Modeling Menu Bundle Designs of Crowdfunding Projects

Yusan Lin
The Pennsylvania State University
University Park, PA 16802, USA
yusan@psu.edu

Peifeng Yin
IBM Almaden Research
San Jose, CA 95120, USA
peifengy@us.ibm.com

Wang-Chien Lee
The Pennsylvania State University
University Park, PA 16802, USA
wlee@cse.psu.edu

ABSTRACT

Offering products in the forms of menu bundles is a common practice in marketing to attract customers and maximize revenues. In crowdfunding platforms such as Kickstarter, rewards also play an important part in influencing project success. Designing rewards consisting of the appropriate items is a challenging yet crucial task for the project creators. However, prior research has not considered the strategies project creators take to offer and bundle the rewards, making it hard to study the impact of reward designs on project success. In this paper, we raise a novel research question: understanding project creators' decisions of reward designs to level their chance to succeed. We approach this by modeling the design behavior of project creators, and identifying the behaviors that lead to project success. We propose a probabilistic generative model, Menu-Offering-Bundle (MOB) model, to capture the *offering* and *bundling* decisions of project creators based on collected data of 14K crowdfunding projects and their 149K reward bundles across a half-year period. Our proposed model is shown to capture the offering and bundling topics, outperform the baselines in predicting reward designs. We also find that the learned offering and bundling topics carry distinguishable meanings and provide insights of key factors on project success.

CCS CONCEPTS

• **Information systems → Social recommendation**; **Web applications**; • **Applied computing → Electronic commerce**; **Marketing**;

KEYWORDS

menu bundle; crowdfunding; offering decision; bundling decision

1 INTRODUCTION

Crowdfunding has emerged as a popular means for entrepreneurs to pledge funding for their creative projects. Among various types of crowdfunding platforms, the reward-based one is considered as the most popular [3]. In such a platform, e.g., Kickstarter, project *creators* establish a funding request including i) the description of their ongoing project, ii) the amount of funding needed (*pledging goal*) and iii) the expiration date.[1] Investors (also known as *backers*) choose to contribute a particular amount of money according to the *reward menu*. In return, corresponding rewards will be delivered, either during the pledging or after the pledging, to the backers. The funding process is successful if the total contributions exceed the preset pledging goal before expiration date.

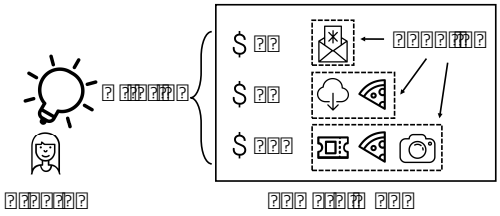


Figure 1: Illustration of menu bundle design process.

One natural and vital question to crowdfunding is what factors may affect the fund-raising result, i.e., success or failure. To answer this question, previous works have studied several aspects, such as project description [16], promotion [10], interaction between creators and potential backers [18], and so on. However, few works study the impact of rewards [9]. A reward item is a tangible or virtual item offered by creators in exchange for backers' contribution. Usually various items are *bundled* as reward packages for backers to make different contributions. Our analysis (detailed in Section 3) reveals that the way rewards are designed and bundled is crucial in affecting the success of a project funding. For example, we find that projects that offer more rewards are more likely to succeed, so as projects that offer rewards in forms of bundles.

A plausible reward design process is illustrated in Figure 1. When designing the menu bundles, a project creator first decides how she wants to offer the rewards by determining the number of rewards and prices of each. Next, for each reward on the menu, the creator designs how the reward items are bundled. In this work, we aim to model the process of reward design, also known as *menu bundle design* in [2]. Essentially there are two decisions to be considered, namely *offering* and *bundling*, as shown in Figure 1. The former determines different contribution levels (reward prices) while the latter decides the combination varieties of corresponding rewards. With understanding of the behaviors of project creators in these two design decisions, we can further establish their correlations with funding success and even identify good menu bundle design.

Modeling bundle menu design decisions is a new and challenging research.[2] Design behaviors in both offering and bundling decisions are potentially affected by many factors. For example, a music project of a rock band's new CD, Sweet Little Bitter by Bad Reed, sets a relatively low price for each contribution level due to their young

[1] https://www.kickstarter.com/

[2] To the best knowledge of the authors, there is a lack of study reported in the literature investigating this problem.

CIKM'17 , November 6–10, 2017, Singapore, Singapore

ACM ISBN 978-1-4503-4918-5/17/11…$15.00
https://doi.org/10.1145/3132847.3132958

fans' limited affordability.[3] As for a technology project developing a new model of light-weight laptop, CruxSKUNK, it is unlikely to offer rewards at low price due to the nature of the product.[4] Consider these two projects again. The music project offers a variety of rewards including t-shirt with the band's logo, CD in digital, physical, and vinyl forms, and a personal concert, etc., all of which are highly correlated with the product. For the technology project, there is a smaller variety of rewards since the laptop is the sole focus.

In preparation for the modeling of menu bundle design process, we conduct a series of data analyses to unveil the important insights regarding the aforementioned two design decisions. Then based on the analysis result, we develop a probabilistic generative model, namely, *Menu-Offering-Bundle model (MOB)*, that captures the offering and bundling design decisions. In MOB, both offering and bundling decisions are represented as latent topics, which generate the reward prices and words modeled as observations. Therefore, different menu bundle designs are abstract as vectors of latent topic distribution. In this way, patterns of menu bundle design can be learned. A comprehensive evaluation shows our method significantly outperforms baseline models when making predictions for reward menu creation. Furthermore, with learned MOB, we explore the correlation of menu bundle design and the project success. It is found that different offering and bundling strategies need to be properly combined to increase the chance of success. Finally, we study successful and unsuccessful projects on basis of learned MOB and discuss key insights.

The contributions andfi ndings of this paper are summarized as below.

(1) We raise a novel research question on how creators make decisions in terms of *offering* and *bundling* in the process of *reward menu bundle design* for reward-based crowdfunding projects. A series of analyses on real dataset is conducted to reveal important factors that support the design decisions for offering and bundling.
(2) We develop a probabilistic generative model, namely, *Menu-Offering-Bundle model (MOB)*, to model the project creators' design process of reward bundles, by capturing the offering and bundling design decisions. With MOB, we provide insights of what type of offering and bundling decisions are more likely to lead to project success.
(3) We leverage MOB as a predictor to predict reward designs for project creators. Through a comprehensive evaluation, the MOB model is shown to outperform LDA by 29.3% when predicting reward prices and 23.5% when predicting reward words in terms of F1-score. With the model, we then study the offering and bundling topics learned by MOB, and examine successful and unsuccessful projects based on their offering and bundling decision.

