstanding differences between models. If the model has a very low NMSE, it shows good performance in both space and time. On the other hand, high NMSE values do not necessarily mean that the model is totally wrong. This may be caused by time and / or moving area. Moreover, it must be pointed out that the difference in the peak is higher than the NMSE, rather than the difference in other values.The confidence interval for NMSE cannot be calculated from known distributions. Must use the bootstrap technique The same filtering for FAa calculations is used for NMSE calculations.

Conclusion

To summarize in this article, we propose a new strategy for pulling dynamic feedback models by collecting votes from the population. We developed a dialogue model and then voted as a model for voting to observe, in which voting will be made after the discussion period.

References
[1] C. of the EU. (2016) European parliament plenary.[Online].   Available: http://www.europarl.europa.eu/plenary/en/votes.html
[2] M. H. DeGroot, "Reaching a consensus," Journal of the American Statistical Association, vol. 69, no. 345, pp. 118–121, 1974.
[3] A. De, S. Bhattacharya, P. Bhattacharya, N. Ganguly, and S. Chakrabarti, "Learning a linear influence model from transient opinion dynamics," CIKM '14, pp. 401–410, 2014.
[4] A. Das, S. Gollapudi, and K. Munagala, "Modeling opinion dynamics in social networks," in Proc WSDM, 2014, pp. 403–412.

[5] A. G. Chandrasekhar, H. Larreguy, and J. P. Xandri, "Testing models of social learning on networks: evidence from a framed field experiment," Working Paper, 2012.

[6] D. Acemoglu and A. Ozdaglar, "Opinion dynamics and learning in social networks," Dynamic Games and Applications, vol. 1, no. 1, pp. 3–49, 2011.

[7] M. E. Yildiz and A. Scaglione, "Computing along routes via gossiping," IEEE Trans. on Signal Process., vol. 58, no. 6, pp. 3313–3327, 2010.

[8] W. Ben-Ameur, P. Bianchi, and J. Jakubowicz, "Robust Average Consensus using Total Variation Gossip Algorithm," in VALUETOOLS, 2012, pp. 99–106.

[9] U. A. Khan, S. Kar, and J. M. F. Moura, "Higher dimensional consensus: Learning in large-scale networks," IEEE Transactions on Signal Processing, vol. 58, no. 5, pp. 2836–2849, May 2010.

[10] P. Jia, A. MirTabatabaei, N. E. Friedkin, and F. Bullo, "Opinion dynamics and the evolution of social power in influence networks," SIAM review,

[10] P. Jia, A. MirTabatabaei, N. E. Friedkin, and F. Bullo, "Opinion dynamics and the evolution of social power in influence networks," SIAM

review, vol. 57, no. 3, pp. 367–397, 2015.

[11] C. Chamley, A. Scaglione, and L. Li, "Models for the diffusion of beliefs in social networks: An overview," Signal Processing Magazine, IEEE, vol. 30, no. 3, pp. 16–29, 2013.

[12] H. T. Wai, A. Scaglione, and A. Leshem, "Active sensing of social networks," IEEE Transactions on Signal and Information Processing over Networks, vol. 2, no. 3, pp. 406–419, September 2016.

[13] F. Chong, T. Chua, and E.-P.Lim, "Trust network inference for online rating data using generative models," in Proc KDD, 2010.

[14] W. Tang, H. Zhuang, and J. Tang, "Learning to infer social ties in large networks," in Proc ECML PKDD, 2011.

[15] J. Tang, H. Gao, H. Liu, and A. D. Sarma, "etrust: Understanding trust evolution in an online world," in Proc KDD, 2012.

[16] X.-Y. Zhang, "Simultaneous optimization for robust correlation estimation in partially observed social network," Neurocomputing, pp. 1–1, 2016.

[17] X. Zheng, Y. Wang, and M. A. Orgun, "Contextual sub-network extraction in contextual social networks," in IEEE Trustcom/BigDataSE/ISPA, vol. 1. IEEE, 2015, pp. 119–126.

[18] J. Huang, F. Nie, H. Huang, Y. Lei, and C. H. Ding, "Social trust prediction using rank-k matrix recovery," in IJCAI, 2013.

[19] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," IEEE Signal Processing Magazine, vol. 30, no. 3, pp. 83–98, 2013.

[20] M. Timme, "Revealing network connectivity from response dynamics," Physical Review Letters, vol. 98, no. 22, pp. 1–4, 2007.