Expert Systems With Applications 111 (2018) 100–106

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Mining social lending motivations for loan project recommendations

Jiaqi Yan[a], Kaixin Wang[a], Yi Liu[b], Kaiquan Xu[c,*], Lele Kang[a], Xi Chen[c], Hong Zhu[c]

[a] School of Information Management, Nanjing University, Nanjing 210046, China
[b] Rennes School of Business, Rennes 35065, France
[c] School of Business, Nanjing University, 22#, Hankou Road, Nanjing 210093, China

ARTICLE INFO

Article history:
Received 30 April 2017
Revised 14 August 2017
Accepted 5 November 2017
Available online 6 November 2017

Keywords:
Project recommendation
Data analytics
Social lending
Lending motivation
Big data
Text data

ABSTRACT

Online social lending has facilitated the ability of borrowers to reach lenders for financing support. With the increasing number of social lending projects, it is becoming very difficult for lenders to find appropriate projects to invest in, and for borrowers to get the funds they need. Project recommendation techniques provide a promising way to solve this problem to some degree, by recommending borrowers' projects to lenders who are able to invest. Unfortunately, current loan project recommendations only explore some structured information to match borrowers and lenders, so they cannot achieve a satisfactory way to solve the problem very well. In this study, we innovatively mine a huge amount of unstructured data, the text data of borrowers' and lenders' motivations, to provide loan project recommendations that solve the problem of mismatches between borrowers and lenders. We present a motivation-based recommendation approach that uses text mining and classifier techniques to identify borrowers' and lenders' motivations. Using a dataset from the well-known social lending platform Kiva, our experiment results show that, compared with prior works, the proposed approach improves project recommendations in inactive lender groups and unpopular loan groups, which shows the superiority of the proposed approach in addressing data sparsity and cold start problems in loan project recommendations. This study thus initiates an attempt to solve the information overload problem and improve matching between borrowers and lenders through mining big unstructured text data found in a large number of P2P platforms.

1. Introduction

Online social lending has experienced significant growth since its introduction in 2005. It is now an essential financing method for individual borrowers to reach individual lenders. Lenders decide for themselves which loan projects they are interested in supporting (Herzenstein, Dholakia, & Andrews, 2011). However, the sustainability of social lending institutions has been questioned because the demand for funding services outstrips the supply (Bogan, 2012), and the problems caused by unmatched expectations and information exchange between borrowers and lenders that can affect the repayment issues (Cadsby, Du, Song, & Yao, 2015) and trust (Duarte, Siegel, & Young, 2012). In particular, with the increasing number of social lending projects, lenders are often overwhelmed with project information, and cannot find the projects that might interest them. Many loan projects fail to reach their funding goals, and, in fact, statistics show that only 43% of projects succeed in reaching their pledge goals (Rakesh, Choo, & Reddy, 2015). Thus, improving the success rate of fund collection and contributing to the sustainability of social lending institutions have become important issues in the social lending literature (Malekipirbazari & Aksakalli, 2015).

A recommender system is one solution that can alleviate the information overload problem by providing lenders with personalized information (Adomavicius & Tuzhilin, 2005). Recommenders have been used in various areas, such as online product recommendation (Zhang, Liu, & Zeng, 2017), travel destination recommendation (Binucci, De Luca, Di Giacomo, Liotta, & Montecchiani, 2017), and researcher collaboration recommendation (Xu et al., 2012). In the social lending context, loan project recommendations can help lenders find loan projects that interest them, by aligning the right loan projects with the right lenders (Choo, Lee, Dilkina, Zha, & Park, 2014; Zhao, Wu, Liu, Ge, & Chen, 2014). However, as compared with traditional product recommendation, there are several challenges in designing recommender systems for loan projects on social lending platforms (Choo et al., 2014; Zhao et al., 2014). First, big data should be used in project recommendation. Given the rapid growth in the social lending business, there are more than 6000 P2P platforms in China alone (Wangdaizhijia, 2017), and the number of lenders and borrowers on some platforms reaches one million (Kiva, 2017). Hence, a huge amount of information is generated in online social lending,

* Corresponding author.
E-mail addresses: jiaqiyan@nju.edu.cn (J. Yan), yi.liu@esc-rennes.com (Y. Liu), xukaiquan@nju.edu.cn (K. Xu), lelekang@nju.edu.cn (L. Kang).