In the rest of this paper, wefi rst review related works in Section 2, including those in crowdfunding and probabilistic generative models. We then analyze a real dataset of reward bundles collected from Kickstarter, one of the most popular reward-based crowdfunding platform, in Section 3 and propose the MOB model that captures the offering and bundling topics of project creators in Section 4. We evaluate the MOB model and two baselines through some prediction tasks in Section 5. In Section 6, we analyze the learned offering and bundling design topics, study their underlying meanings, use them to identify great strategy of reward design, and further study selected projects to understand their reasons of success and failure. Finally, we conclude this paper and discuss future work in Section 7.

[3]https://www.kickstarter.com/projects/badreedband/sweet-little-bitter-by-bad-reed
[4]https://www.kickstarter.com/projects/spywire/cruxskunktm-powerful-ipad-laptop

2 RELATED WORK

In this section, we review related works in crowdfunding and probabilistic generative models.

2.1 Crowdfunding

Due to the booming interests and activities in crowdfunding, there have been growing research on crowdfunding in recent years. In general they can be classified into two categories: i) backer-oriented and ii) project-oriented. The backer-oriented research targets on supporting backers, and considers the crowdfunding platforms as a new environment for recommendation. Rakesh et al. propose a recommendation system to match the investors and projects using temporal, personal, location, and network features [14]. An et al. conduct a series of statistical tests to confirm multiple recommendation hypotheses, e.g., "a project with high pledging goal is likely to befi nanced by frequent investors" [1]. Besides recommendation, Shafqat et al. analyzes the language on Kickstarter to detect scams [16].

The project-oriented research mainly aims on prediction of project success with respect to different features. Greenberg et al. evaluate simple project features such as goal, category, whether connected to Twitter, and sentence counts [8]. Lu et al. investigate the effects of promotions via social networks [10]. Mitra and Desai et al. utilize the language used in Kickstarter projects [6, 12].

Among these early efforts, there has been a lack of work for supporting creators, e.g., how the designs of reward bundles can affect the project success, not to mention providing recommenders dedicated for their needs. While it is claimed that the prediction of project success using project features can help improve project design [8], no practical feedback is provided to the project creators. Our work in this paper aims tofi ll this gap. We propose a novel generative model for capturing the decisions in menu bundle design, and study the set of decisions that tend to lead to project success. Ourfi ndings about reward design can be well combined with previous works to achieve better project success.

2.2 Probabilistic Generative Models

Deerwester et al.fi rst propose latent semantic analysis (LSA), a probabilistic generative model (or aspect model) for capturing the topic of documents generating words [5]. Blei et al. later extend this work by introducing the plate structure, and propose latent Dirichlet allocation (LDA), a probabilistic generative model that captures not only the topics of documents, but also the topics of words used in the documents [4]. Afterwards, abundant generative models have been proposed based on LDA to apply on various scenarios and applications. The more recent works include: Ma et al. model the public opinions on urban affairs [11], Paul et al. model the health topics in social media [13], Zhu et al. model the emotion evolution on news [20], Yin et al. model users' mobile app

Table 1: Dataset statistics of the Kickstarter projects

	# of projects	# of rewards	# of backers
Successful	7,169	81,896	952,041
Unsuccessful	7,683	67,272	173,043
Total	14,852	149,168	1,125,084

download [19], and Trouleau et al. model people's binge watching behavior [17], just to name a few.

Probabilistic generative models are also applied in crowdfunding recently. Rakesh et al. develop a generative model that recommends Kickstarter projects to group of backers [15]. Gao et al. apply LDA to study the project descriptions and further extract features for project success prediction [7]. Xu et al. use LDA to model the project updates as well as its impact on project success [18]. However, none of these research attempts to capture the menu bundle designs, which is the gap we aim to bridge in this work.

3 REWARD BUNDLE ANALYSIS

In this section, we first describe the dataset used in this paper, and then conduct a series of data analyses to show the importance of reward menu bundle design in crowdfunding.

Kickstarter is currently the most popular reward-based crowdfunding platform. Since founded in April, 2009, Kickstarter has helped 103K project creators pledge over $2.29 billion, which has involved 10 million backers with in total 29 million backings.[5]

We collect data from January 1st, 2014 to June, 30th, 2014. Within this period, we have 14,852 projects and 149,168 reward bundles. We process the textual description of the reward bundles by converting all letters to lowercase, stemming the words using Porter stemmer, and removing stop words. We also remove words that occur less than 10 times. This gives us 11,483 unique words. Detailed statistics are summarized in Table 1.

Here we analyze the collected data to cast light on how the reward menu bundle design affects project success from the following aspects: number of reward bundles, the bundling of rewards, number of reward items, and project categories.

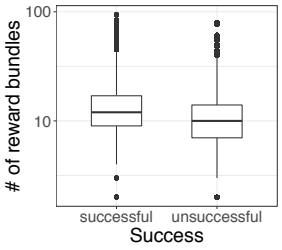


Figure 2: Number of reward bundles v.s. project success.

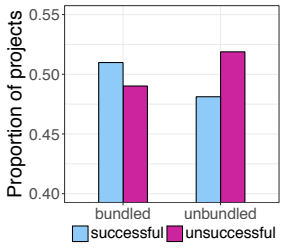


Figure 3: Projects offering bundles v.s. success rates.

Number of reward bundles: Project creators need to consider the offering in menu bundle design. A factor naturally arises is *the number of bundles to offer*. Does the number of reward bundles affect the success of projects? As shown in Figure 2, on average the projects successful in achieving pledging goal generally have more reward bundles than unsuccessful projects.

[5]https://www.kickstarter.com/help/stats

Bundling of rewards: In reward-based crowdfunding platform, it is a common scenario that multiple items are bundled as a reward package. We explore how the concept of *bundling* helps to attract backers. As shown in Figure 3, projects with bundled rewards generally have a higher success rate than those with unbundled rewards.[6]

Number of reward items: Within a reward bundle, we study how the number of included items would affect the number of attracted backers.[7] As shown in Figure 4, it is a negative correlation. In general, the more items included in a reward, the less backers support that reward. This is reasonable because most people have a limited budget and the price increases as the bundled reward grows, shown in Figure 5. We further examine for reward bundles offered at the same price, whether bundles including more items are supported by more backers. As Figure 6 shows, when conditioned on the prices $25, $200, and $1000, reward bundles with more items tend to attract more backers.

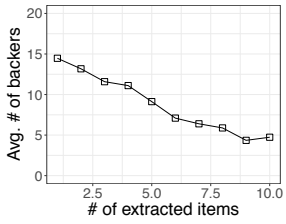


Figure 4: Number of items v.s. number of backers.

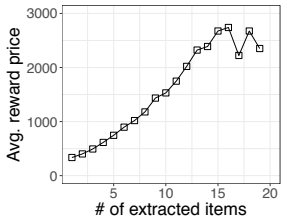
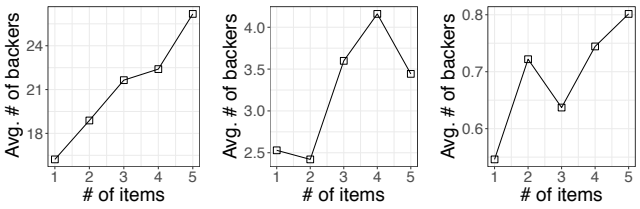


Figure 5: Number of items v.s. reward price.



