Make the Right Choices: Dynamic Prediction on the Success of Crowdfunding Projects

Boshu Zhang, Chang Liu, Wei Xu

(School of Information, Renmin University of China, Beijing, 100872, China)

Introduction

With the rapid advance of Internet finance, the past decade has had a surge in considerable financial platforms, such as the crowdfunding. According to Oxford Dictionary, crowdfunding is a kind of practice of funding a real project or a venture through raising small money separately from thousands of people, especially via Internet. As a consequence of the penetration of crowdfunding around the world, crowdfunding provides different benefits for those who are expecting the financial support, for instance, quick access to capital and deep understanding what the market is asking for (Mollick, 2014).

Generally, the crowdfunding has been developed four forms: donation-based, reward-based, lending-based and equity-based crowdfunding. In addition, Ordanini et al. (2011) put forward that the crowdfunding model is composed by three mainactors: the project initiators who launch the project (we call them creators in this paper), individuals or groups who invest the project (we call them investors in this paper), and the platform which offers the creators and investors opportunities to get together. Nowadays, hundreds of crowdfunding platforms act as facilitators for crowdfunding (for instance, the Kickstarter (2009), IndieGoGo (2008) etc. which are globally famous). Creators are able to launch a project in the platform if satisfying the relative requirement. Investors can invest those projects they are interested in via any money they want. Moreover, a project is successful when it has reached the goal money in the given time. Besides, there exists two types of crowdfunding platforms: one is "Keep-it-All" (KiA), which means the creators can keep all money raised in the given time no matter whether the project is successful or not; another is "All-or-Nothing" (AoN), which means that the creators can keep nothing until they reach the launched goal in the given time.

It is a hot issue on the prediction of a project's success rate recently.Previous research mainly focuses on two kinds of information: hard information and soft information. Greenberg et al. (2013) apply different machine learning classifiers by using the hard information of KickStarteras input of their classifiers, and the accuracy of which is almost 68%. At the meantime, Mollick (2014) put forward that the factors that reflect the quality of the projects can also predict the success of projects, for instance, the number of video, duration and update times. Besides the potential factors to influence the final status of a project noted before, Etter et al. (2013) state that the time series of pledged money play an important role in predicting the success of projects, and propose a dynamic prediction model.Independent of hard information of projects, soft information of creators (e.g. personal networks) determines the success of projects as well (Etter et al. 2013; Mollick, 2014; Sorensen and Fassiotto, 2011; Stam and Elfring, 2008). Belleflamme et al. (2014) assume that investors are willing to invest for the reason that they enjoy the community benefits to increase their utility. Bayus (2013) studies the role of soft information in the dynamic behavior of investors, while Etter et al. (2013) feed the information of tweets and projects/investors graph to a support vector machine (SVM) to predict the success of the projects, which is proved to perform well.

However, to the best of my knowledge, there is only a few studies on the dynamic prediction on the success of Chinese crowdfunding projects. We use the data from a Chinese crowdfunding platform and our experiment is divided into two distinct parts: prediction model at the very beginning of the crowdfunding projects and the dynamic prediction model at each time point during the projects. To investigate the accuracy of our dynamic prediction model, we admit the dynamic prediction model proposed by Etter et al. (2013)as the baseline in our empirical analysis. Our empirical results validates that proposed prediction model indeed outperforms and we also discover some interesting phenomena in the context of Chinese crowdfunding, which will be discussed later.

The rest of this paper is organized as follows. Section 2will introduce the framework for crowdfunding prediction is illustrated. For the assessment of the efficiency and effectiveness of the proposed framework, empirical experiments and the experimental results are reported in Section 3. Finally, we offer concluding remarks and summarize the future directions of our research work.

The Proposed Method

An overview of proposed model is illustrated in Figure 1.

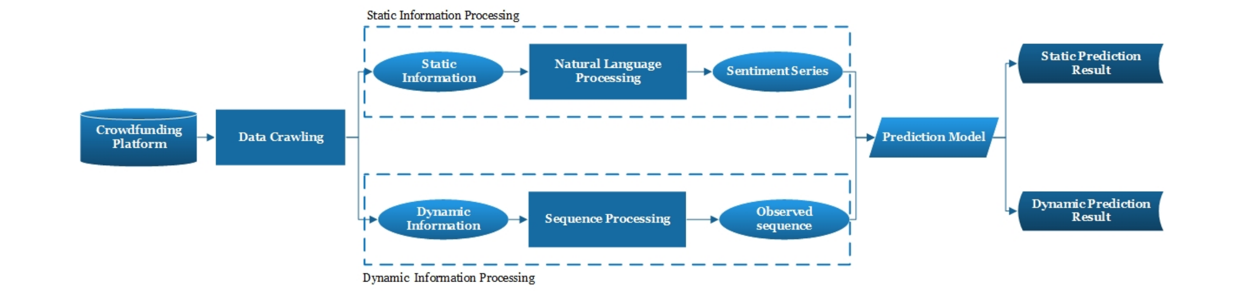

Figure 1. The Overview of Proposed Prediction Model

According to Figure 1, the first step is to crawl data from crowdfunding platform and divide the data into static and dynamic information respectively where static information indicates that all underlying natures we can get in each project while dynamic information presents those information updated every day.

