Probabilistic Group Recommendation Model for Crowdfunding Domains

Vineeth Rakesh
Wayne State University
Detroit, MI - 48202
vineethrakesh@wayne.edu

Wang-Chien Lee
Pennsylvania State University
State College, PA - 16801
wlee@cse.psu.edu

Chandan K. Reddy
Wayne State University
Detroit, MI - 48202
reddy@cs.wayne.edu

ABSTRACT

Crowdfunding has gained a widespread popularity by fueling the creative minds of entrepreneurs. Not only has it democratized the funding of startups, it has also bridged the gap between the venture capitalists and the entrepreneurs by providing a plethora of opportunities for people seeking to invest in new business ventures. Nonetheless, despite the huge success of the crowdfunding platforms, not every project reaches its funding goal. One of the main reasons for a project's failure is the difficulty in establishing a linkage between it's founders and those investors who are interested in funding such projects. A potential solution to this problem is to develop recommendation systems that suggest suitable projects to crowdfunding investors by capturing their interests. In this paper, we explore Kickstarter, a popular reward-based crowdfunding platform. Being a highly heterogeneous platform, Kickstarter is fuelled by a dynamic community of people who constantly interact with each other before investing in projects. Therefore, the decision to invest in a project depends not only on the preference of individuals, but also on the influence of groups that a person belongs and the on-going status of the projects. In this paper, we propose a probabilistic recommendation model, called *CrowdRec*, that recommends Kickstarter projects to a group of investors by incorporating the on-going status of projects, the personal preference of individual members, and the collective preference of the group . Using a comprehensive dataset of over 40K crowdfunding groups and 5K projects, we show that our model is effective in recommending projects to groups of Kickstarter users.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database applications-Data Mining; I.2.6 [**Artificial Intelligence**]: Learning; H.3.3 [**Information Search and Retrieval**]: Information filtering

Keywords

Recommendation; crowdfunding; probabilistic models; topic models; group recommendation.

1. INTRODUCTION

Crowdfunding has emerged as "the next big thing" in entrepreneurial financing. By providing the much-needed seed capital for new business ventures, it has created plenty of new job opportunities and revived lost business ventures. In 2014 alone, the crowdfunding platforms raised over $16.2 billions of dollars worldwide, thereby becoming a viable alternative to banks, brokers, and other financial intermediaries for people seeking funding to jump-start their business ventures. The concept of crowdfunding is analogous to micro-financing where the required funds are collected by pooling small amounts of money from several individuals. Since its launch in April 2009, Kickstarter has grown to become one of the most popular crowdfunding platforms. Kickstarter terms the founders of a project as *creators* and the investors as *backers*. The creators express their ideas by posting a detailed description about their project. Usually, the description contains videos, images and textual information that explains the novelty of the project. In addition, the creators provide a detailed timeline, funding goal, and the rewards for different pledge levels.

The need for recommender systems in crowdfunding platforms. Although the interest in using Kickstarter for crowdfunding has been outstanding, the success rate of projects is not very impressive. Statistical reports show that only about 37% of the projects succeeded in reaching their funding goal [19], which means over 60% of the crowdfunding projects in Kickstarter have failed. One of the main reasons for these failures is the lack of publicity [17]. We argue that recommendation systems that suggest suitable projects to crowdfunding investors can address this problem. Thus, in this paper, we propose a probabilistic recommendation model, called *CrowdRec*, that recommends projects to a *group* of potential backers. To the best of our knowledge, very few studies have explored the crowdfunding domain from a data mining perspective [3, 13, 23, 34].

Recommendation challenges in crowdfunding. Recommendation in crowdfunding poses a number of challenges. Based on our preliminary study on analyzing Kickstarter user behaviors [28], we found that a diverse set of factors collectively influence the users' decision to back a project. Hence, recommendation cannot be based on some simple set of straightforward features that are directly available from the projects. Consider the decision making of a
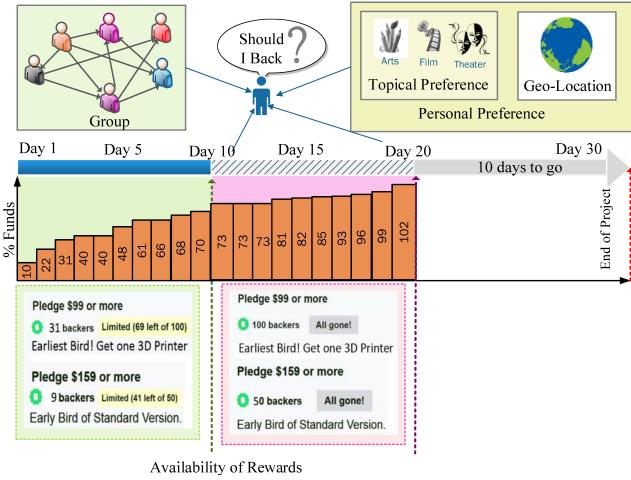
Figure 1: Impact of project, personal, and social network based features on Kickstarter users.

Kickstarter investor, as illustrated in Figure 1; we can see that the user is influenced by (1) his own personal interest, (2) the group (or community) that the user is associated with, and (3) the real-time status of the project. The personal interest of a backer is attributed to his preferences over the topic or the geo-location of the project. The group's influence on the other hand is amplified by the pervasive growth of social media such as Twitter and Facebook, where users' decision to back a project depends not only on their personal interests, but also on their relationship to a social-group of peer investors they communicate with. Therefore, when designing a recommender system for crowdfunding, it is important to incorporate the group's influence. Finally, due to the *transient* nature of crowdfunding projects, the real-time status of the project plays a critical role in determining the backing habits of a user. Notice that, in conventional recommendation such as movies or books, it is reasonable to apply collaborative filtering techniques since the recommended items usually can serve many users for several years. This is not the case in crowdfunding; in Kickstarter, once the project expires after its posting period, we cannot recommend the project to any user any more. For instance, as shown in Figure 1, the project has a duration of 30 days. Recommending it after expiration makes no sense. This transient nature of projects gives rise to some interesting real-time properties such as (a) *the popularity of projects* and (b) *the availability of rewards*. In Kickstarter, the popularity of a project could be measured based on the percentage of funds collected at a specific time. In Figure 1, we see that, on Day 10 the project has already collected over 70% of it's goal amount, indicating high popularity. Therefore, there is a high chance for a user to fund such popular projects. Nonetheless, the popularity is not the sole deciding factor. In the same figure, we can also see that, on Day 20, although the project remains popular, most of the rewards are already sold out (denoted by the phrase "All Gone") due to the demand. This in-turn means that people might not be interested in backing such projects (with no availability of certain type of rewards), despite it's popularity.

Overview of the proposed approach. To create a personalized recommendation system for crowdfunding, we propose a *group recommendation* model called *C*rowdRec. Using a probabilistic generative framework, we incorporate various heterogeneous features related to the project, users, and their social groups to precisely model the interests of investors. Notice that in the crowdfunding scenario, projects are like *living entities that survive for a finite duration, and are affected by real-time actions such as popularity and reward availability.* Therefore, we need to leverage the information about the dynamic on-going status of a project when recommending it to the users.

Research Contributions. The major contributions of this paper are summarized as follows:

1. We propose a group-recommendation model for crowdfunding domains, which incorporates the dynamic-status of the on-going projects to recommend Kickstarter projects for a group of investors.