(a) Price $25 (b) Price $200 (c) Price $1000

Figure 6: Number of items v.s. number of backers at different prices.

Project categories: Taking a closer look at the difference of reward bundles among project categories, we plot the project categories, their average offering price divided by project goal and average number of backers per reward bundle divided by number of backers per project in Figure 7. There are two observations. Firstly, in general, a higher price attracts a smaller number of backers. This is confirmed in earlier analysis (Figure 4 and Figure 5) as most people cannot afford expensive backings. A second observation is that price differs a lot among different project categories. For example, reward bundles of journalism are often offered at a lower price,

[6]Rewards are considered as unbundled if only one item is included, as bundled if two or more items are included.

[7]We leverage reward items extracted by BILOU labeling scheme in Named Entity Recognition (NER). Number of reward items is determined by the reward items extracted.

Figure 7: Project categories v.s. the average offering prices and number of backers per reward bundle.

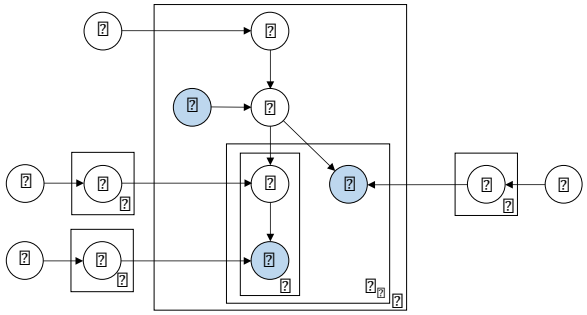


Figure 8: Menu-Offering-Bundle (MOB) model.

and can attract many backers; reward bundles of game projects are often offered at a higher price, and hence attract less backers.

With the above analyses, we can see the effect different designs of rewards may have on the projects' pledging results. To properly capture design decisions behind menu bundle design, we propose a probabilistic generative model in the next section.

4 MENU-OFFERING-BUNDLE MODEL

In this section, we describe the details of our proposed probabilistic generative model, Menu-Offering-Bundle model (MOB), that captures the previously introduced offering and bundling design decisions. Before moving forward, we list the definition of each symbol in Table 2 for clarity.

4.1 Modeling Offering and Bundling Behaviors

We propose the Menu-Offering-Bundle (MOB) model to capture the menu bundle design process involving the decisions on offering and bundling, as introduced in Section 1. In MOB, both design decisions are modeled as latent topics, namely *offering topic* and *bundling topic*. When an offering topic is decided based on the creator's personal preference over the offering topics and the category of project, the creator plans out the number of reward bundles as well as price for each of them. With fixed reward price, a bundling topic is selected to determine what word should be included to describe the reward. In this process, the selected offering topic will affect the choice of bundling topics. This design comes from the observation that reward words are associated with price, as shown in Figure 5. Furthermore, our analysis finds that the median reward prices of each project are significantly different across the project categories, with a p-value < 0.01 by conducting an ANOVA test. Thus we add the project category as an additional observed variable that affects the selection of offering topic.

Figure 8 shows the graphic representation of MOB, where the observed variables are shaded. As shown, for each project p, an offering topic x is chosen from a distribution over offering topics (θ) based on its category q. The offering topic x represents p's design behavior of offering bundles in the proposed project, which further generates the project's bundles' prices c (chosen from a distribution over price ψ), and the p's bundling topic y. For each bundle offered by p, with a price c generated earlier by x, and based on the bundling topic y, a word w describing the reward is generated (chosen from a distribution over words ϕ).

Table 2: Symbol definitions of MOB

Symbol	Description
P	set of all projects
W	set of all reward items
B_p	the bundles offered by project p
θ	$(M \times S) \times K$ matrix, distribution of projects over offering topics
ψ	$K \times V$ matrix, distribution of offering topics over prices
λ	$K \times L$ matrix, distribution of offering topics over bundling topics
ϕ	$L \times T$ matrix, distribution of bundling topics over words,
α	prior of θ
β	prior of ψ
γ	prior of λ
η	prior of ϕ
x	offering topic variable
y	bundling topic variable
q	project category variable
c	bundle price variable
w	word variable
k	index for offering topics
l	index for bundling topics
s	index for project categories
m	index for projects
n	index for bundles
t	index for words
v	index for prices

The joint probability of the observed and latent variables, given the distribution priors, can be written as follows.

$$Pr(\vec{w}, \vec{c}, \vec{q}, \vec{x}, \vec{y} | \alpha, \beta, \gamma, \eta)$$
$$= \prod_{p}^{P} Pr(x|\alpha, q) \prod_{b}^{B_p} Pr(c|x, \beta) \prod_{w}^{W} Pr(y|x, \gamma) Pr(w|y, \eta) \quad (1)$$

As shown in Figure 8, the parameters in the model are θ, ψ, λ, and ϕ, which we design their distributions to be generated from Dirichlet distributions given the priors, and the corresponding latent variables to be drawn from multinomial distributions given the parameters. The dimensions of the parameters are specified in Table 2.

The likelihood of the model is expressed as follows.

$$
\begin{aligned}
&Pr(\vec{w}_m, \vec{c}_m, \vec{q}_m | \vec{\alpha}, \vec{\beta}, \vec{\gamma}, \vec{\eta}) \\
&= \int_{\theta} \int_{\psi} \int_{\lambda} \int_{\phi} Pr(\vec{\phi}_{y_{n,t}} | \vec{\eta}) Pr(\vec{\lambda}_{m,n} | \vec{\gamma}) Pr(\vec{\psi}_{m,n} | \vec{\beta}) Pr(\vec{\theta}_m | \vec{\alpha}) \\
&\prod_{t=1}^{W} \sum_{x_m} \sum_{y_{m,n}} Pr(w_{n,t} | \vec{\phi}_{y_{n,t}}) Pr(y_{n,t} | \vec{\lambda}_{m,n}) Pr(c_{m,n} | \vec{\psi}_{m,n}) \\
&\qquad Pr(x_m | q_m, \vec{\theta}_m) \, d\vec{\theta}_m \, d\underline{\Psi} \, d\underline{\Phi}
\end{aligned} \tag{2}
$$

which we will use to search for the best number of x and y.

In summary, for a project p, the generative process of MOB is as follows.

