The determinants of reward-based crowdfunding project delivery performance: A configurational model based on Latent Dirichlet Allocation

Yijing Wang¹*, Yi Feng¹ and Junhao Hu¹

¹School of Management and Economics, University of Electronic Science and Technology of China, Chengdu, Sichuan, 610054, China

*Corresponding author’s e-mail: 15528432779@163.com

Abstract. Reward-based crowdfunding (BRCF) projects as the most popular crowdfunding model, the funders expect the project owners to deliver the promised rewards within the specified delivery time. Previous studies examined various factors such as level and amount of funding received, project category and size that might influence the reward delivery performance. However, textual information of projects has rarely been studied for analyzing on time or late delivery. The main contribution of our research is applying text analytical framework that can extract latent semantics from the textual descriptions of projects to predict the reward delivery performance. More specifically, we use the Latent Dirichlet Allocation (LDA) topic model for effective extraction of topical features to form a configuration model, and Qualitative Comparative Analysis (QCA) was conducted for the antecedent factors. The findings indicate that the configurational model has high consistency and coverage, which indicate sufficient antecedent conditions for on-time and late delivery. The asymmetric analysis of this study also supports the application of complexity theory in the field of crowdfunding, which offers insights into the setting of appropriate antecedent conditions to achieve on-time delivery. The managerial insights are also discussed in the paper.

1. Introduction
Crowdfunding is a novel Internet-based financial model which raises project funds from netizens through the form of group purchase and pre-purchase. There are four main crowdfunding models such as donation-based, loan-based, equity-based, and reward-based [2]. This study focuses on reward-based crowdfunding in which the sponsors get non-monetary returns for their investments, such as products service, or even a thank-you letter from the entrepreneurs for their investments as a reward for backing a project [3].

In the previous studies, the choice of an operational approach needs to consider essential project characteristics which are deemed to influence delivery performance, though some studies explored the impact of numerical features (e.g., funding amount, duration of project, etc.) on crowdfunding delivery performance [2,9-11], none of the previous studies examined the influence of topical features (i.e., latent semantics) mined from textual descriptions of projects on fund raising success.

As a contribution to the research gap, this research applies latent Dirichlet allocation (LDA) [1] extracted a set of semantically related terms. Based on an unsupervised Bayesian learning algorithm, LDA can capture the latent topics from the project risk and challenge description. A topic-based text analytic method is different from the traditional key word-based approach in that a topic consists of a set of semantically coherent words, whereas the keyword-based method assumes the independence
among words [12,13]. We propose applying Jensen-Shannon divergence (JSD) (Lin, 1991) for topic dissimilarity, the perplexity measure [1] for the generalization performance of the model to get a proper number of topics and more interpretable results. Moreover, one novelty of our research is to combine topical features with common numerical features and apply a fuzzy set-Qualitative Comparative Analysis (fs-QCA) to explore the configurations of projects delivery performance [15].

2. Research framework and Method

2.1. Research framework
In this section, we describe the research framework, the LDA and the fs-QCA methods applied in the paper. The framework is shown in Fig. 1, which summarizes the processes and steps of how we extract the latent topics from project descriptions data and use fs-QCA to analyze them.

![Fig.1. A text analytics framework for crowdfunding analysis.](image)

2.2. LDA method
Latent Dirichlet allocation (LDA) is a three-layer Bayesian probability model which utilizes two Dirichlet distributions, namely topic-word distribution and document-topic distribution to describe the generation of words in documents. Essentially, each document is considered as a mixture of some topics and each topic is comprising a mixture of words. By following Blei et al. (2003), the graphical model representation of LDA is shown in Fig. 2. The boxes are “plates” representing replicates. The outer plate represents description documents, while the inner plate represents the repeated choice of topics and words within a description. The description-topic distribution is denoted as $\theta$, and each is drawn independently from a symmetric Dirichlet prior $\alpha$. The topic-word distribution is denoted as $\phi$, and each is drawn from a symmetric Dirichlet prior $\beta$.

LDA assumes the following generative process for each description in a corpus $D$ [1]:
Step 1. Choose $N \sim \text{Poisson}(\xi)$.
Step 2. Choose a topic distribution $\theta_d \sim \text{Dir}(\alpha)$.
Step 3. For each of the $N$ words $w_n$:
   a. Choose a topic $z_n \sim \text{Multinomial}(\theta)$.
   b. Choose a word $w_n$ from $p(w_n|z_n, \beta)$, a multinomial probability based on topic $z_n$. 
With LDA, we can identify a reasonable number of topics, label the topics and find the antecedent conditions based on the topic features.