2. We use a diverse set of features about the projects. These features include (1) topical preference (2) geo-location preference (3) social-network links of backers and (4) various temporal information about the projects to incorporate rich prior information into our probabilistic model.

3. Using comprehensive evaluation techniques, we show that our model outperforms a number of baselines and a state-of-the-art group-recommendation model to provide effective and meaningful recommendations for backer groups in Kickstarter.

The rest of this paper is organized as follows. We review the related work on crowdfunding and group recommendation models in Section 2. In Section 3, we discuss the notion of groups in Kickstarter and the challenges associated with group recommendation. The CrowdRec model and it's generative process are introduced in Section 4, followed by the derivation of the model parameters. In Section 5, we show the different ways of incorporating various prior information. In Section 6, we explain the data collection methodology and report the experimental results for performance evaluation. Finally, we conclude the paper in Section 7.

2. RELATED WORKS

In this section, we review two lines of related research namely, crowdfunding and recommender systems.

2.1 Crowdfunding Litrature

Crowdfunding and Kickstarter. Since crowdfunding is still an emerging research domain, most works in this area are relatively new. The dynamics of Kickstarter are examined in a recent survey [20], while various types of crowdfunding platforms are compared in [26]. In [14, 17], the authors analyze the crowdfunding platforms to learn the motivation behind the users who create and invest in crowdfunding projects. Moreover, the effect of frequent updates over the success rate of projects is explored in [34]; and the impact of social network on Kickstarter projects is delineated in [23].

Recommendation in Crowdfunding Platforms. So far, there are very few studies on developing recommendation models for crowdfunding platforms. In [11], a personalized loan recommendation system for a micro-financial platform called *Kiva.org* is proposed. Recently, an SVM classifier is trained using updates, comments, Facebook friends and other features from Kickstarter to recommend investors

to Kickstarter projects [3]. In our previous work on Kickstarter [28], a variety of traits based on backer personality, geo-location, project quality, and social network have been analyzed and incorporated into a gradient boosting tree model to recommend backers to Kickstarter projects. Finally, in our recent work [21], we formulate a survival analysis problem to predict the success of Kickstarter projects.

2.2 Recommender Systems

Basic Recommendation Techniques. The core of all recommender systems is to obtain a utility function that estimates the preference of a user towards an item. Essentially, recommender systems can be divided into two main categories: collaborative filtering methods and content-based methods. Collaborative filtering-based techniques utilize the user rating of items to derive the similarity between the users (or items) to recommend relevant items to users. Content-based approaches, on the other hand, utilize the content features of the users (or items) to make the recommendation. These features might include user's topical interest, age, gender, etc. For a comprehensive summary of collaborative filtering techniques, the readers are referred to survey articles [1, 30]. In the context of crowdfunding, the heterogeneity in Kickstarter crowdfunding suggests hybrid recommendation techniques, which are gaining a lot of attention in recent years. A naive way of using hybrid recommendation is to use collaborative and content-based methods in parallel [5, 12, 31]; nonetheless, these approaches do not incorporate the influence of social groups to model the interests of users. In this work, we adopt probabilistic latent modeling (discussed below) to seamlessly integrate the collaborative and content-based filtering approaches into a unified group recommendation model.

Probabilistic Models for Recommendation. The original aspect model is proposed by Hofmann for Probabilistic Latent Semantic Analysis (PLSA) [15, 16]. Since then, more complex machine learning models have been proposed, including multinomial mixture models, latent dirichlet allocation (LDA), markov models and latent factor models [8, 24, 29]. The PLSA aspect model has been widely used in information retrieval and data mining applications [27]. For example, in [10], the aspect model is used for recommending communities, while in [35], an additional latent variable has been added to the aspect model to capture the influence of friends on a user's topical interest.

Group Recommendation. Different from conventional recommendation which typically recommends an item to a user, the group recommendation aims to either (a) recommend a group (or community) to a user or (b) recommend an item to a group of users. In type (a), one usually recommends a set of groups (i.e., communities) to a user based on various measures such as (1) topical similarity between the community and the user, (2) popularity of the community at a given time, (3) proximity of geo-location between the user and the members of the community, etc., [9, 10, 32, 33, 37]. Our definition of the group recommendation in this paper belongs to type (b), which is new to the data mining research and thus very few studies have explored this research direction [22, 25, 36]. In [35], a probabilistic aspect model is proposed to learn the interests of users using both their personal preference and the influence from their social network. Using score aggregation strategies such as average and least-misery techniques [4], recommendation

of items is made for a targeted group of users. In [22], a unified group recommendation model is proposed and later extended by incorporating a group-topic distribution [36] to provide an improved recommendation. The group recommendation model proposed in this paper is uniquely different from the aforementioned works as we incorporate the "live-status" of projects along with user preference and group influence in the model.

3. GROUPS IN CROWDFUNDING

To elucidate the need for group recommendation in Kickstarter, we obtain backers who have their Twitter profiles and retrieve their friends and followers from Twitter search API. We then create communities of these backers to analyze *whether the backing habits of these investors are influenced by their relationship to a community*[1]. The process of community creation is detailed in our previous work [28]. To measure the influence of a group over the backer, we calculate their *Affinity* score as follows:

$$Affinity(b, g) = |F(b) \cap F(g)| \qquad (1)$$

where $F(b)$ and $F(g)$ denote the set of all the followers and followees of a backer b and a community g, respectively; $|F(b) \cap F(g)|$ indicates the number of mutual friends between this backer and the members of the community. Figure 2(a) shows the outcome of this analysis, where $p(M(b,v)|M(g,v))$ indicates the probability of a user b backing a project v, given that v is backed by the members of the group g. In this notation, $M(b,v)$ denotes the action of a backer investing in project v and $M(g,v)$ denotes the same by a community g. From this figure, one can see that the stronger *affinity* of a user towards a group leads to a greater chance for this user to back the same project that was backed by the group. Therefore, given a project, the goal of our CrowdRec recommendation model is to *identify a group of users who may potentially back this project*.
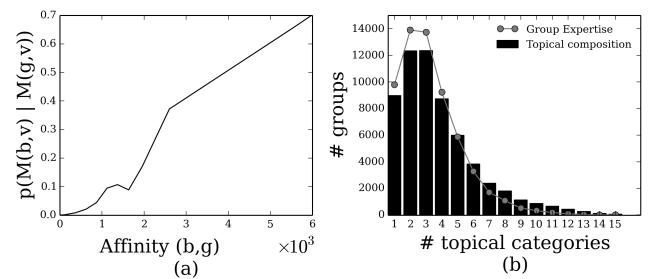
Figure 2: Characteristics of groups in Kickstarter. (a) shows the influence of communities (groups) over the backing habits of users and (b) shows the topical composition in groups.

Next, we analyze the preferences of individual members in a group. To proceed, for every group in our dataset, we calculate the number of unique topical categories of projects backed by the members of the group. As shown in Figure 2(b), although there are groups consisting of members who are interested in just one single topic, *a majority of them have diverse interests, i.e., the members have backed projects*

---

[1]In this paper, we also refer these communities as *groups*, so these two terms are used interchangeably.

from multiple topical categories. In addition, we analyze the expertise of individual group members by calculating the number of backers who are experts in a specific topical category of projects. We say a backer to be an expert in a topic if she has backed at least 3 projects from the same category and has over 15 backings in total. As depicted in Figure 2(b), a *majority of groups consist of backers who are experts in multiple topical categories*. We aim to exploit these observations in the proposed CrowdRec model.