(1) Draw an offering topic x from the multinomial distribution given θ and its category q
(2) For each bundle b offered by p:
  (a) Draw a price c from the multinomial distribution ψ given the previously chosen x
  (b) For each word in the vocabulary:
    (i) Draw a bundling topic y from the multinomial distribution λ given the previously chosen offering topic x
    (ii) Draw a word w from the multinomial distribution ϕ given the chosen bundling topic y

4.2 Model Inference

Due to its simplicity and fast convergence, we adopt Gibbs sampling to infer the parameters. Using Gibbs sampling, we construct a Markov chain, converging to the posterior distribution on the latent variables, x and y. The idea is to repeatedly draw x and y from their distribution conditioned on the rest of the variables, integrating out the parameters, θ, ψ, λ, and ϕ. To do so, wefi rst calculate the posterior of x as follows.

$$
\begin{aligned}
&Pr(x_i = k | \vec{x}_{\neg i}, \vec{y}, \vec{q}, \vec{c}, \vec{w}) \\
=&Pr(x_i = k | \vec{x}_{\neg i}, \vec{y}, \vec{q}, \vec{c}) \\
&\frac{Pr(\vec{x}|\vec{q})}{Pr(\vec{x}_{\neg i}|\vec{q}_{\neg i}) Pr(q_i)} \cdot \frac{Pr(\vec{c}|\vec{x})}{Pr(\vec{c}_{\neg i}|\vec{x}_{\neg i}) Pr(\vec{c}_i)} \cdot \frac{Pr(\vec{y}|\vec{x})}{Pr(\vec{y}|\vec{x}_{\neg i})} \\
&\frac{n^{(k)}_{m,s,\neg i} + \alpha_k}{\sum_{k=1}^{K} n^{(k)}_{m,s,\neg i}} \cdot \frac{n^{(v)}_{k,\neg i} + \beta_v}{\sum_{v}^{V} n^{(v)}_{t,\neg i} + \beta_v} \cdot \frac{n^{(l)}_{k,\neg i} + \gamma_l}{\sum_{l}^{L} n^{(l)}_{k,\neg i} + \gamma_l}
\end{aligned} \tag{3}
$$

where $n^{(k)}_{m,s,\neg i}$ is the number of times project m of category s is observed with offering topic k, excluding the case of i; $n^{(v)}_{k,\neg i}$ is the number of times price is observed with offering topic k, excluding the case of i; $n^{(l)}_{k,\neg i}$ is the number of times bundling topic l is observed with offering topic k, excluding the case of i.

We then calculate the posterior of y as follows.

$$
\begin{aligned}
&Pr(y_j = l | \vec{y}_{\neg j}, \vec{x}, \vec{q}, \vec{c}, \vec{w}) \\
=&Pr(y_j = l | \vec{y}_{\neg j}, \vec{x}, \vec{w}) \\
&\frac{Pr(\vec{y}|\vec{x})}{Pr(\vec{y}_{\neg j}|\vec{x}_{\neg j}) Pr(y_j)} \cdot \frac{Pr(\vec{w}|\vec{y})}{Pr(\vec{w}_{\neg j}|\vec{y}_{\neg j}) Pr(\vec{y}_j)} \\
&\frac{n^{(l)}_{k,\neg j} + \gamma_l}{\sum_{l}^{L} n^{(l)}_{k,\neg j} + \gamma_l} \cdot \frac{n^{(t)}_{l,\neg j} + \eta_t}{\sum_{t}^{T} n^{(t)}_{l,\neg j} + \eta_t}
\end{aligned} \tag{4}
$$

where $n^{(l)}_{k,\neg j}$ is the number of times bundling topic l is observed with offering topic k, excluding the case of j; $n^{(t)}_{l,\neg j}$ is the number of times word t is observed with bundling topic l, excluding the case of j.

After convergence of the Gibbs sampling process, we can then estimate the parameters of the model as follows.

$$
\theta_{(m,s),k} = \frac{n^{(k)}_{m,s} + \alpha_k}{\sum_{k=1}^{K} n^{(k)}_{m,s} + \alpha_k} \tag{5}
$$

$$
\psi_{k,v} = \frac{n^{(v)}_{k} + \beta_v}{\sum_{v=1}^{V} n^{(v)}_{k} + \beta_v} \tag{6}
$$

$$
\lambda_{k,l} = \frac{n^{(l)}_{k} + \gamma_l}{\sum_{l=1}^{L} n^{(l)}_{k} + \gamma_l} \tag{7}
$$

$$
\phi_{l,t} = \frac{n^{(t)}_{l} + \eta_t}{\sum_{t=1}^{T} n^{(t)}_{l} + \eta_t} \tag{8}
$$

The process of learning MOB using Gibbs Sampling as the parameter learning strategy is summarized in Algorithm 1.

Algorithm 1 Learning MOB

Input: Projects P, bundles B, reward prices C, categories Q, reward words W, number of offering topics K, number of bundling topics L
Output: θ, λ, ψ, ϕ

```
Initialize θ, λ, ψ, φ
X  a |P| × K matrix
Y  a K × L matrix
while X and Y not converged do
    for i in P do                          ▷ Gibbs sampling process
        for k in K do
            X[i,k]  posterior computed by (3)
            for l in L do
                Y[k,l]  posterior computed by (4)
θ, λ, ψ, φ  update by (5), (6), (7) and (8), respectively
```

5 MODEL EVALUATION

MOB is designed to capture frequent patterns of offering and bundling design decisions. In this section, we evaluate this model via two tasks, i.e., prediction of reward prices and words. Two existing topic models are also included as benchmark for comparison.

Task 1: Reward price prediction. Ourfi rst task corresponds to thefi rst step in the menu bundle design process, offering decision, by predicting different prices a project creator uses to offer her reward bundles. To do this, we estimate the posterior of bundle prices given project category. Formally, given the i^{th} project, which is in the u^{th} category, and the words used to describe the reward

bundles $\vec{t}$, the probability for offering a bundle at the j^{th} price is

$$\begin{aligned}
&Pr(c = j|p = i, \theta, \psi, \lambda, \phi, \vec{t}) \\
=&\frac{Pr(c = j, \vec{t}|p = i, \theta, \psi, \lambda, \phi)}{Pr(\vec{t}|p = i, \theta, \psi, \lambda, \phi)} \\
&\sum_{k=1}^{K}\sum_{l=1}^{L}\sum_{t=1}^{\vec{t}} Pr(c = j|x = k, \psi)Pr(x = k|q = u, \theta, y = l)\cdot \\
&\qquad Pr(y = l|\lambda, w = t)Pr(w = t|\phi) \\
=&\sum_{k=1}^{K}\sum_{l=1}^{L}\sum_{t=1}^{\vec{t}} \psi_{k,v}\theta_{(i,s=u),k}\ \ _{k,l,s=u}\phi_{l,t} \qquad (9)
\end{aligned}$$

where is a $K \times L \times S$ matrix created by θ and λ to represent $Pr(x|q, y)$, and θ here is a $M \times K$ matrix with $s = u$. We omit the denominator $P(\vec{t}|...)$ because it is a constant independent of c.