Since LDA is a cluster model, we employ the Jensen-Shannon divergence (JSD), which is a symmetrized and smoothed version of the Kullback-Leibler divergence [4], to solve the topics dissimilarity and maximize the distance between every pair of topic-term distributions. The general definition of the JSD is as follows:

$$\text{JSD}_{\pi_1, \pi_2, \ldots, \pi_n} = H\left(\sum_{i=1}^{n} \pi_i p_i\right) - \sum_{i=1}^{n} \pi_i H(p_i)$$

where $\pi_1, \pi_2, \ldots, \pi_n$ are weights that are selected for the probability distributions $P_1, \ldots, P_n$, and $H(P)$ is the Shannon entropy for distribution $P$, $\pi_i$ is set to $1/n$ and $P_n$ is the topic-word distribution of topic $n$.

At the same time, the perplexity measure is used to avoid the over-clustering problem. The perplexity is formulated as follows:

$$\text{Perplexity} = \exp\left\{\frac{\sum_{d=1}^{M} \log p(w_d)}{\sum_{d=1}^{M} N_d}\right\}$$

where $N_d$ denotes the total number of terms occurred in the $d_{th}$ document. A lower perplexity score means a higher prediction power of the model.

2.3. $fs$-QCA method

Fuzzy sets QCA, currently developed by Ragin [15], prolonging and expanding the logic of QCA, allows the researcher to analyse fuzzy variables, which can be defined in terms of degree of membership in a well-defined set. Based on the equifinality concept, fuzzy set-QCA configurational models establish the antecedent conditions that represent different causal paths to achieve an outcome, and the paths leading to another outcome are unique and not the mirror opposite of ones leading to a different outcome. Further, every single cause has its own separate, independent impact on the outcome which means that one condition cannot solely produce an outcome. As a case-oriented approach, fuzzy-set QCA examines relationships between antecedent conditions and outcomes and uses coverage and consistency to evaluate the results.

3. Data collection and Text pre-processing

Our crowdfunding dataset is collecte from the Kickstarter crowdfunding platform (https://www.kickstarter.com) from 2012 to 2018. We use Random.org to generate numbers for the random selection of projects from the categories of 31,102 successful projects.

Following prior studies [16,18], text pre-processing in this study includes word text tokenization, converting words to lower-case, removing punctuation characters and numbers, removing stop words...
and omitting words with a length below a certain minimum (3 in our experiments) Beyond that, we filter words that occurred in less than three description texts. We use the Stanford Topic Modelling Toolbox version 0.4.0 [21] to implement text pre-processing. After cleaning the data for erroneousness and insufficient facts, 690 projects remained for this experiment analysis, which is considered an adequate sample size.

Table 1. Distribution of the project categories in the sampled crowdfunding projects.

| Project category    | Frequency | Percentage (%) |
|---------------------|-----------|----------------|
| Product design      | 198       | 28.7           |
| Technology product  | 156       | 22.6           |
| Publishing          | 69        | 10.0           |
| Film & video        | 48        | 7.0            |
| Fashion             | 42        | 6.1            |
| Craft               | 65        | 9.4            |
| Art                 | 59        | 8.6            |
| Game                | 53        | 7.7            |
| Total               | 690       | 100            |

4. Experiment

4.1. Topical feature selection

We compute the JSD scores and implement 5-fold cross validation for the perplexity scores of the datasets. The results are respectively showed in Fig. 3. To avoid over-clustering and improve clustering results, we set the number of topics as 10.

Fig.3 JSD scores and Perplexity scores of project description datasets

We apply LDA model to extract and label the topics of project description. Following previous studies [17,18], we name each topic based on the logical connection between their top words and the corresponding relative weights. For instance, based on the word “years” that is weighted 10.9%, “designers” that is weighted 9.3%, and “products” that is weighted 5.5%, this topic named as team experience. The final results of the features extracted from the topic are shown in table 2. The features are set as antecedent conditions for fs-QCA analysis.
Table 2. Extracted features from topics with their top ten words