4. CROWDREC: A PROBABILISTIC GENERATIVE MODEL

In this section, we introduce CrowdRec, a probabilistic generative model for recommending crowdfunding projects to groups of users. The recommendation model aims to capture the following observations: (1) A crowdfunding group may support projects from multiple topical categories, (2) A user's backing decision is based not only on her personal preference but also on the collective preferences of her groups; (3) A group's collective preference to support a project is strongly correlated with the personal preferences of topically authoritative users (i.e., users expertise) within the group, (4) The dynamic status of a project impacts both the individual investor's personal preferences and the group's collective preferences in backing crowdfunding projects.
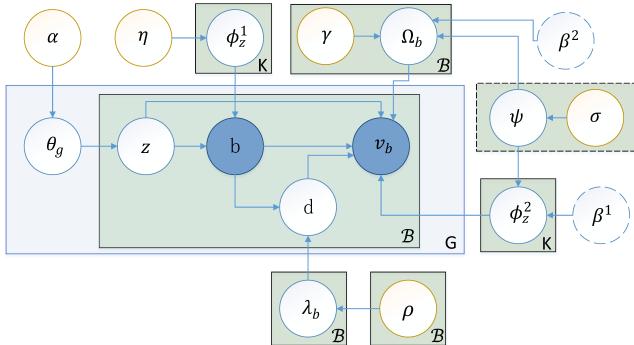
Figure 3: Graphical representation of the CrowdRec model.

In this section, we first formulate the modeling problem and then present the generative process captured in our CrowdRec model. Next, we show how the dynamic status of projects is incorporated into the model. Finally, we derive the parameters of the model to facilitate model learning via Gibbs sampling.

Problem Statement: Given a set of *projects* $V = \{v_1, v_2, ..., v_{|V|}\}$, a set of *backers* $B = \{b_1, b_2, ..., b_{|B|}\}$, and a set of *groups* $G = \{g_1, g_2, ..., g_{|G|}\}$, let $B^g \subset B$ denote a group of backers in group g, i.e., $B^g = \{b_1^g, b_2^g, ..., b_{|B^g|}^g\}$. The action of a group g backing a project v is denoted by $M(B^g, v) = \{(b, v) | b \in B^g\}$, where (b, v) refers to an individual b from a group g choosing to back a project v. The goal of our CrowdRec model is to recommend a ranked list of V projects to a target (or new) group $\tilde{g}$.

4.1 Generative Process

The Graphical representation of the CrowdRec model is shown in Figure 3. We describe the generative process in our model as follows.

- Each group $g \in G$ is composed of members who are interested in certain particular project categories. Therefore, the collective preference of the group is captured by θ_g, which is represented as a distribution over a universal set of latent topics in projects. Here, θ_g follows a symmetric Dirichlet distribution, i.e., $\theta_g \sim Dirichlet(\alpha)$.
- Based on θ_g, the group g chooses a single topic z and nominates a user b to decide whether to back a project v. The distribution ϕ_z^1 captures the expertise of this backer over the topic z.
- To decide whether to back a project v, the nominated user relies on either her own personal interest or the collective preference of the group. This decision is governed by the random variable D, which takes the binary value 0 or 1. Thus, we model d using a binomial distribution with beta prior.
- If d is 0:
  - The user picks the project based on (1) *the group's influence* (i.e., collective preference), which is a multinomial over the distribution ϕ_z^2 and (2) *the current status of the project*. The current status of the project (in the dotted box) is denoted by a ψ, $\psi \sim Dirichlet(\sigma)$.
- If d is 1:
  - The user picks the project based on (1) *her own topical interest* (personal preference), which is a multinomial over the distribution Ω_b and (2) *the current status of the project*.

Incorporating Dynamic Status of Projects. Notice that the status of a project plays a key role in the generative process described above. It is an external factor determined by the progression of the project over time. In Figure 3, this is indicated by the *project-status* distribution ψ, which is constrained by a dirichlet prior σ. ψ in-turn affects the *project-topic* distribution ϕ^2 and *user-project* distribution Ω. When the user relies on the group's influence to back a project v, the project-topic distribution ϕ^2 (i.e. $V \times K$ matrix) is multiplied by the project-status distribution ψ (i.e. $1 \times V$ matrix). In other words, ψ becomes a prior for ϕ^2. Alternatively, if the user backs a project based on her own preference, Ω (i.e. $B \times V$ matrix) is multiplied by ψ.

In the CrowdRec model, β^1 and β^2 are concentration scalars that affects the extent to which a group (or a user) relies on project status to make the backing decision. When β^1 is high, the group strongly relies on the on-going status to back the project, which means ϕ^2 becomes similar to ψ. Alternatively, if β^1 is low, the group's decision to back a project is independent of the on-going status of the project. Same applies for the scalar β^2, which affects the variable Ω. In the literature, this type of formulation is known as the *hierarchical Polya-Urn model*, which has been used to model the global and local topic distributions in LDA [6, 2, 7]. Algorithm 1 summarizes the complete generative process and Table 1 provides the list of symbols used in this paper.

4.2 Parameter Estimation

To learn the parameters in the CrowdRec model, the eastimation of the posterior is given by:

$$p(z, d|b, v, \alpha, \eta, \gamma, \sigma, \rho) = \frac{p(z, d, b, v|.)}{p(b, v|.)} \quad (2)$$

Table 1: List of notations used in this paper.

Symbol	Description
$\mathcal{B} = \{b_i\}$	set of backers, b_i indicates a single backer
$\mathcal{V} = \{v_i\}$	project set, v_i indicates a single project
$G = \{g_j\}$	crowdfunding groups, g_j indicates a single group
D	a binary decision variable, representing d=1 or d=0
$Z = \{z_i\}$	latent topics assigned to projects in Z
K	number of topics specified as parameter
$i = (b, v)$	a tuple that indicates backer b picks project v
θ_g	topic distribution of a group g
ϕ_z^1	$B \times K$ latent matrix for a backer b
ϕ_z^2	$G \times K$ latent matrix for a group g
Ω_b	$B \times V$ matrix for a backer b
λ_b	prior for the binary decision variable D
ψ	dynamic status distribution of a project v
$\alpha, \eta, \gamma, \sigma, \rho$	parameters of $\theta, \phi^1, \Omega, \psi, \lambda$
β_1, β_2	concentration scalars for ϕ^2, Ω
$c_{k,g,i}$	# times i is assigned to topic k in group g
$c_{k,g,b}$	# times backer b is assigned to topic k in group g
$c_{k,g,v}$	# times project v is assigned to topic k in group g
$c_{b,g,v}$	# times project v is assigned to backer b in group g
$c_{b,g,d}$	# times choice d chosen by a backer b in group g