Task 2: Reward word prediction. The second task corresponds to the second step in the menu bundle design process, bundling decision, by predicting the reward words the project creator uses to describe the reward bundles. To approach this task, we estimate the posterior of reward words given project category. Similar to Task 1, let $p = i$ denote the i^{th} project, which is in the u^{th} category, the probability for using the j^{th} reward word in the reward bundles is

$$\begin{aligned}
&Pr(w = j|p = i, \theta, \psi, \lambda, \phi) \\
=&\sum_{k=1}^{K}\sum_{l=1}^{L} Pr(w = s|y = l, \psi)Pr(y = l|x = k, \lambda)Pr(x = k|p = i, \theta) \\
=&\sum_{k=1}^{K}\sum_{l=1}^{L} \phi_{l,s}\lambda_{k,l}\theta_{(i,s=u),k} \qquad (10)
\end{aligned}$$

where θ is a $M \times K$ matrix with $s = u$.

To see how MOB performs against other models, we use two baseline probabilistic generative models, *latent semantic analysis (LSA)* and *latent Dirichlet allocation (LDA)*, as our benchmarks. The project-offering-reward history is denoted as $H = \ s, m, \ , n, t$, meaning the m^{th} project in the s^{th} category offers an n^{th} bundle at the $^{\text{th}}$ price, using the t^{th} word. In total, we have 134,169 records in H. We conduct a 10-fold cross validation of each model on H.

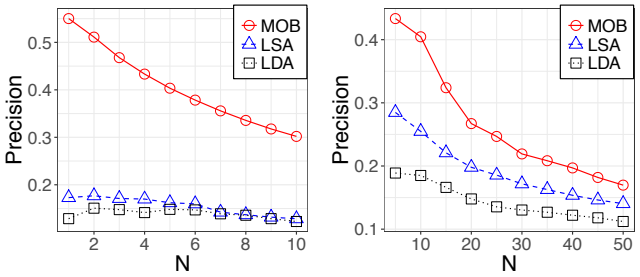
For Task 1, MOB takes in records in as $s, m, \ , n, t$, while LSA and LDA take in records as $m, \ $. The prediction is done by ranking the posterior probability of the m^{th} project offering rewards at the $^{\text{th}}$ price. For Task 2, MOB also takes in records as $s, m, \ , n, t$, while LSA and LDA take in records as m, t . The prediction is done by ranking the posterior probability of the m^{th} project using the t^{th} word to describe reward bundles. Both tasks are evaluated using the three performance metrics below.

$$precision@N = \frac{|\{top\ N\ recommendations\} \quad \{true\ items\}|}{|\{top\ N\ recommendations\}|} \qquad (11)$$

$$recall@N = \frac{|\{top\ N\ recommendations\} \quad \{true\ items\}|}{|\{true\ items\}|} \qquad (12)$$

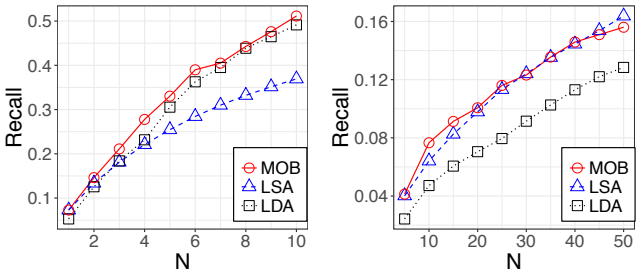
$$f1@N = 2 \times \frac{precision@N \times recall@N}{precision@N + recall@N} \qquad (13)$$

where N is the number of retrieved items, i.e., prices and words, from the recommendation list.



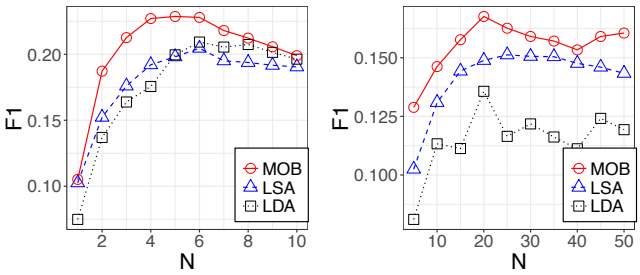
(a) Recommend prices (b) Recommend words

Figure 9: Precision of reward price and word predictions.



(a) Recommend prices (b) Recommend words

Figure 10: Recall of reward price and word predictions.



(a) Recommend prices (b) Recommend words

Figure 11: F1 of reward price and word predictions.

As shown in Figures 9, 10, and 11, when predicting reward prices to projects, MOB consistently outperforms the other two benchmarks throughout various N. Similarly, when predicting reward words to projects, MOB also outperforms the other two benchmarks. One thing to note regarding to the performance of reward price prediction is that, unlike predicting the words to describe reward bundles, it is not practical to predict with large N. In our dataset, less than 4% of projects offer more than 20 rewards. Also, while we can recommend words to a project that it already uses in the training dataset, it makes less sense to recommend prices that a project already uses. This is because in our dataset, most projects do not use duplicate prices when offering rewards. Even if they do, less than 10% of the projects reuse prices more than twice. These results further show the effectiveness of our proposed MOB model.

6 TOPICAL ANALYSIS

As an analytical tool, MOB mayfi nd frequent strategies of menu bundle design and the correlation between design strategies and project success, both qualitatively and quantitatively. Specifically, by empirical study, wefi nd that the maximum likelihood of MOB is achieved when the number of offering and bundling topics are set to 10. Therefore in the analysis, MOB is trained with $|x| = 10, |y| = 10$. Note that when learning MOB, the model has no knowledge of the final project success. However, with an ANOVA test on thefinal amount raised among all combinations of project categories and offering topics, wefi nd that they are significantly different with a p-value = 0.0135. This indicates the potential correlation between the learned offering decisions and thefi nal outcomes of projects.

In this section, we analyze the offering and bundling topics learned by MOB. Wefi rst investigate the contents within each offering and bundling topics, then look at the association between offering and bundling topics. Finally, we examine the correlation between these topics and the project success.

6.1 Offering Topics

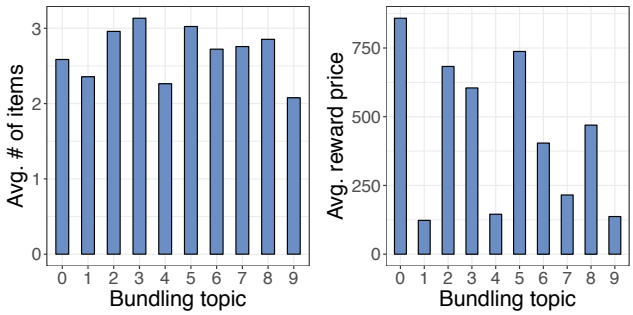
Recall that in the generative process of MOB, offering topic generates a project's reward prices and the corresponding bundling topics. We show the offering topic's learned multinomial distribution over prices and bundling topics in Table 3 and Figure 13.