https://doi.org/10.1016/j.eswa.2017.11.010
0957-4174/

which can be used for project recommendation. In addition to the high volume of data, the variety of data, especially unstructured data, is also very high. Lenders and borrowers prefer to describe their objectives with text data, and these text data contain rich information for project recommendation. Second, lenders' funding decisions are affected by many factors, which are always captured as unstructured data that must be processed to be used in a recommender system. Hence, recommendations cannot be based on some simple sets of straightforward features that are directly available from the projects. For instance, borrowers' photographs may affect lenders' perception of the reliability of loan receivers. As a result, categorizing the photographs becomes an analytical issue to be dealt with so that the information can be used in project recommendation. If there are analytical findings suggesting that some lenders may favor lighter-skinned and less obese borrowers (Jenq, Pan, & Theseira, 2015), then the problem is addressed by classifying the photographs accordingly. Similarly, rich text information in social lending, such as loan descriptions and lender self-state descriptions, must be analyzed for project recommendation. Third, analytical findings need to be incorporated into project recommendation to address data sparsity and cold start problems. Traditional recommendation techniques, such as collaborative filtering, assume that users would probably choose items that similar users like, and the recommended items are reusable, that is, an item can serve many users for several years. This is not the case for online social lending, as a project can only be recommended until the opportunity expires (Choo et al., 2014). In other words, the project data cannot be reused in recommendations and the lender data may also be sparse, which leads to a cold start problem in project recommendation.

To address these challenges in loan project recommendation, we propose to recommend loan projects based on analyzing the big data of lenders and borrowers. In particular, we argue that the alignment of motivations on why borrowers need financial support and why lenders fund projects is important for project recommendations, and we propose to improve project recommendations based on data analytics of the huge amount of text data addressing borrowers' and lenders' motivations. In this paper, we present a motivation-based recommendation system on the Kiva platform. We employ text classifiers to classify the lenders' self-stated motivations, and incorporate this classification of lenders' motivations into the project recommender system based on word embedding techniques.

The paper is organized as follows. In the next section, related studies of project recommendation and the motivations for crowdfunding participation are introduced. Section 3 presents a system framework for motivation-based project recommendation. In Section 4, we apply the proposed approach in the data analytics of lenders' motivations on Kiva, and present a recommender approach based on the analytical results. We conclude the work in Section 5.

2. Related studies

2.1. Project recommendation in online lending

Online social lending platforms could use relevant information (e.g., previous or ongoing loan projects, lender profile, etc.) to give lenders personalized recommendations to increase the success rates of loan projects (Belleflamme, Omrani, & Peitz, 2015). A recommendation system is likely to be the most visible and useful personalization tool that can be used to enhance the experience of online service users (Kumar & Benbasat, 2006). For project recommendation in online lending, platform designers employ all available resources to display appropriate funding projects to potential lenders. The objective of the recommendation system is to reduce the uncertainty for lenders and increase the probability of lending actions.

For profit-oriented platforms, Rakesh et al. (2015) studied how various features influence recommendations. They grouped features into four categories and built a supervised learning framework to learn a model for recommendations. Rakesh, Lee, and Reddy (2016) proposed a probabilistic model called *CrowdRec* to recommend projects by incorporating features including the status of projects, the lenders' preferences, and the collective preferences of the group. Experiment results show that their model performs better than the state-of-the-art group recommendation model and some collaborative filtering algorithms. Some researchers also proposed methods for recommendations with multiple objectives (Zhao, Liu, Wang, Ge, & Chen, 2016). Malekipirbazari and Aksakalli (2015) developed a Random Forest–based classification method to predict borrower status, which performed better than FICO credit scores.

For nonprofit platforms like Kiva, Choo et al. (2014) proposed a novel feature extraction method using joint nonnegative matrix factorizations for loan project recommendation. As common recommendation methodologies may favor some loans over others, Lee et al. (2014) suggested a fairness-aware recommendation system to diversify the distribution of donations to reduce the inequality of loans. Experiment results show that with little AUC (Area Under the Curve, which measures how much higher positive samples are ranked than negative samples) value sacrifice, their model can provide a significant reduction in the standard deviation of recommendations, which shows improvement in the fairness of loan project recommendation.