Algorithm 1: Generative process of `CrowdRec` model

for *each project $v \in \mathcal{V}$* **do**
  Draw $\psi_v \sim Dirichlet(\sigma)$
end
for *each topic $z_k, k \in \mathcal{K}$* **do**
  Draw $\phi_z^1 \sim Dirichlet(\eta)$
  Draw $\phi_z^2 \sim Dirichlet(\psi\beta^1)$
end
for *each backer $b \in \mathcal{B}$* **do**
  Draw $\Omega_b \sim Dirichlet(\psi\beta^2\gamma)$
  Draw $\lambda \sim Beta(\rho)$
end
for *each group $g \in G$* **do**
  Draw $\theta_g \sim Dirichlet(\alpha)$
  for *each backer,b in group g* **do**
    Draw $z \sim Multinomial(\theta_g)$
    Draw $b \sim Multinomial(\phi_z^1)$
    Draw the *decision $d \sim Bernoulli(\lambda_b)$*
    if *$d = 0$* **then**
      Draw $v \sim Multinomial(\phi_z^2)$
    end
    if *$d = 1$* **then**
      Draw $v \sim Multinomial(\Omega_b)$
    end
  end
end

The likelihood of the above equation is expanded as follows:

$$p(z, d, b, v|.)$$
$$= \underbrace{\int p(z|\theta)p(\theta|\alpha)d\theta}_{(A1)} . \underbrace{\int p(b|z,\phi^1)p(\phi^1|\eta)d\phi^1}_{(A2)} . \underbrace{\int p(d|\lambda)p(\lambda|\rho)d\lambda}_{(A3)} .$$
$$\underbrace{\int\int\int p(v|b,d,z,\Omega,\phi^2)p(\Omega|\gamma,\psi,\beta^2).p(\phi^2|\psi,\beta^1)p(\psi|\sigma)d\Omega d\phi^2 d\psi}_{(A4)}$$
$$(3)$$

To infer the parameters $\phi^1, \Omega, \psi, \phi^2$ and λ, we obtain samples from this high-dimensional distribution using collapsed Gibbs sampling-based approach. It is important to note that there are complex relationships between the latent-topic variable Z and the latent-decision variable D. To overcome this problem, we adopt the two-step Gibbs sampling method proposed by Yuan et. al. [36] by decomposing the expression $(A4)$ of Equation (3) as follows:

$$\int\int\int p(v|b,d,z,\Omega,\phi^2).p(\Omega|\gamma,\psi,\beta^2).p(\phi^2|\psi,\beta^1).p(\psi|\sigma)d\Omega d\phi^2 d\psi$$
$$= \underbrace{\int\int p(v^0|z,d^0,\phi^2)p(\phi^2|\psi,\beta^1)p(\psi|\sigma)d\phi^2 d\psi}_{(B1)}$$
$$\underbrace{\int\int p(v^1|b,d^1,\Omega)p(\Omega|\gamma,\psi,\beta^2)p(\psi|\sigma)d\Omega d\psi}_{(B2)} \qquad (4)$$

where expression $(B1)$ corresponds to the decision variable $d = 0$; in other words, when a user chooses to back a project v^0 based on his group's interest and expression $(B2)$ corresponds to the decision $d = 1$ i.e., when a user chooses a project v^1 based on her own interest. Therefore, using Gibbs sampling, we sample the latent variable d for two different cases: (a) when $d = 0$ and (b) when $d = 1$. Similarly, we sample the latent variable z when (a) project v^0 is chosen and (b) when project v^1 is chosen.

The derivation of the collapsed Gibbs sampling equation for the topic-latent variable z and the decision-latent variable d is similar to [36]. The probability of a tuple $i = (b, v)$ belonging to a latent topic z is derived as follows:

$$p(z_{g,i} = k|Z^{-(g,i)}, v^0, b) \propto$$
$$\frac{c_{k,g,i*}^{-(g,i)} + \alpha_k}{c_{k*,g,i*}^{-(g,i)} + \alpha_{k*}} \cdot \frac{c_{k,g*,b}^{-(g,i)} + \eta_b}{c_{k,g*,b*}^{-(g,i)} + \eta_{b*}} \cdot \left(\frac{c_{k,g*,v}^{-(g,i)}}{c_{k,g*,v*}^{-(g,i)}} + \beta^1 \frac{c_{k*,g*,v}^{-(g,i)} + \sigma_v}{c_{k*,g*,v*}^{-(g,i)} + \sigma_{v*}} \right)$$
$$(5)$$

$$p(z_{g,i} = k|Z^{-(g,i)}, v^1, b) \propto \frac{c_{k,g,i*}^{-(g,i)} + \alpha_k}{c_{k*,g,i*}^{-(g,i)} + \alpha_{k*}} \cdot \frac{c_{k,g*,b}^{-(g,i)} + \eta_b}{c_{k,g*,b*}^{-(g,i)} + \eta_{b*}}$$
$$(6)$$

In the above, the variable of type $c_{x,y,z}$ indicates a count as described in Table 1, and the symbol $*$ over the subscript variables denotes the summation over the respective subscript variables. For example, $c_{k,g,v}$ indicates the number of times the project v is assigned to topic k in group g and $c_{k,g*,v}$ is the same variable that is summed across all the groups $g \in G$. The superscript symbol $-(g,i)$ means that we exclude the ith tuple for group g when sampling. The probability of a tuple $i = (b, v)$ choosing a decision d is derived as follows:

$$p(d_{g,i} = 0|D^{-(g,i)}, Z, v, b) \propto$$
$$\frac{c_{b,g*,d_0}^{-(g,i)} + \rho_0}{c_{b,g*,d_1}^{-(g,i)} + c_{b,g*,d_0}^{-(g,i)} + \rho_0 + \rho_1} \cdot \left(\frac{c_{k,g*,v}^{-(g,i)}}{c_{k,g*,v*}^{-(g,i)}} + \beta^1 \frac{c_{k*,g*,v}^{-(g,i)} + \sigma_v}{c_{k*,g*,v*}^{-(g,i)} + \sigma_{v*}} \right)$$
$$(7)$$

$$p(d_{g,i} = 1|D^{-(g,i)}, Z, v, b) \propto$$

$$\frac{c_{b,g^*,d_1}^{-(g,i)} + \rho_1}{c_{b,g^*,d_1}^{-(g,i)} + c_{b,g^*,d_0}^{-(g,i)} + \rho_0 + \rho_1} \cdot \left(\frac{c_{g^*,b,v}^{-(g,i)} + \gamma_v}{c_{g^*,b,v^*}^{-(g,i)} + \gamma_{v^*}} + \beta^2 \frac{c_{k^*,g^*,v}^{-(g,i)} + \sigma_v}{c_{k^*,g^*,v^*}^{-(g,i)} + \sigma_{v^*}} \right)$$
$$(8)$$