Table 3 shows the topfi ve prices of each offering topic with corresponding weight ($Pr(c|x) \times 10^5$). As shown, adopted prices and bundling topics are very different across different offering topics. For example, offering topic #4 tends to offer prices that are in the lower range, while offering topic #6 tends to offer prices that are in wider price range, even up to $10,000. Taking a close look at the projects that adopt these offering topics, wefi nd that a photography project [8] adopting offering topic #4 offers rewards at prices $25, $50, and $100, while a theatre project [9] adopting offering topic #6 offers rewards at prices $1, $10, $20, $30, $50, $100, and $1000. We also show two of the offering topics' price distributions in bar charts, illustrating the difference in price range between offering topics #1 and #2.

As for the distribution of bundling topics in Figure 13 (with top three bundling topics of each offering topic highlighted), one can see that how each offering topic adopts the bundling topics vary greatly. Some offering topics use certain bundling topics more than the others, such as offering topic #2, while some offering topics use all bundling topics rather equally, such as offering topic #5. We will get into examples in the following discussion.

6.2 Bundling Topics

In MOB's generative process, bundling topic generates the words describing reward bundles. Table 4 shows topfi ve words of selected bundling topics with corresponding weight ($Pr(w|y) \times 10^5$). As shown, bundling topic #7 consists of mainly free access to certain resources and early bird benefits, while bundling topic #8 consists of signed or autographed products, records, and tickets to events. For example, a $45 reward offered by a tablet case project [10] offers



(a) Number of items (b) Price

Figure 12: Bundling topics v.s. reward bundle attributes.

"*Early Bird Special - Receive one canvas tablet case in grey*", which uses bundling topic #6, and a $75 reward offered by a country music album project [11] offers "*a hard copy of the new record, download a bonus song, download two of Wilder's albums, a handwritten note PLUS a signed photo*", which uses bundling topic #8. These results show that the bundling topics can capture both the words used to describe the rewards and the items that are included in the rewards.

Recall in MOB design (Figure 8), the bundling topic directly generates the number of items and indirectly correlates with price via the offering topic. We investigate such correlations in Figure 12(a) and Figure 12(b), respectively. As Figure 12(a) shows, the number of items included are significantly different across bundling topics (with ANOVA test of p-value $< 2 \times 10^{-16}$). As for price, Figure 12(b) shows a significant difference of prices across bundling topics (with ANOVA test of p-value $< 2 \times 10^{-16}$).

6.3 Association between Offering and Bundling Topics

To further study the correlation between offering and bundling topics, we leverage the *lift* metric defined as below.

$$lift_{k,l} = \frac{Pr(y = l|x = k)}{Pr(y = l|x = \neg k)} = \frac{\lambda_{kl}}{\lambda_{\neg kl}} \tag{14}$$

where λ is one of the parameter matrices learned by MOB.

Given a pair of offering and bundling topics, the lift measures the ratio of the generation probability over other offering topics. A large value thus suggests a frequent strategy adopted.

Wefi nd that the most associated combination is offering topic #5 and bundling topic #7, with a lift of 0.160. An example is the video game project, Universe Rush, offering a long list of 15 reward bundles, ranging from $5 to $5000.[12] For the rewards at $10 and $35 levels, they both offer early access to the game before it releases, which correspond to the words shown in Table 4.

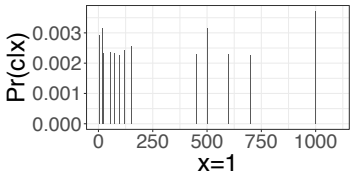
However, not all highly associated offering and bundling topics lead to project success. We therefore study each combination of strategy's success rate in the dataset.

[8] https://www.kickstarter.com/projects/652250349/jumping-in-with-both-feet-showcasing-photography/
[9] https://www.kickstarter.com/projects/1028060887/original-cast-album-for-our-glorious-cause/
[10] https://www.kickstarter.com/projects/1307647270/tablet-case
[11] https://www.kickstarter.com/projects/wilderembry/record-catalog-4-smolderoldingpictureaid
[12] https://www.kickstarter.com/projects/1326034892/universe-rush

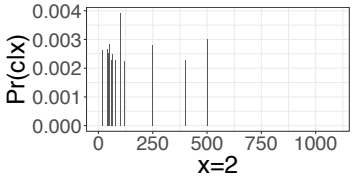
Table 3: Top five words of bundling topics and sample word clouds

#0	#1	#2	#3	#4
$50 (870.050)	$1,000 (840.363)	$100 (921.479)	$250 (775.441)	$5 (832.683)
$30 (696.325)	$20 (620.048)	$500 (562.709)	$100 (666.956)	$50 (819.334)
$10,000 (633.693)	$500 (607.865)	$50 (489.401)	$1,500 (476.967)	$0 (687.786)
$100 (597.242)	$5 (524.458)	$250 (475.199)	$75 (454.548)	$25 (579.126)
$60 (483.765)	$150 (382.460)	$40 (421.314)	$50 (406.959)	$150 (554.586)
#5	**#6**	**#7**	**#8**	**#9**
$100 (1086.169)	$10 (1463.324)	$25 (749.139)	$25 (910.866)	$15 (839.889)
$5,000 (628.003)	$1,000 (966.535)	$250 (652.717)	$500 (686.674)	$250 (729.250)
$10 (487.259)	$200 (499.434)	$500 (649.650)	$30 (616.453)	$75 (478.062)
$40 (443.887)	$10,000 (482.876)	$75 (562.429)	$2,500 (585.693)	$35 (470.563)
$0 (387.127)	$45 (413.834)	$35 (502.094)	$200 (462.858)	$200 (459.830)

(a) Offering topic #1



(b) Offering topic #2



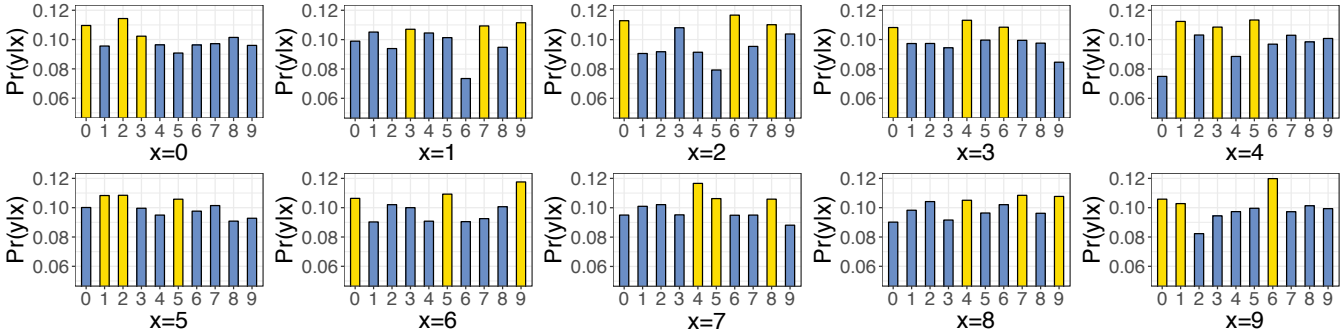


Figure 13: All offering topics' probability distributions of bundling topics.