2.2. Motivations for crowdfunding participation

The importance of studying individuals' motivations is widely recognized in various online contexts. Generally, motivational theories argue that motivation serves as a driver of behavior and captures individuals' willingness to act. Motivations are generally categorized into intrinsic and extrinsic motivations, which could coexist and simultaneously influence individuals' behavior (Vallerand, 1997). Intrinsic motivation is defined as the psychological force arising from the individual's innate needs (Deci, 1975). Intrinsic motivation relates to individuals' perception of pleasure, satisfaction, and interest derived from the task itself (Deci & Ryan, 2000). Meanwhile, extrinsic motivation is defined as motivation that comes from outside the individual, and is instrumental in achieving valued outcomes that are distinct from the activity (Venkatesh, 1999). Extrinsic motivation relates to gaining rewards or avoiding punishment by performing a behavior. The motivations of individuals can be analyzed to predict their behaviors and preferences. For instance, online gamers are segmented based on their motivations, which could predict an intent to purchase virtual assets in online games (Tseng, 2011), and the motivations of social media users have been found to affect their social media marketing responses (Chi, 2011). In the context of online lending, both of these theoretical motivations are helpful in predicting lenders' funding behavior.

In contrast to profit-oriented social lending platforms where lenders expect to receive monetary benefits from loan projects, not-for-profit social lending platforms, such as Kiva, attract lenders with different motivations. Lenders perceive the online social lending service on Kiva as a way to support certain kinds of entrepreneurial, responsible individuals with neoliberal views (McKinnon, Dickinson, Carr, & Chávez, 2013). Through analyzing lenders' self-stated motivations on Kiva, lenders' motivations for crowdfunding participation have been categorized into (1) general altruism, (2) group-specific altruism, (3) empathy, (4) reciprocity, (5) quality and social safety net, (6) social responsibility and social

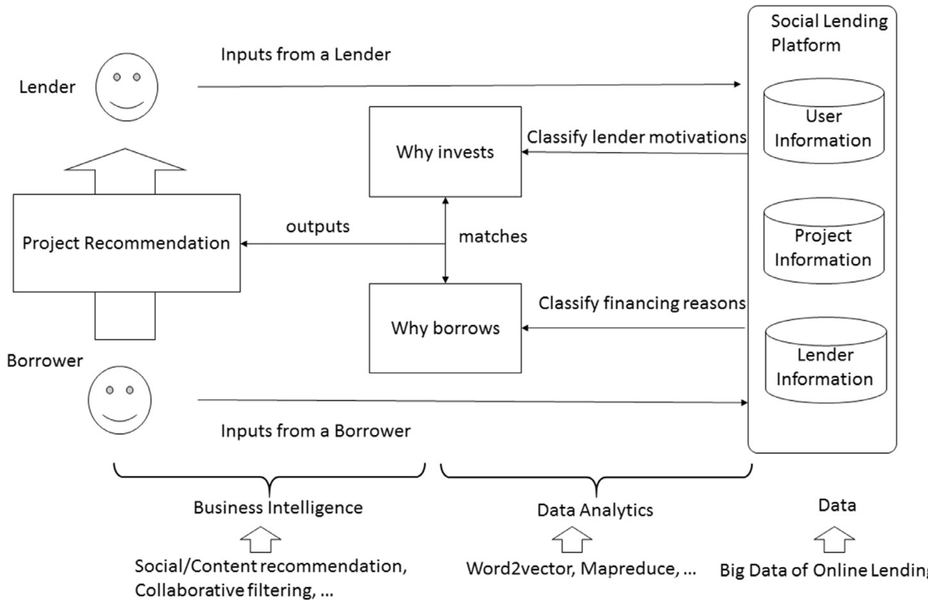

Fig. 1. The motivation-based recommendation framework.

norms, (7) effective development tool, (8) personal satisfaction, (9) religious duty, and (10) external reasons (Liu, Chen, Chen, Mei, & Salib, 2012). All of these classified motivations could capture both intrinsic and extrinsic motivations, and could be used to predict lending behavior. For instance, lenders motivated by religious duty or effective development tool make more loans, while lenders motivated by general altruism, group-specific altruism, or external reasons make fewer loans (Liu et al., 2012). In another study, the motivations of lenders from Team Canada on Kiva are realized via analyzing lenders' self-stated motivations. The desire to make a difference in the lives of families to alleviate poverty is the most popular motive (Mittelman & Rojas-Mendez, 2013). Other popular motives include altruism, social justice, and existential guilt. Egoism, sympathy, religious commitment, and empathy are also identified as less popular motives (Mittelman & Rojas-Mendez, 2013).