After obtaining sufficient number of samples using the above Gibbs update rules, we can finally infer the parameters $\phi_z^1, \Omega_b, \psi, \phi_z^2$ and λ_b as follows:

$$\phi_{z,b}^1 = \frac{c_{k,g^*,b} + \eta_b}{c_{k,g^*,b^*} + \eta_{b^*}} \tag{9a}$$

$$\Omega_{b,v} = \frac{c_{g^*,b,v} + \gamma_v}{c_{g^*,b,v^*} + \gamma_{v^*}} + \beta^2 \frac{c_{k^*,g^*,v} + \sigma_v}{c_{k^*,g^*,v^*} + \sigma_{v^*}} \tag{9b}$$

$$\psi_v = \frac{c_{k^*,g^*,v} + \sigma_v}{c_{k^*,g^*,v^*} + \sigma_{v^*}} \tag{9c}$$

$$\phi_{z,v}^2 = \frac{c_{k,g^*,v}}{c_{k,g^*,v^*}} + \beta^1 \frac{c_{k^*,g^*,v}^{-(g,i)} + \sigma_v}{c_{k^*,g^*,v^*}^{-(g,i)} + \sigma_{v^*}} \tag{9d}$$

$$\lambda_b = \frac{c_{b,g^*,d_1} + \rho_1}{c_{b,g^*,d_1} + c_{b,g^*,d_0} + \rho_0 + \rho_1} \tag{9e}$$

Recommending projects: To recommend a set of projects to a new group $\widetilde{g}$, we need to learn the group-topic distribution $\theta_{\widetilde{g}}$. This is done by estimating the posterior distribution of topics $\widetilde{z}$, given the backers $\widetilde{b}$ and the estimated backer-topic distribution ϕ_z^1 and the hyperparameter α_k that was obtained from our CrowdRec model $\mathcal{M}$.

$$p(\widetilde{z}_{\widetilde{g},j} = k|\mathcal{B}_j = b_j, \widetilde{Z}^{-(\widetilde{g},b)}, \mathcal{B}^{-(j)}; \mathcal{M}) \propto \phi_{k,b}^1(c_{k,\widetilde{g},i^*}^{-(\widetilde{g},i)}) \quad (10)$$

Once we sample the topics for this new group $\widetilde{g}$, we recommend a new project based on the following equation.

$$p(v|\widetilde{g}) = \prod_{b \in \mathcal{B}^{\widetilde{g}}} \psi_v \sum_{z \in Z} \theta_{\widetilde{g},z}.(\lambda_b.\Omega_{b,v} + (1 - \lambda_b).\phi_{z,v}^2) \tag{11}$$

The above equation captures the following components when recommending a project to a group: (1) the individual's personal preference over a project Ω, (2) the influence of the group over a backer is captured by ϕ^2, (3) the topical preference of the group and the users are captured by θ and Ω, and (4) finally, the dynamic status (or popularity) of the project ψ.

5. PRIOR INFORMATION

In this section, we discuss the approaches we adopt to estimate the priors incorporated in CrowdRec. Notice that we have priors for two different distributions: (1) a static prior γ for the distribution Ω_b, which is a $B \times V$ matrix indicating the preferences of backers towards the Kickstarter projects; (2) a dynamic prior σ for distribution ψ, which is a $1 \times V$ row matrix indicating the on-going status of the project.

For the first case, we estimate the static prior γ by exploiting the backing history of all the users $b \in B$. This backing history, denoted by $\mathcal{H}_b$, contains the details such as a project's topical category, the geo-location of project, and the person who created the project.[2] For most part,

---

[2]In our experiments, we consider city and state as the geo-location of the projects.

this information remains static. Since it is extracted from the backing history of users, we call γ as *user-specific* prior. On the contrary, in the second case, the prior information σ is not static due to the transient nature of the projects (as explained in Section 1). As σ is changed at regular time intervals, we term it as *dynamic-status* prior. The calculation of these priors are detailed in the following sections.

5.1 User-Specific Priors

In this paper, we incorporate three features from the users' backing history to create the user-specific prior γ, namely: (a) topical preference (b) creator preference, and (c) geo-location preference.

Topical Preference: In our previous work [28], we observed that Kickstarter users have a strong topical preference in their decisions to back a project. We assume users have a tendency to continuously back projects in the same topical category. This tendency can be modelled as a conditional probability of a user b to back a project in topic t, given t is present in the backing history of this user, denoted by $P((b,t)|t \in \mathcal{H}_b)$. Using Bayes' theorem, it is derived as follows:

$$P((b,t)|t \in \mathcal{H}_b) = \frac{P(t \in \mathcal{H}_b|(b,t)) \, P(b,t)}{P(t \in \mathcal{H}_b)} \tag{12}$$

Creator Preference: Users tend to develop an inclination towards creators whom they have backed in the past. We represent this as the conditional probability of a user b to back a creator e, given that e is in the backing history of this user, denoted by $P((b,e)|e \in \mathcal{H}_b)$. It is calculated in a similar way as that of Equation (12).

Geo-location Preference: Geo-location has a strong impact on the success of projects, and the level of impact depends on the topical category of the project [28]. For instance, we find that projects based on technology are relatively less dependent on their geo-location, while projects on theatrical arts are highly dependent. Therefore, we incorporate this information as prior by calculating the probability of a user b to back a project v, given b and v are from the same geo-location. This probability is represented by $p((b,v)|Loc(b,v) = \ell, \tau(v) = t)$, where $Loc(b,v)$ indicates the geo-location of the backer and project, $\tau(v)$ is the topic of the project v. This probability is calculated using Bayes' theorem by expanding the likelihood using the chain rule of probability.

Finally, the interest of a user b towards a project v is obtained as a linear combination of topic-, creator-, and geo-location-based preferences as follows:

$$\gamma_{b,v} = p((b,t)|.) + p((b,e)|.) + p((b,v)|.) \tag{13}$$

5.2 Dynamic-Status Priors

We incorporate two factors in the prior σ for the project status ψ, namely, (1) the popularity of the project, and (2) the availability of popular rewards at specific time t.

Popularity of project: The popularity of a project at time t is derived as follows:

$$\mathcal{P}o_t(v) = \frac{\text{\# pledged amount of project } v \text{ at time } t}{\text{Goal amount of the project } v} \tag{14}$$

where $\mathcal{P}o_t(v)$ is the *popularity score* of project v at time t.

Availability of rewards: As explained earlier in Figure 1, the status of the project cannot be expressed purely based

on it's popularity, but the availability of rewards should also be considered. To verify our claim, we extract the top 3 popular reward categories of projects and obtain the additional percentage of backers backing the project on the day when the rewards are sold out. The result of this analysis is given in Figure 4, which shows that the interest of users in backing a project decays with the depletion of popular reward categories. Therefore, we calculate this prior information as follows:

$$\mathcal{R}_t(v) = \frac{\text{\# rewards sold out at time } t}{\text{\# limited rewards}} \qquad (15)$$

where $\mathcal{R}_t(v)$ is the *reward score* of the project v at time t. Using Equations (14) and (15), we define the dynamic-status prior for a project v as follows:

$$\sigma_{(v,t)} = \mathcal{R}_t(v) \times \mathcal{P}o_t(v) \qquad (16)$$

The score $\sigma_{(v,t)}$ is constantly updated at different time intervals $\{t1, t2..., tn\}$ until the project expires.
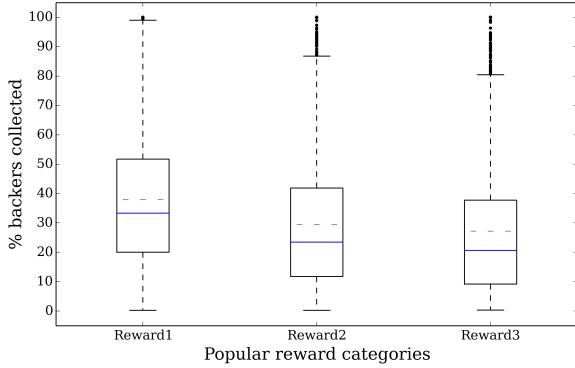
Figure 4: Decay of user interest with the depletion of popular reward categories.