Static Information Processing

According to previous research, the description of each project effects the success rate, so we use NLPIR (Natural Language Processing of Information Retrieval) for text description retrieval, which is a natural language processing technology and performs well in the context of Chinese words segment.

At the same time, in order to analysis the sentiment of description of each project, we admit one of the Chinese sentimental dictionary of National Taiwan University, shorted for NTUSD, classifying 11086 Chinese words into 2810 positive words as well as 8276 negative words respectively. As for the positive words, we label them as 1 and -1 for negative words.

We combine the NLPIR and NTUSD together to calculate the sentiment score of each description based on the counts of sentiment carrying words, where we hold the view that the positive projects are more likely to achieve their goal compared with the negative ones. Furthermore, we normalize each sentiment score into a continuous value between -1 and 1 which is widely adopted in the most sentiment analysis studies.

Dynamic Information Processing

Dynamic information is a kind of discrete information, which needs sequence processing to generate suitable time series for our prediction model. We divide the time into nine steps on the base of duration of the project considering about the pledged money each step, i.e. we will get a 9-dimension vector (10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90% of the duration respectively) and each dimension means the total money raised and the number of investors by the end of the current time point.

Prediction Model

Prediction model is divided into prediction model with only static information and dynamic prediction model with both static and dynamic information. Static information is used as input of prediction model to get static prediction results. Followed by sequence processing in 2.2, we feed the 9-dimension sequences as the input of prediction modelgetting the dynamic prediction results at each time point

Furthermore, to test our prediction model, we will introduce threedifferent activation functions to pick out a most effective one. To evaluate the performance of the classifiers, we accept the approach of 10-fold cross-validation which is widely accepted and will overcome the problems such as overfitting. The original sample is randomly divided into 10 subsamples equally and 9 of them are used to train the prediction model while the remaining one sub-sample is used as the validation data to test the model. The cross-validation will repeat 10 times, where every subsample will be used as the validation data exactly once. Then we will get 10 results which is averaged to calculate the final estimation of our proposed model.

Empirical Analysis

Data Description

We crawled data from a famous crowdfunding website in China named "Dreamore". Dreamore is one of the biggest crowdfunding platforms in China, which is a reward-based and AoN crowdfunding platform. Just like any other crowdfunding platforms, Dreamore provide initiators to post projects for raising funds, with a title, description (including pictures), goal money, duration, expected reward, etc. For another hand, as for the registered users can invest any project they are interested in via the information of each project.

A crowdfunding project is typically open for several days, which is called duration in our paper. According to Dreamore, when launching a project, the time limit is from one to twenty eight days, and it is free when duration is in seven days, it is 2% of total pledged money when the duration is between 8 and 14 days, 3.5% between 15 and 21 days and 5% between 22 and 28 days. As a consequence, it can be confirmed that the duration is concerned with the success rate of one project in some degree.

The average statistics of each project are given in Table 1, which follows the previous studies that the goal money of failed projects is much higher than the average and successful ones. Otherwise, we can also find the average duration of successful projects is lower than the failed while average number of both pictures and investors are higher than the failed, which is consistent with our common sense. However, what is different from previous studies is that the average goal of failed projects is staggering and we speculate that the cause may be the existence of singular data.

Table 1. Projects Statistics of Dreamore Dataset

	Successful	Failed	Total
Goal Money (yuan)	2588.17	115580.79	42124.78
Duration(days)	11.40	13.20	12.03
Final Amount (yuan)	3234.56	381.58	2236.26
Pictures number	3.60	3.09	3.42
Investors number	142	20	99

The dataset we use is scraped for the period between 3rd April and11th June for 70 days.Eventually, we get 2092 projects including 1360 successful and 732 failed. At the meanwhile, we divide the data into two types: static information and dynamic information of each project. As for the static information of project, we collect its ID as the primary key, as well as the title, goal money, category the project belongs to, duration, description and the final state.

On the other hand, dynamic data includes the state at each day, the days left, the amount of money that has been raised and the amount of investors who have funded the projectand we regard the id of project and date as the primary keys. Furthermore, the amount of raised money is normalized with respect to its goal. Then we generate 9-dimension vector through dynamic information processing method.

Empirical Results

We divide our experiment into two parts: prediction model only with static features and the dynamic prediction model with both static and dynamic features together. We run four variations of prediction model, feedingonly static features as input of backpropagation neural network (BPNN) with tansig, logsig and purelinas activation functions, and then we run the prediction model with the sentimentto demonstrate how sentiment effects the effectiveness of prediction model.