The motivations of borrowers could be identified in the *loan description* section on social lending platforms. The represented motivations will be read and evaluated by lenders (Dorfleitner et al., 2016), whose decisions will be determined by their perception of the borrowers' motivations. Thus, the motivations in loan description influence funding success (Feng, Fan, & Yoon, 2015) and default rate (Serrano-Cinca, Gutiérrez-Nieto, & López-Palacios, 2015). In addition to the identified motivations, other characteristics of *loan description* may also influence funding success, including length of the description (Greiner & Wang, 2011), quality (Chen, Lai, & Lin, 2014), and concreteness (Larrimore, Jiang, Larrimore, Markowitz, & Gorski, 2011).

Although the motivations of lenders and borrowers have been separately recognized as influential factors in the funding success of loan projects, the alignment of both parties' (i.e., lenders and borrowers) motivations has not been examined (Bachmann et al., 2015). Hence, it may be possible to improve the funding success rate by matching lenders and borrowers with a recommendation system based on their respective motivations.

3. A framework for motivation-based project recommendation

To recommend a project based on the alignment of lenders' and borrowers' motivations, we propose a framework for motivation-based project recommendation as shown in Fig. 1. It consists of three interrelated components: big data, data analytics, and business intelligence. The collection of big data from social lending is the foundation of data analytics to understand lending motivations, while the results of data analytics provide business intelligence that leads to better project recommendations. In the following, we explain these three components in turn.

- Big Data: A large dataset is the foundation of the whole process. Online lending platforms collect big data and are best positioned to use it efficiently. The databases of online social lending platforms are the collections of borrower information, loan project information, and lender information. Borrowers convey relevant information about themselves and the characteristics of loan projects, while lenders search their credit information and screen the loan applicants. The availability of big data, retrieved from various data sources, with different data formats and sizes, can provide a more complete picture of borrowers (Yan, Yu, & Zhao, 2015), while information about lenders provides an opportunity to understand their preferences in project investment.
- Data Analytics: Data analytics provide the opportunity to discover insights from big data. In this research context, the objective of data analytics is to find out why borrowers borrow online and lenders lend online. On social lending platforms, data analytics have been applied to discover knowledge about borrowers (Lin, Prabhala, & Viswanathan, 2013), loan projects (Herzenstein, Sonenshein, & Dholakia, 2011), and lenders (Lee & Lee, 2012). To understand the motivations of lenders and borrowers, we need to perform data analytics on the data available from online social lending platforms. Borrower motivations could be mainly identified from their loan description, while lender motivations could be recognized by their previous behavior. In this step, some big data techniques are adopted, such as word2vector, Mapreduce, and so on.
- Business Intelligence: Business intelligence is a tool used to recommend lending projects to lenders by matching the motivations of both borrowers and lenders. The goal of the business intelligence tool in our research is to encourage lenders to fund projects and improve loan performance. The business insights from data analytics on social lending have generated great opportunities by using business intelligence to guide project recommendation.

4. An experiment on Kiva.org

In this section, we apply our proposed motivation-based recommendation framework to the social lending platform Kiva.org. Kiva is a non-profit micro-finance organization that acts as an intermediary service to attract people to lend money to underprivileged borrowers. Its lending model is based on a social lending model in which any individual can fund a particular loan by contributing to a loan individually or as a part of a lender team.