Table 4: Top five words of bundling topics and sample word clouds

#0	#1	#2	#3	#4
Signed (11.964)	Personal (16.225)	Thank (18.762)	Album (12.131)	book (15.883)
Copy (11.764)	Digital (13.952)	Name (13.858)	Download (11.662)	Credit (12.579)
DVD (11.582)	Credit (13.678)	Invitation (11.124)	Art (10.814)	Producer (12.284)
Limited (11.449)	Shirt (11.563)	Everything (10.610)	Facebook (10.597)	Limited (12.034)
Video (11.260)	Card (10.814)	Party (10.406)	Show (10.139)	Exclusive(11.433)
#5	**#6**	**#7**	**#8**	**#9**
Thank (17.260)	Shirt (14.871)	Get (15.511)	Signed (13.449)	Cast (10.117)
Choice (15.667)	Website (12.925)	Free (12.113)	Record (10.998)	Performance (9.909)
Monthly (10.356)	Poster (12.028)	Access (10.579)	Autograph (9.741)	Artist (9.878)
Sticker (10.284)	CD (11.414)	Bird (9.478)	Ticket (9.719)	Event (9.470)
Package (9.564)	Original (10.420)	Early (9.416)	Download(9.712)	Character (9.350)

(a) Bundling topic #8

(b) Bundling topic #9

6.4 Successful Menu Bundle Design Strategy

By aligning frequent offering-bundling combination with success metrics, we can quantitatively evaluate whether this strategy is good or bad. We investigate various strategies with three measurement of project performance: success rate, raised-goal ratio, and number of backers.

Success rate: Let z be the final success of a project p, given a pair of offering and bundling topics, the success rate can be calculated as below.

$$\mu_{k,l} = \frac{|\{p|p_x = k, p_y = l, p_z = 1\}|}{|\{p|p_x = k, p_y = l\}|} \quad (15)$$

where p_x is the offering topic of p, p_y is the bundling topic of p, and p_z is thefi nal outcome of p with 1 being successful and 0 otherwise.

As shown in Table 5, offering topic #5 is mostly associated with bundling topic #7, which leads to a success rate of 80% among the projects adopting this strategy combination. These projects are mostly art and design projects. The rewards they offer are often early birds and free access to resources such as artwork. On the other hand, the combination of offering topic #3 and bundling topic #7 leads to only 40% success rate, even though the bundling topic is the same as the one previously mentioned. When taking a closer look, the projects adopting this strategy are often games and technology projects. This contrast tells us that even with the same bundling topic, when adopted under different offering topics and categories, the effect may differ.

Raised-goal ratio: Besides projects'fi nal success, we also look at the raised-goal ratio of projects, which indicate how much funds is collected in regards to the initial goal. As shown in Table 5, a combination of offering topic #2 and bundling topic #5 results in high average of raised-goal ratio. One of the projects using such strategy is a game project, New York 1776, which offers 13 rewards ranging from $2 to $335, with most prices around $100 (as shown in Figure 3).[13] An interesting decision New York 1776 made is the absence of limited edition. Rather than utilizing hunger marketing to attract backers, New York 1776 uses various packages of the game products to design the reward bundles (as shown in Table 4), resulting in a total fund of 6.87 times of the original goal collected.

Number of backers: The last performance metrics we consider is the number of backers. As shown in Table 5, the combination of offering topic #6 and bundling topic #3 results in the highest average number of backers. One of the project using this strategy is a publishing project, Artificial Intelligence for Humans.[14] It offers 12 rewards ranging from $1 to $250, with most prices around $10 to $50 (as shown in Figure 3), and 11 out of the 12 reward bundles include download of ebooks (as shown in Table 4). Due to the affordable price range, the project is able to attract backers to all of the offered reward bundles, with the most popular reward bundle being supported by 146 backers.

According to the above observation, based on a project creator's objective, she can refer to different combination of offering and bundling topics. To maximize the chance of reaching the pledging goal, one should choose based on its project category; to maximize the raised-goal ratio, one should avoid offering reward bundles at low prices, and avoid using limited edition; to maximize the number of backers supporting the project, one should offer various reward bundles at low prices.

In addition to investigating the contents of the learned offering and bundling topics, we also examine whether the offering and bundling topics can help the prediction of project success. We conduct a logistic regression with project success (successful and unsuccessful) as dependent variable, and project category, learned offering and bundling topics as independent variables.[15] The prediction is conducted with 10-fold cross validation, and compared with model only considering project categories. The results are shown in Figure 14. As shown, by adding offering and bundling topics learned by MOB as features, the predictiveness improves when predicting project success.

[13] https://www.kickstarter.com/projects/1456271622/new-york-1776

[14] https://www.kickstarter.com/projects/jeffheaton/artificial-intelligence-for-humans-vol-2-nature-al

[15] Since the goal of this analysis is not to predict the project success, we do not incorporate abundant independent variables to pursue high accuracy.

Table 5: Combinations of offering and bundling topics with project performance metrics

x	y	success rate
5	7	80.00%
2	1	66.67%
0	5	65.22%
3	9	64.71%
2	8	63.16%

x	y	raised-goal
2	5	3.94
0	5	3.15
6	3	2.61
4	2	2.08
5	0	2.07

x	y	backers
6	3	227.00
2	5	193.00
8	1	164.49
1	2	154.80
0	7	149.67

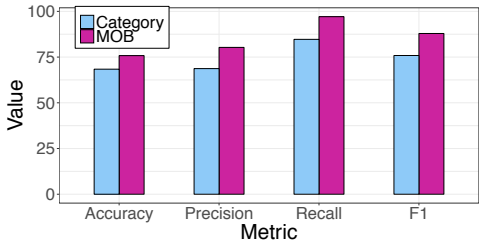


Figure 14: Performance of project success prediction.

6.5 Menu Bundle Design Suggestion

With MOB's ability to obtain the offering and bundling topics, we select three unsuccessful projects' designs of menu bundles, compare them with successful projects' designs in the same categories, and discuss potential mistakes a project creator may make.

Table 6 shows two unsuccessful publishing projects that have different offering topics from the successful one. The table shows their funds raised over the initial goals, the reward prices, and the number of backers received by each reward in parentheses. As shown, both the unsuccessful projects provide reward lists shorter than what the successful project provides. Also, the unsuccessful projects offer rewards with narrow price ranges, while the successful project offers rewards with a wide range of prices from $10 to $5000. Projects using this strategy are able to cater various backers with different budget. To improve, the unsuccessful projects should offer a longer list of rewards, and widen their price range.