6. EXPERIMENTS

In this section, we conduct comprehensive experiments to evaluate the performance of the CrowdRec model in comparison with state-of-the-art models. In the following, we first describe the datasets used in our experiments and then report the experimental results.

6.1 Dataset Description

For our experiments, we obtain six months of Kickstarter data (12/15/13 to 06/15/14) from *kickspy*, which consists of 27,270 projects. We remove projects that were canceled or suspended, with less than one backer, or with less than $100 of pledged amount. Next, we build a web-crawler to fetch backers from this filtered set of projects. As such, we obtain over 1 million backers for the remaining 18,143 projects after the removal process.

In our previous study [28], we show that backers in Kickstarter are distinguished by two important features: (a) the presence of social network profile,[3] and (b) the backing frequency. Therefore, in this paper, we leverage this information to categorize the backers $b \in \mathcal{B}$ into two types: (i) those who have linked their Kickstarter profiles to Twitter, and (ii) those without a Twitter profile. We

---

[3]we only consider Twitter profiles since Facebook data is not publicly available.

denote these two types by Twt and Kck, respectively. In addition, we classify the backers into *occasional* backers who have backed 2-10 projects (denoted by *Occ*) and *experienced* backers who have backed over 10 projects (denoted by *Exp*). Thus, we have four different datasets: (1) **Twt-Occ** = $\{b|b \in \text{Twt} \cap (2 < Backings(b) < 10)\}$; (2) **Twt-Exp** = $\{b|b \in \text{Twt} \cap (Backings(b) > 10)\}$; (3) **Kck-Occ** = $\{b|b \in \text{Kck} \cap (2 < Backings(b) < 10)\}$ and (4) **Kck-Exp** = $\{b|b \in \text{Kck} \cap (Backings(b) > 10)\}$

Group Creation. A group consists of backers who have backed the same project. For the four datasets described above, we create groups of Kickstarter users using the methodology described in [4], i.e., in our group creation process, we calculate the inner group similarity between the group members using Pearson correlation co-efficient (PCC) and filter out groups which have PCC less than 0.2. The statistics of our backer-group datasets are shown in Table 2.

Table 2: Statistics of Kickstarter groups formed by frequent and occasional backers.

Dataset	#Bkrs/Grp	#Grps	Avg. #Prjs/Grp	#Prjs
Kck-Occ	10	30,100	2	1,609
Kck-Exp	10	20,675	4	1,397
Twt-Exp	5	3,373	3	959
Twt-Occ	5	3,513	2	1,104

6.2 Performance Evaluation

In the following, we first discuss the performance metrics employed in our evaluation and a number of recommendation models examined for comparison. Our evaluation is performed over all the four datasets *Twt-Occ*, *Twt-Exp*, *Kck-Exp*, and *Kck-Occ* by randomly holding-off 20% of the ground truth for testing. For most of our experiments we set the parameters β_1 and β_2 to be 0.5, the topic parameter K to 200 and the dynamic-status prior $\sigma_{(p,t)}$ is calculated using the 50% of the total project duration. The CrowdRec model and all other baselines are implemented using python's numpy numerical module, and scikits machine learning module.[4] The codes of our model are publicly hosted in the Github page.[5]

6.2.1 Evaluation Metrics

To evaluate the performance of ranking, we use the standard information retrieval measures. For every project, we compute: 1) P@N: *precision at rank N* is defined as the fraction of rankings in which the true backers are ranked in the *top-N* positions, 2) MRR: The *mean reciprocal rank* is the inverse of the position of the first true backer in the ranked set of backers produced by a recommendation model, 3) S@N: The *success at rank N* is the probability of finding at least one true backer in the *top-n* ranked set, and 4) DCG: The *discounted cumulative gain* [18] is based on the fact that highly relevant backers are more important than marginally relevant ones.

6.2.2 Baseline Methods for Comparison

We compare the performance of our model with a simple collaborative filtering-based approach that uses various aggregation strategies for group recommendation and other

---

[4]http://scikit-learn.org
[5]https://github.com/magnetpest2k5/Crec

state-of-the-art group recommendation models as described below:

- **Collaborative Filtering with averaging (*CFA*)**: First we learn user-project preference using user-based collaborative filtering. We then take the average of the recommended scores for a group and rank the preference scores to recommend a project to the group.
- **Collaborative Filtering with least-misery strategy (*CFL*)**: First we learn the user-project preference using user-based collaborative filtering. We then take the least scores as the recommended scores for a group and rank the preference scores to recommend a project to the group.
- **Collaborative Filtering with relevance disagreement (*CFR*)**: First we learn user-project preference using user-based collaborative filtering, and then takes the relevance score as CFA and the disagreement is calculated as the difference between the preference scores of individuals within a group.
- **COM Model:** The state-of-the-art group recommendation model that does not include the dynamic-status component [36].

6.2.3 Experimental Results

Overall Performance: Figure 5 shows that the performance of CrowdRec and COM are distinctly better than the collaborative filtering-based group recommendation techniques. Although, the collaborative filtering with averaging technique (CFA) produces better results than CFL and CFR, the approaches that heuristically aggregate the individual scores of backers to determine the groups' preference towards projects do not produce reasonable results.[6] It is also clear that the CrowdRec performs better than COM model in terms of both precision and recall. This shows that users strongly rely on the on-going status information of projects to make backing decisions.
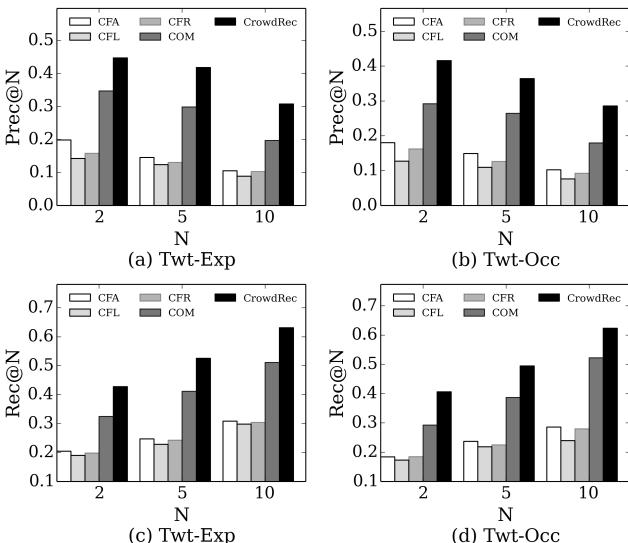
Figure 5: The precision and recall performance over experienced and occasional backers with Twitter profiles.