Project recommendation plays a significant role in attracting people involved in non-profit lending behavior (Choo et al., 2014). In selecting charitable loan projects, most lenders do not focus on the profit return. Instead, lenders are more interested in why the borrowers need financial support (Herzenstein, Dholakia et al., 2011; Herzenstein, Sonenshein et al., 2011). We believe such a context can provide a suitable demonstration for the application of the motivation-based recommendation framework. Kiva.org provides a big data set of heterogeneous information about borrowers, loan projects, lenders, funding time, and other micro-financing information. We access Kiva's data through their daily snapshots and the API. For a demonstration of our proposed approach, we focus only on mining the rich textual features on Kiva (*loan because* on lender's side and *loan description* on borrowers' side) to provide project recommendation. The experiment consists of three major stages: textual features preprocess, lender motivation mining, and project recommendation.

4.1. Textual features preprocess

To process the text data, we average embedding word vectors to obtain a fixed-length vector of a sentence rather than the traditional bag-of-word technique used in prior project recommendations (Choo et al., 2014), which we believe can exploit deeper semantic relations for our textual feature preprocessing. We simply average word vectors to obtain the vector of a sentence, and use pre-trained word vectors trained on Twitter corpus instead of news or Wikipedia, as the lenders' motivation statements or loan descriptions are less formal and have more similarities with Tweets.

Before computing vectors, we tokenize sentences, remove punctuation, convert English words to lower case, and obtain a list of tokens for each training example with the help of the Python nltk package. The reason to convert words to lower case is that the word vectors data only support lower case words, even though this may result in the loss of some information. Then, we retrieve the corresponding vector for each token and average them to get sentence vector. Tokens that are not in the Twitter vocabulary are simply ignored. We notice there are a lot of short sentences that are not helpful for recommendations, because they appear in many training examples and have meanings that are too broad, such as "I can," "I help," and "It is great." Thus, we discard training examples whose text features have less than 5 tokens. As we have only 894 hand-coded examples for training motivation classifier, we decide to use 100-dimension word vectors to make a tradeoff between performance and overfitting issues. Either the *loan because* feature or the *loan description* feature can then be transformed into a 100-dimension vector. No dimension reduction technique is applied.

4.2. Mining Kiva users' motivations

Mining users' motivations is to discover users' loan or borrow reasons by classifying lenders' self-stated motivation and borrowers' loan description. For a demonstration of our proposed approach, we adopt the classification framework of lending motivation proposed by Liu et al. (2012) to classify Kiva lenders' motivations.

Table 1
Motivation classification results.

Motivation Category	$F_{0.5}$	Precision	Recall
General altruism	0.61	0.58	**0.78**
Empathy	0.50	0.57	0.33
Reciprocity	0.56	0.55	**0.60**
Equality and social safety net	**0.82**	**0.94**	**0.56**
Social responsibility and social norms	**0.86**	**0.93**	0.67
Effective development tool	0.59	0.58	**0.60**
Personal satisfaction	0.53	0.56	0.45
Religious duty	0.68	0.69	0.65
External reasons	**0.73**	0.70	**0.88**

Table 2
Motivation classification results using resampling.

Motivation Category	$F_{0.5}$	Precision	Recall
General altruism	0.51	0.75	0.22
Empathy	0.19	0.16	0.67
Reciprocity	0.23	0.20	0.70
Equality and social safety net	0.36	0.33	0.48
Social responsibility and social norms	0.64	0.65	0.62
Effective development tool	0.50	0.50	0.51
Personal satisfaction	0.53	0.56	0.45
Religious duty	0.61	0.75	0.35
External reasons	0.50	0.67	0.25

Three of our researchers each manually classify 3000 lenders' loaning reasons into motivation categories. Then, we take the 894 unanimous results to serve as the training data. Next, we use techniques discussed in Section 4.1 to obtain feature vectors. We employ several classification algorithms to train a motivation classifier, including the SVM (support vector machine) used in Liu et al. (2012). We find that a soft-voting ensemble of an LDA (linear discriminant analysis) model and a logistic regression model can give a comparably good result as in Liu et al. (2012), while simply using the SVM or other classification algorithms does not work well. We use a one-vs-rest scheme and the results are listed in Table 1.

Our model performs better in some categories but worse in others compared with prior work on motivation classification (Liu et al., 2012). There are two possible explanations. First, we have only half hand-coded data. With more data, our model would probably work better. Second, our data is severely imbalanced, which may account for poor performance. For example, we have only 40 examples each in the "Empathy" and "Reciprocity" categories out of 894 training examples. To address the imbalance issue, the resampling technique is used, and the results appear in Table 2. It seems the results are mixed. Although our motivation classifier performs poorly in some categories, it can still give acceptable predictions.