| topic name          | top words                                                                 | Extracted features               |
|---------------------|----------------------------------------------------------------------------|----------------------------------|
| team experience     | designers, ears, products, development, experience, partner, technology, software, industry, engineers | Project team experience          |
| lead time           | timeline, schedule, confident, delays, keep, deliver, unforeseen, arise, informed, ability |                                  |
| delivery plan       | pledge, refund, fees, order, rewards, information, ship, address, duties, policy | Promised leading time            |
| supply chain        | chain, fulfillment, supply, things, delays, longer, happen, issues, orders, components |                                  |
| project difficulty  | possible, making, risk, deliver, goal, soon, done, before, ready, hands     |                                  |
| production          | manufacturer, prototypes, ensure, products, testing, components, technology, suppliers, demand, ready | Logistics (sourcing approach production approach) |
| sourcing plan       | material, delays, delivered, only, source, campaigns, previous, fulfillment, successfully, rewards |                                  |
| goods               | handwork, printing, take, tools, need, print, done, work, risk, before       |                                  |
| Delivery situation  | challenges, many, over, successful, help, deliver, work, possible, overcome, carefully | General description             |
| funders             | people, watch, support, directly, raise, launch, life, risk, challenges, future |                                  |

4.2. The fs-QCA analysis

According to fs-QCA requirements, before structural analysis, the qualitative information should be transferred to quantitative values. Hence, antecedent conditions and outcome variables in the sample should be calibrated to produce values ranging from 0 to 1 [24]. The calculations of the calibration for the membership scores were performed using the fuzzy set score in fs-QCA software. After the variables were calibrated, the data built a “Truth Table” to indicate all possible configurations of the different antecedent conditions, which were thought to affect delivery performance [19]. Using the truth table as the starting point for analysis, fs-QCA software can use Boolean algebra's operational logic to get three types of solutions: complex connections, intermediate solutions, and simple solutions. This paper uses fsQCA3.0 software to analyze and obtain the corresponding configuration (interpretation based on the intermediate solution).

The QCA configurational model explains the combinations of the antecedent conditions for delivery performance. The high coverage value indicates that a configuration model explains most of on-time or late delivery performance [20]. In Tables 3 and 4, all the configuration models have consistency values exceeding 0.74 and coverage not exceeding 0.65, which means the antecedents forming the configurations are sufficient conditions to achieve on-time and late delivery performance outcomes [24].

Table 3. Models for on-time delivery performance.

| Configuration | Raw coverage | Unique coverage | Consistency |
|---------------|--------------|-----------------|-------------|
| PLT~PS*PCS~IHP*OSP~PTE | 0.4012 | 0.1053 | 0.8750 |
| ~PLT*PS*PCS*IHP*OSP*PTE | 0.3609 | 0.0902 | 0.8727 |
| ~PLT*PS*PCS*IHP~OSP*PTE | 0.2782 | 0.0075 | 0.8409 |
| PLT*PS*PCS*IHP~OSP*PTE | 0.3008 | 0.0301 | 0.8000 |

solution coverage: 0.703759
solution consistency: 0.87013
Table 4. Models for late delivery performance.

| Model                  | Raw coverage | Unique coverage | Consistency |
|------------------------|--------------|-----------------|-------------|
| PLT*~PS*PCS*~IHP*~OSP*--PTE | 0.4190       | 0.0419          | 0.9375      |
| PLT*PS*--PCS*IHP*--OSP*--PTE | 0.2542       | 0.0279          | 0.8922      |
| PLT*PS*--PCS*--IHP*OSP*PTE  | 0.3631       | 0.1117          | 0.9220      |
| ~PLT*--PS*PCS*--IHP*OSP*PTE | 0.4106       | 0.0503          | 0.8571      |

solution coverage: 0.622905
solution consistency: 0.888446

5. Analysis and Conclusion

To explore the delivery performance of crowdfunding projects, we use an asymmetric analysis method. The models in Table 3 consist of different paths with unique approaches to achieve on-time delivery. Model one offers some insight into a project team without experience when they adopt post-campaign sourcing. It is essential to set long promised lead times and use the outsourcing production to support on-time delivery. Model two offers an understanding that an inexperienced team with both in-house and outsourcing production approach can reduce the adverse effect of a short lead time. This model offers flexibility in an operational approach toward on-time delivery. Model three and model four have a significant implication to crowdfunding projects with experienced teams. The positive effect of proactive sourcing and in-house production would reduce the adverse effect of promised leading time.