We also observe that the performance of CrowdRec over experienced backers (Figures 5(a) and 5(c)) is better than

---
[6]In [36] the authors report similar observations.

occasional backers (Figures 5(b) and 5(d)), because the former have a higher backing count, which provides a richer set of prior information about the backer's preference over topics, creators and geo-location compared to the occasional backers. Figure 6 shows similar set of results for the Kck dataset. In comparison to results shown in Figure 5, we see that the performance of our model slightly decreases in the dataset Kck. This is because the backers with Twitter profiles (i.e. dataset Twt) constantly receive Tweets about Kickstarter projects from their friends and followees, which leads to a better communication with their group members. Since one of the key components of our model is to effectively incorporate the groups' influence, it has a stronger impact over these type of users. Finally, in Table 3, we show the superior performance of CrowdRec over all other models by averaging the Success@N measure over all four datasets.
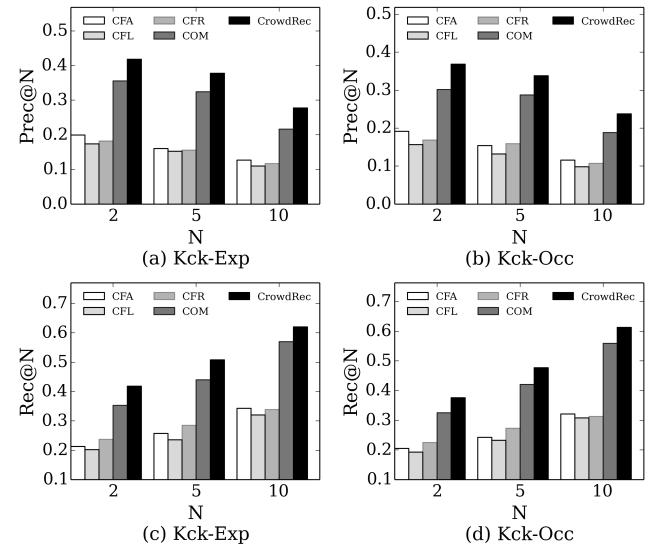

Figure 6: The precision and recall performance over experienced and occasional backers without Twitter profiles.

Table 3: The Average Performance over all datasets using Success @ N.

Model	Success@2	Success@5	Success@10
CFA	0.2638	0.3014	0.3352
CFR	0.2518	0.2857	0.3017
CFL	0.2698	0.3138	0.3369
COM	0.6347	0.7143	0.7584
CrowdRec	**0.6926**	**0.7436**	**0.7832**

Impact of Group Size: Depending on the number of backers in a group, a group can become more diverse or conservative in terms of their topical-preference. Therefore, we show the impact of group size on the performance of our model in Figure 7. Due to space constraints, we only show the recall performance for top 10 recommended projects. We again observe that CrowdRec performs better than COM and all other models irrespective of the group sizes. The performance of CrowdRec increases as we move from group size 2 to group size 5. However, this improvement becomes insignificant as the group size further increases. In fact, we observe that the performance slightly reduces for group sizes of 15 and above, mainly due to the sparsity of the data.[7]

[7]Groups greater than 15 members are extremely few in num-
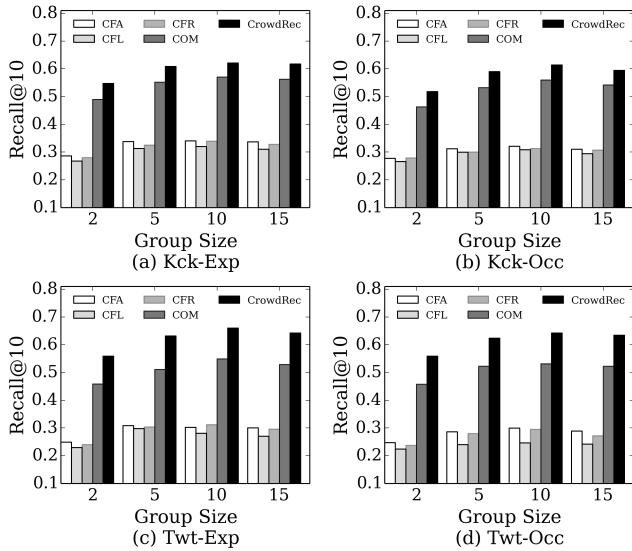
Figure 7: Effect of group sizes of Kickstarter users over the recall performance for all datasets.

Effect of Topic Size: To study the effect of topic size K over the performance of our model, in Figure 8, we plot the DCG scores of the top 10 recommended projects by varying K from 25 to 300. Similar to our prior observations, we see that the CrowdRec outperforms COM and all other models. Additionally, both CrowdRec and COM performs better on the backers with Twitter profiles (i.e. Twt-Exp and Twt-Occ) when compared to backers without Twitter profile (i.e. Kck-Exp and Kck-Occ). It is also important to note that the DCG values for experienced backers (Figure 8(a)) are higher than the occasional backers (Figure 8(b)), mainly because experienced backers have a much higher backing count (and thus richer prior information) than the occasional backers. Finally, we observe that the increase in the number of topics does not necessarily translate to a better performance. Although the DCG scores improves with the increase in topic count, the performance becomes static at about 200 topics. In fact, there is a slight decrease in the DCG scores when the topic count goes past 200.
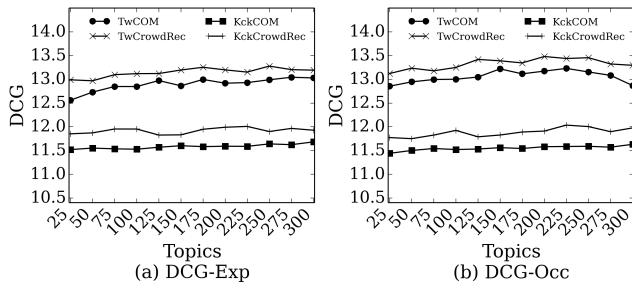

Figure 8: Effect of topics on the DCG measure.

Effect of Prior Information: Figure 9 shows the effect of various user- and project-based features that were used as priors in our model. In this figure, the y-axis denotes

ber.

the priors and the x-axis indicates the MRR scores of the top 10 recommended projects. We begin with a symmetric prior (*s-prior*) and gradually add other features to the prior distribution, which is indicated by the *+* symbol. For instance, Topic[+] implies that we are including the topical-preference of users, which was calculated in Equation (12); similarly, Cr[+] implies we add creator preference into the prior information. We observe that simply by including the topic prior provides a significant boost to the MRR scores indicating that backers strongly depend on their topical interest to fund a project. Although, the addition of creator (*Cr*) and geo-location (*Geo*) preferences of backer improves the performance of the model, though it is not as significant as the topical-preference. Finally, the inclusion of the popularity prior (*Pop*) provides a significant boost yet again, which shows the importance of the information about the on-going status of the Kickstarter projects.
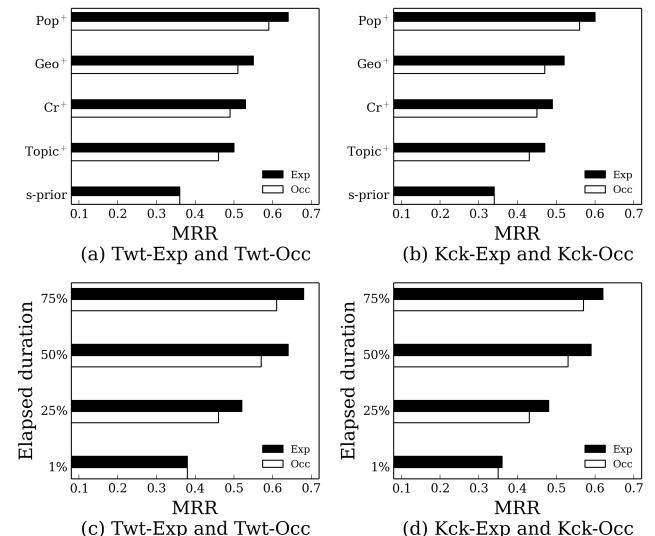

Figure 9: Effect of prior information ((a) and (b)) and project duration ((c) and (d)) on the recommendation performance. The terms *Topic*, *Cr*, *Geo* and *Pop* indicates the topical, creator, geolocation and popularity based priors, respectively.