4.3. Kiva project recommendation

We formulate the recommendation problem as a binary classification problem according to prior work on project recommendation (Choo et al., 2014). Although there are a lot of features that would affect the recommendation, we only exploit textual data about lending motivation to demonstrate the motivation-based recommendation approach. We employ the trained motivation classifier to generate motivation feature and represent it in a one-hot scheme. Next, the Random Forest algorithm (Breiman, 2001) is employed to train the data using the *loan because* and *loan description* features.

We group lenders and loans into 4 groups according to their activities and train a classifier on each group. Then, we train a classifier on each combination of 16 and discover some interesting results. We compare the results obtained by directly using the

Table 3
Lender groups and loan groups.

Lender				
Group	1	2	3	4
Range (# of loans)	0–5	5–15	15–50	50–Max
Count	44,059	40,436	30,326	14,476
Loan				
Group	1	2	3	4
Range (# of lenders)	0–5	5–15	15–35	35–Max
Count	205,554	224,137	63,437	11,448

Table 4
AUC results.

AUC	LoanGroup1	LoanGroup2	LoanGroup3	LoanGroup4
LenderGroup1	0.6832	0.7187	0.6983	0.6342
LenderGroup2	0.6628	0.6767	0.6536	0.6258
LenderGroup3	0.6578	0.6753	0.6617	0.6167
LenderGroup4	0.6375	0.6566	0.6385	0.5886

Table 5
Precision and recall results.

	LoanGroup1		LoanGroup2		LoanGroup3		LoanGroup4	
	Prec	Recalll	Prec	Recalll	Prec	Recalll	Prec	Recalll
	0.6628	0.6375	0.6628	0.6375	0.6628	0.6375	0.6628	0.6375
LenderGroup1	0.6929	0.6470	0.7451	0.6647	0.7033	0.6710	0.6390	0.6127
LenderGroup2	0.6556	0.6741	0.6726	0.6832	0.6541	0.649	0.6387	0.5906
LenderGroup3	0.6492	0.6645	0.6700	0.6875	0.6735	0.6346	0.6272	0.5903
LenderGroup4	0.6349	0.6603	0.6573	0.6681	0.6389	0.634	0.5878	0.5756

loan because and *loan description* vectors to train the model. The comparison shows that motivation classification improves the AUC value by 0.03 on average. Another thing to mention is that the model trained without motivations still performs better than prior work on project recommendation (Choo et al., 2014), thereby proving the advantages of embedding word vectors that utilize deeper semantic relations. We will explain the details in the following sections.

4.3.1. Data preparation

First, we process XML files downloaded from Kiva's website and generate a loan table and a lender table. Next, lenders with empty *loan because* features or loans with empty *loan description* features are dropped. Then, we group both lenders and loans into 4 groups according to their activities/popularities (measured respectively by the number of loans he/she has made and by the number of lenders who funded this loan). The group results are presented in Table 3.

Now we have $4 \times 4 = 16$ combinations of lender–loan groups. For each combination, there will be a matrix indicating whether a particular lender funded a particular loan. We randomly select 5000 positive examples and 5000 negative examples for each combination. So we have 16 training sets in total, and each of them comprises 10,000 examples. Each example has a *loan because* textual feature, a *loan description* textual feature, and a target 0/1 label indicating whether the lender funded the loan.

4.3.2. Transforming and training

We use techniques discussed in Section 4.1 to transform textual features into vectors. After transformation, each training set has 201 dimensions. Then, we apply trained motivation classifiers to predict motivation using the *loan because* vector feature and represent the predicted category in a one-hot scheme, which gives us 210-dimension training sets. Finally, we drop the 100-dimension *loan because* vector feature and use remaining features to train the model. Fig. 2 shows a specific example of generating the 109-dimension feature vector.

We find that the Random Forest algorithm performs better on our datasets than Gradient Boosting, which is used in other literature about recommendations in online lending. As the Random Forest algorithm randomly chooses a feature subset when splitting a node, we run the algorithm on each dataset 50 times to obtain average metrics. The AUC metric, which measures how much higher positive samples are ranked than negative samples, is used to evaluate the model. Also, precision and recall metrics are used for evaluating the performances. The experimental results are listed in Tables 4 and 5 and Fig. 3.