Table 7 shows one unsuccessful food project that shares the same offering topic but very different bundling topics as the successful project. As shown, the two projects have very similar offering behavior in terms of number of rewards offered and the price range. However, how they incorporate reward items into bundles are significantly different. For the unsuccessful project, three out of seven rewards include only single reward item, which are all experiences rather than actual product; for the successful project, even though three out of nine rewards include only one reward item, they are all actual product except the hand written thank you note offered at $5. Another characteristic that the successful project has is its utilization of reward hierarchy. As the price increases, the reward items add on top of the ones in the previous level. From a potential backer's point of view, this design gives a stronger justification of the pricing for each reward, and thus make it easier to choose among rewards. However, the unsuccessful project does not adopt such strategy. For improvement, the unsuccessful project could

Table 6: Offering suggestion for publishing projects

Unsuccessful projects		Successful project
Biography project $490/$8000	GPS story project $1,186/$6,000	Poetry project $10,087/$10,000
$10 (1) $25 (3) $50 (0) $100 (1) $250 (1)	$1 (1) $25 (13) $50 (10) $750 (0)	$10 (25) $25 (41) $50 (12) $100 (8) $200 (3) $500 (1) $1000 (2) $5000 (0)

Table 7: Bundling suggestion for food projects

Unsuccessful project	Successful project
Coffee project $501/$7000	Caramel project $2520/$1500
$5 (1) y=4: thank you, a hug, 8 oz Americano $15 (0) y=9: your name on our website, a bag of Diosa Rosquillas $30 (4) y=0: 12 oz bag of coffee, a bag of Diosa Rosquillas $60 (0) y=0: thank you, our coffee, your name on our website, a bag of Diosa rosquillas, T shirt $125 (1) y=4: 2-guest complete gourmet cafe style meal $300 (0) y=5: 4-guest complete gourmet cafe style meal $1000 (0) y=6: Coffee farm tour for two	$5 (2) y=4: hand written thank you note $10 (26) y=6: 1/4 pound bag gourmet caramels $25 (30) y=8: 1/2 pound bag gourmet caramels, half pound bag brittle/toffee $50 (11) y=0: 1 & 1/2 pound bag gourmet caramels/toffee/brittle $75 (4) y=7: 2 pounds gourmet caramels/brittles/toffee. $100 (5) y=5: 3 pounds gourmet caramels/toffees/brittles, coffee mug $250 (0) y=3: 5 month subscription of 1 pound gourmet caramels/toffee/brittle, coffee mug $500 (0) y=8: 10 month subscription of gourmet caramel/toffee/brittle, coffee mug $1000 (0) y=6: everything from the $500 reward tier, baseball cap, t-shirt, custom candy

break down all the available reward items, and rearrange them in a hierarchical manner like the successful project does.

7 CONCLUSION

In crowdfunding, reward menu bundle design can significantly influence the project pledging results. Our analyses show that projects offering more rewards and leveraging bundling are more likely to succeed. We alsofi nd that rewards with more items bundled together tend to be more expensive, and in turn attracts less backers. The reward menu design process involves decisions on offering and bundling. We develop a Menu-Offering-Bundle (MOB) model to capture behaviors in these two decisions as well as their interactions with project category, reward price and reward content.

Using a real-world data set from Kickstarter for training and testing, we demonstrate that the trained MOB model can not only help predict the projectfi nal result, but also predict reward designs, with better performance than baseline methods. Moreover, leveraging MOB, we present a diagnosis system to investigate problematic menu bundle designs and provide improvement suggestions. Case studies demonstrate the actionable feedbacks obtainable from MOB.

For future works, we plan to incorporate the actions taken by backers on the platform to see how different backers react to different offering and bundling decisions of projects. We will also explore the effect of offering and bundling design decisions over time. Furthermore, we plan to expand the study on the design behaviors of offering and bundling decisions beyond crowdfunding to better understand their impacts in menu design for other domains.

ACKNOWLEDGMENTS

This work was supported in part by National Science Foundation grant No.: 1717084.

REFERENCES

[1] J. An, D. Quercia, and J. Crowcroft. Recommending investors for crowdfunding projects. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 261–270, New York, NY, USA, 2014. ACM.
[2] Y. Bakos and E. Brynjolfsson. Bundling information goods: Pricing, profits, and e$ ciency. *Management science*, 45(12):1613–1630, 1999.
[3] S. Bannerman. Crowdfunding culture. *Journal of Mobile Media*, 7(01):1–30, 2013.
[4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.
[5] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391, 1990.
[6] N. Desai, R. Gupta, and K. Truong. Plead or pitch? the role of language in kickstarter project success.
[7] Q. Gao and M. Lin. Lemon or cherry? the value of texts in debt crowdfunding. 2015.
[8] M. D. Greenberg, B. Pardo, K. Hariharan, and E. Gerber. Crowdfunding support tools: predicting success & failure. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 1815–1820. ACM, 2013.
[9] Y. Lin, W.-C. Lee, and C.-C. H. Chang. Analysis of rewards on reward-based crowdfunding platforms. In *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on*, pages 501–504. IEEE, 2016.
[10] C.-T. Lu, S. Xie, X. Kong, and P. S. Yu. Inferring the impacts of social media on crowdfunding. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 573–582. ACM, 2014.
[11] B. Ma, N. Zhang, G. Liu, L. Li, and H. Yuan. Semantic search for public opinions on urban affairs: A probabilistic topic modeling-based approach. *Information Processing & Management*, 52(3):430–445, 2016.
[12] T. Mitra and E. Gilbert. The language that gets people to give: Phrases that predict success on kickstarter. In *CSCW*, pages 49–61. ACM, 2014.
[13] M. J. Paul and M. Dredze. Discovering health topics in social media using topic models. *PloS one*, 9(8):e103408, 2014.
[14] V. Rakesh, J. Choo, and C. K. Reddy. Project recommendation using heterogeneous traits in crowdfunding. In *ICWSM*, pages 337–346, 2015.
[15] V. Rakesh, W.-C. Lee, and C. K. Reddy. Probabilistic group recommendation model for crowdfunding domains. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pages 257–266. ACM, 2016.
[16] W. Shafqat, S. Lee, S. Malik, and H.-c. Kim. The language of deceivers: Linguistic features of crowdfunding scams. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 99–100. International World Wide Web Conferences Steering Committee, 2016.
[17] W. Trouleau, A. Ashkan, W. Ding, and B. Eriksson. Just one more: Modeling binge watching behavior. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, pages 1215–1224, 2016.
[18] A. Xu, X. Yang, H. Rao, W.-T. Fu, S.-W. Huang, and B. P. Bailey. Show me the money!: An analysis of project updates during crowdfunding campaigns. In *SIGCHI*, pages 591–600. ACM, 2014.
[19] P. Yin, P. Luo, W. Lee, and M. Wang. Silence is also evidence: interpreting dwell time for recommendation from psychological perspective. In *The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013, Chicago, IL, USA, August 11-14, 2013*, pages 989–997, 2013.
[20] C. Zhu, H. Zhu, Y. Ge, E. Chen, Q. Liu, T. Xu, and H. Xiong. Tracking the evolution of social emotions with topic models. *Knowledge and Information Systems*, 47(3):517–544, 2016.