Effect of Dynamic Status: Lastly, we show the effect of dynamically varying prior information in Figures 9(c) and 9(d). We calculate the status prior $\mathcal{R}_t$ using the Equation (14) at various intervals of the projects' duration ranging from 1% to 75% of the total project duration, which is indicated by the y-axis of this figure. It can be seen that, in general, the recommendation performance increases with the progression of the project. This is because, as the project progresses, we can obtain a much accurate estimation about it's status both in-terms of popularity and the availability of rewards.

7. CONCLUSION

In this paper, we introduced a recommendation framework for a popular crowdfunding platform, i.e., Kickstarter. We point out the challenges arising in Kickstarter, where the backing habits of its users depend on a diverse set of features, including topical, geo-location, temporal, and social traits. By exploiting the notion of groups, we proposed a recommendation model that effectively incorporates all these features when recommending projects to groups of Kickstarter

users. Using a real dataset, we conducted a comprehensive evaluation to show that our model outperforms other state-of-the-art group-recommendation models in terms of a variety of performance metrics. Finally, we also studied the impact of various prior information and show that the ongoing status (or popularity) of the projects plays an important role in improving the recommendation performance.

Acknowledgements

8. REFERENCES

[1] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):734–749, 2005.

[2] A. Ahmed, Y. Low, M. Aly, V. Josifovski, and A. J. Smola. Scalable distributed inference of dynamic user interests for behavioral targeting. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 114–122, 2011.

[3] J. An, D. Quercia, and J. Crowcroft. Recommending investors for crowdfunding projects. In *Proceedings of the 23rd international conference on World wide web*, pages 261–270, 2014.

[4] L. Baltrunas, T. Makcinskas, and F. Ricci. Group recommendations with rank aggregation and collaborative filtering. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 119–126, 2010.

[5] D. Billsus and M. J. Pazzani. User modeling for adaptive news access. *User modeling and user-adapted interaction*, 10(2-3):147–180, 2000.

[6] D. Blackwell and J. B. MacQueen. Ferguson distributions via pólya urn schemes. *The annals of statistics*, pages 353–355, 1973.

[7] Y. Cha, B. Bi, C.-C. Hsieh, and J. Cho. Incorporating popularity in topic models for social network analysis. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 223–232, 2013.

[8] K. Chen, T. Chen, G. Zheng, O. Jin, E. Yao, and Y. Yu. Collaborative personalized tweet recommendation. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 661–670, 2012.

[9] W.-Y. Chen, J.-C. Chu, J. Luan, H. Bai, Y. Wang, and E. Y. Chang. Collaborative filtering for orkut communities: discovery of user latent behavior. In *Proceedings of the 18th international conference on World wide web*, pages 681–690, 2009.

[10] W.-Y. Chen, D. Zhang, and E. Y. Chang. Combinational collaborative filtering for personalized community recommendation. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 115–123, 2008.

[11] J. Choo, C. Lee, D. Lee, H. Zha, and H. Park. Understanding and promoting micro-finance activities in kiva. org. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 583–592, 2014.

[12] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin. Combining content-based and collaborative filters in an online newspaper. In *Proceedings of ACM SIGIR workshop on recommender systems*, volume 60, 1999.

[13] V. Etter, M. Grossglauser, and P. Thiran. Launch hard or go home!: predicting the success of kickstarter campaigns. In *Proceedings of the first ACM conference on Online social networks*, pages 177–182, 2013.

[14] E. M. Gerber, J. S. Hui, and P.-Y. Kuo. Crowdfunding: Why people are motivated to post and fund projects on crowdfunding platforms. In *CSCW Workshop*, 2012.

[15] T. Hofmann. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 50–57, 1999.

[16] T. Hofmann. Unsupervised learning by probabilistic latent semantic analysis. *Machine learning*, 42(1-2):177–196, 2001.

[17] J. S. Hui, M. D. Greenberg, and E. M. Gerber. Understanding the role of community in crowdfunding work. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 62–74, 2014.

[18] K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446, 2002.

[19] KickStarterStats. Kickstarter stats. *Kickstarter* https://www.kickstarter.com/help/stats. accessed 06-01-2015.

[20] V. Kuppuswamy and B. L. Bayus. Crowdfunding creative ideas: The dynamics of project backers in kickstarter. *SSRN Electronic Journal*, 2013.

[21] Y. Li, V. Rakesh, and C. K. Reddy. Project success prediction in crowdfunding environments. In *In Proceedings of the 9th ACM International Conference on Web Search and Data Mining*, 2016.

[22] X. Liu, Y. Tian, M. Ye, and W.-C. Lee. Exploring personal impact for group recommendation. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 674–683, 2012.

[23] 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, 2014.

[24] B. Marlin. *Collaborative filtering: A machine learning perspective*. PhD thesis, University of Toronto, 2004.

[25] J. F. McCarthy. Pocket restaurantfinder: A situated recommender system for groups. In *Workshop on Mobile Ad-Hoc Communication at the 2002 ACM Conference on Human Factors in Computer Systems*, 2002.

[26] E. Mollick. The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing*, 29(1):1–16, 2014.

[27] A. Popescul, D. M. Pennock, and S. Lawrence. Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 437–444, 2001.

[28] V. Rakesh, J. Choo, and C. K. Reddy. Project recommendation using heterogeneous traits in crowdfunding. In *Ninth International AAAI Conference on Web and Social Media*, 2015.

[29] A. Sadilek, H. Kautz, and J. P. Bigham. Finding your friends and following them to where you are. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 723–732, 2012.

[30] X. Su and T. M. Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009:4, 2009.

[31] T. Tran and R. Cohen. Hybrid recommender systems for electronic commerce. In *Proc. Knowledge-Based Electronic Markets, Papers from the AAAI Workshop, Technical Report WS-00-04, AAAI Press*, 2000.

[32] V. Vasuki, N. Natarajan, Z. Lu, and I. S. Dhillon. Affiliation recommendation using auxiliary networks. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 103–110, 2010.

[33] J. Wang, Z. Zhao, J. Zhou, H. Wang, B. Cui, and G. Qi. Recommending flickr groups with social topic model. *Information retrieval*, 15(3-4):278–295, 2012.

[34] 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 *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 591–600, 2014.

[35] M. Ye, X. Liu, and W.-C. Lee. Exploring social influence for recommendation: a generative model approach. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 671–680, 2012.

[36] Q. Yuan, G. Cong, and C.-Y. Lin. Com: a generative model for group recommendation. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 163–172, 2014.

[37] W. Zhang, J. Wang, and W. Feng. Combining latent factor model with location features for event-based group recommendation. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 910–918, 2013.