Compared with AUC results in prior literature (Choo et al., 2014), our model displays better performance. We also find some interesting facts about loan groups. From Fig. 2, we can see best performance in mid-popularity loan groups across all 4 lender groups. On the other hand, we find that the more active a lender is, the harder it is to predict his/her behavior based on motivations.

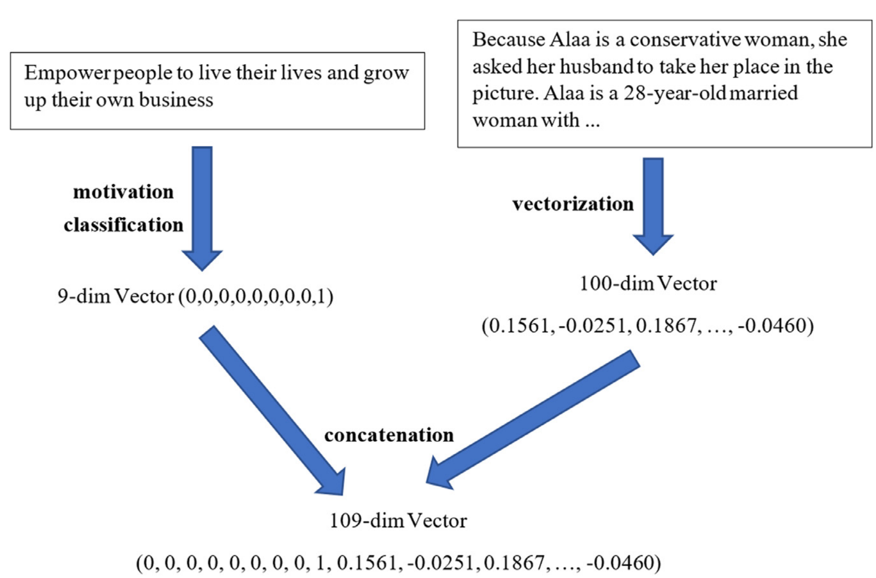

Fig. 2. An example of generating the 109-dimension feature vector.