Prediction Model

We compare the forecast performance of four variations of prediction model and other common data mining modelsonly with static information we mentioned before, and the results are showed in Table 2:

Table 2. The Prediction Performance of Data Mining Models

Data mining model	TP rate	FP rate	Accuracy
BPNN-*tansig*	0.8569	0.2568	0.8225
BPNN-*logsig*	0.6206	0.0929	0.7208
BPNN-*purelin*	0.2044	0.0150	0.4775
BPNN-*tansig*without sentiment	0.7912	0.2965	0.7886
Logistic	0.691	0.435	0.678
Naïve Bayes	0.486	0.316	0.703
SMO	0.67	0.453	0.654
Random Tree	0.732	0.373	0.725
REP Tree	0.691	0.451	0.678
RBF Network	0.666	0.581	0.699

As can be seen from Table 1, three variations outperform the other models and BPNN with tansig as the activation function enhances the accuracy 6%. In our proposed models, the FP rates of four proposed models are all low, which indicates a strong ability to recognize the negative items. Taking the dataset into consideration, we think there are two reasons for this phenomenon: (1) the goal money of some projects is too high to reach. Some initiators do not evaluate the projects objectively, and overstate the attraction of them. For example, several projects label the goal money as one million or five million, which leads to the failure obviously. (2) The duration is too short to attract more investors. The Dreamore rules that a short duration (shorter than seven days) project can free from procedure fee, which results in a lot of short duration projects, and a lot of failed projects for less exposure to the investors.

In general,BPNN with tansig obtains a high TP rate, and BPNN with logsig and that with purelin perform better on FP rate. In practical application, the first model can be used to select likely successful projects, and BPNN with logsig and purelin can be used to avoid likely failed projects, which may be significant for further relevant application such as the recommendation of crowdfunding projects.

It is a pilot study to regard the sentiment of text description of a project as a fundamental factors when it comes to success prediction. In order to testify the sentiment effect, we feed two different inputs into BPNN and the result confirm that the emotion reflected in the text description does improve the prediction accuracy almost 3%.

Dynamic Prediction

We investigate the performance of our model dynamically in this part and we train BPNN with tansig as activation function for 50 times to choose the best performance as our empirical result.

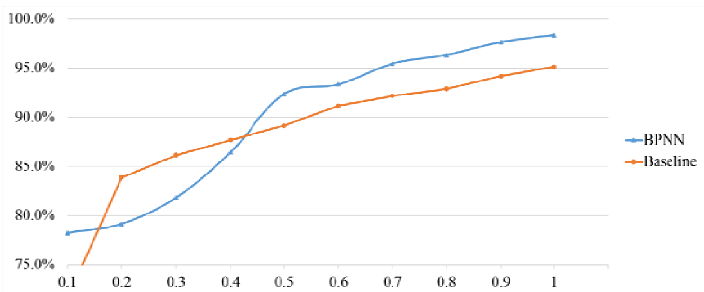

Figure 2. The Accuracy of Dynamic Prediction

Figure 2 shows the prediction accuracy of BPNN with both static and dynamic attributes, along with the baseline of Etter et al. (2013). It is obvious that the features used in the time spot = 0 is exclusive of dynamic ones and the result shows that our prediction model outperforms the baseline. The performance of BPNN is not so good at first four steps, and the part of the explanations for it is that the duration in Dreamore is generally short. However, what should be noted that there is no social attributes in Dreamore, which is much different from the dataset of baseline, so there is no condition for us to testify the impacts of social factors.

Conclusions

This paper illustrates a further study on crowdfunding prediction yielding novel insights. At the same time, the other improvement we have done is integrating our prediction with dynamic data. We use both static and dynamic data to generate time-series sequences as the initial inputs of prediction model and our empirical results showwell, which has the great practical significance to potential investors that they can know whether a crowdfunding project is successful or not at any time. Furthermore, our investigation extend the study on herding behaviors via our dynamic prediction via the phenomenon of relative sharp raise of prediction accuracy, which has been discussed in detailed in our empirical results. This finding also throws a light on the future work on herding behaviors of crowdfunding.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China (No. 91224008), Humanities and Social Sciences Foundation of the Ministry of Education (No. 14YJA630075), Beijing Nova Program (No. Z131101000413058).

References

Belleflamme, P., Lambert, T., and Schwienbacher, A. 2014. "Crowdfunding: Tapping the Right Crowd," Journal of Business Venturing (29:5), pp. 585-609.

Etter, V., Grossglauser, M., and Thiran, P. 2013. "Launch Hard or Go Home!,"), pp. 177-182.

Fisk, R. P., Ordanini, A., Miceli, L., Pizzetti, M., and Parasuraman, A. 2011. "Crowd - Funding: Transforming Customers into Investors through Innovative Service Platforms," Journal of Service Management (22:4), pp. 443-470.

Greenberg M. D., Pardo B., Hariharan K., Gerber E., and Pardo B. 2013. "Crowdfunding support tools: Predicting success & failure," CHI'13 Extended Abstracts on Human Factors in Computing Systems. ACM, pp. 1815-1820.

Mollick, E. 2014. "The Dynamics of Crowdfunding: An Exploratory Study," Journal of Business Venturing (29:1), pp. 1-16.

Sørensen, J. B., and Fassiotto, M. A. 2011. "Organizations as Fonts of Entrepreneurship," Organization Science (22:5), pp. 1322-1331.

Stam, W., and Elfring, T. 2008. "Entrepreneurial Orientation and New Venture Performance: The Moderating Role of Intra- and Extraindustry Social Capital," Academy of Management Journal (51:1), pp. 97-111.