The findings on the paths to late delivery performance differ from previous crowdfunding research findings on the causes of late delivery [2,16,17]. The four models in table 6 suggest that the absence of appropriate production approach among the antecedent conditions lead to late delivery. Specifically, a project team without experience decides to produce all rewards in-house but do not determined the sourcing approach before the start of the campaign would deliver late. These findings are in line with the outcome of the interview reported in CNNMoney that manufacturing challenges are among the primary causes of a project's late delivery [14].

The high consistency and coverage values of the models make the antecedent conditions examined and verified. Through the analysis, project founders should embrace proactive sourcing and set an appropriate lead time to ensure on time delivery in the implementation phase. Conversely, when post-campaign sourcing adopted, founders can consider a combination of multiple production methods to take advantage of cost minimization. Furthermore, Crowdfunding platforms should encourage project founders to make adequate preparation of the sourcing and production before the funding phase begins. Advanced preparation may promote a better delivery performance and further boost crowd confidence in crowdfunding project.

Acknowledgement

The authors are partially supported by Youth Team Program for Technology Innovation of Sichuan Province (No. 2016TD0013).

References

[1] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. (2003) Latent dirichlet allocation. Journal of machine Learning research., 3: 993-1022.
[2] E. Mollick. (2014) The dynamics of crowdfunding: an exploratory study, Journal of Business Venturing., 29: 1–16
[3] J.-A. Koch, M. Siering. (2015) Crowdfunding success factors: the characteristics of successfully funded projects on crowdfunding platforms, Proc. Twenty-Third Eur. Conf, Inf. Syst., pp.1–15.
[4] Burnham, Kenneth P., and David R. Anderson. (2001) Kullback-Leibler information as a basis for strong inference in ecological studies. Wildlife research., 28.2: 111-119.
[5] C. Sascha Kraus, A. Richter, C.-F. Brem, M.-L. Chan Cheng. (2016) Strategies for reward-based crowdfunding campaigns, Journal of Innovation & Knowledge., 1: 13–23.
[6] H. Yuan, R.Y.K. Lau, W. Xu. (2016) The determinants of crowdfunding success: a semantic text analytics approach, Decision Support Systems., 91: 67–76.
[7] A. Lukkarinen, J.E. Teich, H. Wallenius, J. Wallenius. (2016) Success drivers of online equity crowdfunding campaigns, Decision Support Systems., 87: 26–38.
[8] A. Cordova, J. Dolci, G. Gianfrate. (2015) The determinants of crowdfunding success: evidence from technology projects, Procedia-Social and Behavioral Sciences., 181: 115–124.
[9] J.A. Hauge, S. Chimahusky. (2016) Are promises meaningless in an uncertain crowd-unding environment, Economic Inquiry., 54: 1621–1630.
[10] Mollick, Ethan R, and V. Kuppuswamy. (2014) After the Campaign: Outcomes of Crowdfunding. SSRN Electronic Journal.
[11] S. Ryu, Y.G. Kim. (2016) A typology of crowdfunding sponsors: birds of a feather flock together? Electronic Commerce Research and Applications., 16: 43–54.
[12] Lau, Raymond YK, Yunqing Xia, and Yunming Ye. (2014) A probabilistic generative model for mining cybercriminal networks from online social media. IEEE Computational intelligence magazine., 9.1: 31-43.
[13] Wang, Tao, et al. (2014) Product aspect extraction supervised with online domain knowledge. Knowledge-Based Systems., 71: 86-100.
[14] Regin, Charles C. (2009) Redesigning social inquiry: Fuzzy sets and beyond. University of Chicago Press.
[15] Braumoeller, Bear. Fuzzy-Set Social Science. By Charles C. Regin. (University of Chicago Press., 2000.). Journal of Politics., 70.1: 291-292.
[16] Tirunillai, Seshadri, and G. J. Tellis. (2014) Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation. Journal of Marketing Research (JMR)., 51.4: 463-479.
[17] Jun, Liu, W. Quzhen, and M. Qingguo. (2011) The effects of project uncertainty and risk management on IS development project performance: A vendor perspective. International Journal of Project Management., 29.7: 923-933.
[18] Guo, Yue, S. J. Barnes, and Q. Jia. (2017) Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation. Tourism Management., 59: 467-483.
[19] Rihoux, B.C. Regin. (2009) Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques.
[20] Ramage, Daniel. (2011) and Evan Rosen. Stanford topic modeling toolbox.