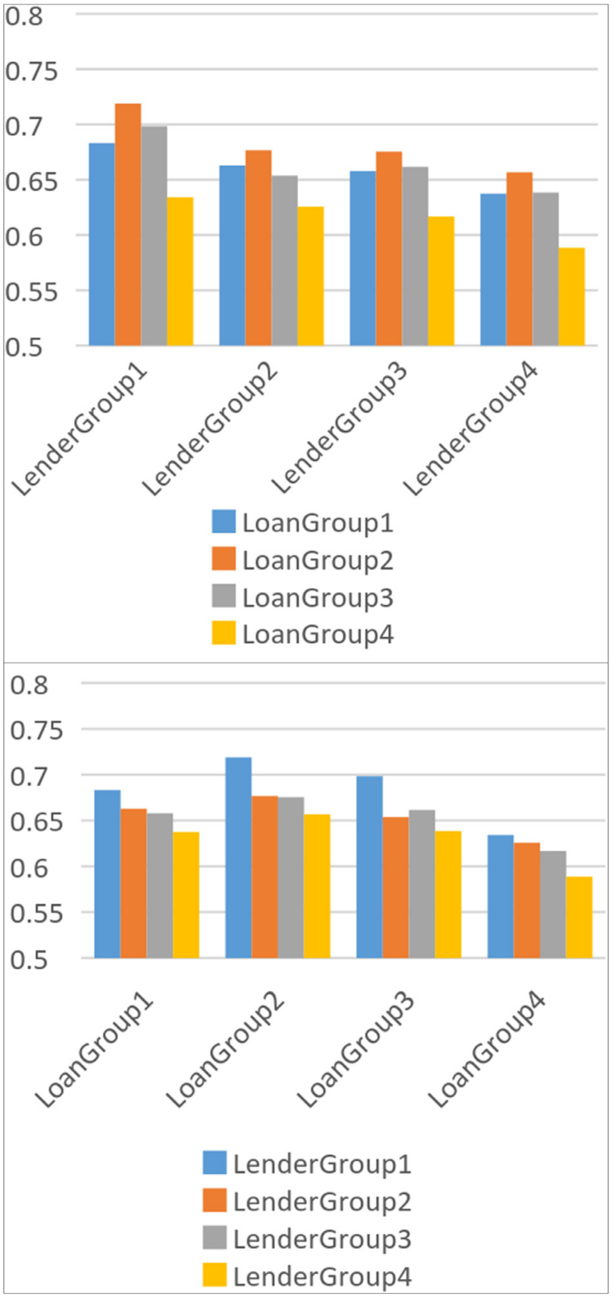

Fig. 3. AUC results over 0.5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Table 6
AUC results of our model using different grouping strategy.

Group	m = 5	m = 10	m = 15	m = 20	m = 25	m = 50
AUC	0.7114	0.6839	0.7011	0.7021	0.6925	0.6862

This is slightly different from the findings in Choo et al. (2014), who report that mid-active lenders' behaviors are hardest to predict. To further investigate this difference, we regroup lenders without grouping loans. We run experiments on each training set 50 times to obtain average metrics. Results are presented in Table 6 and illustrated in Fig. 3. In Fig. 3, the blue bins show the results in Choo et al. (2014). The orange bins show the difference compared to our model. The results indicate that our model performs almost equally well in 6 groups, and the model has better performance over Choo et al. (2014), especially in groups where 10 <= m <= 25.

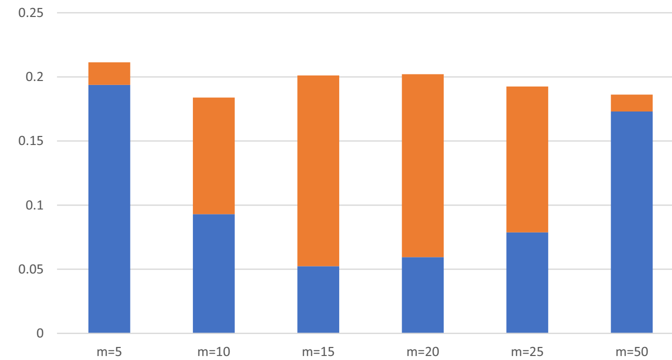

Fig. 4. AUC results over 0.5 of our model using different grouping strategy.

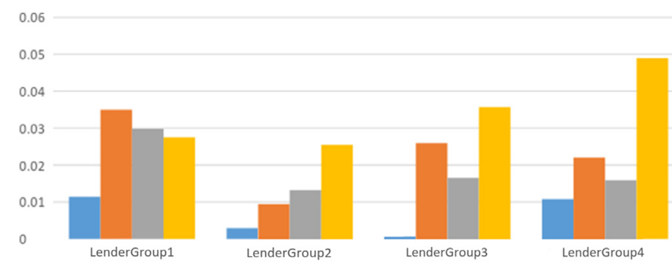

Fig. 5. AUC results of our model using motivation over directly using vectors.

Table 7
Comparison between using motivation and directly using vectors.

	Loan Group 1		Loan Group 2		Loan Group 3		Loan Group 4	
	Motive	Vector	Motive	Vector	Motive	Vector	Motive	Vector
Lender1	**0.6832**	0.6557	**0.7187**	0.6889	**0.6983**	0.6633	**0.6342**	0.6228
Lender2	**0.6628**	0.6373	**0.6767**	0.6635	**0.6536**	0.6442	**0.6258**	0.6229
Lender3	**0.6578**	0.6221	**0.6753**	0.6588	**0.6617**	0.6358	**0.6167**	0.6161
Lender4	**0.6375**	0.5885	**0.6566**	0.6407	**0.6385**	0.6165	**0.5886**	0.5778

Table 8
Comparison between using motivation and directly using vectors (using the gradient boosting tree algorithm).

	Loan Group 1		Loan Group 2		Loan Group 3		Loan Group 4	
	Motive	Vector	Motive	Vector	Motive	Vector	Motive	Vector
Lender1	**0.6238**	0.6099	**0.5934**	0.5922	**0.5939**	0.5740	**0.5743**	0.5529
Lender2	0.6344	0.6348	0.5952	0.6161	0.6164	0.6248	**0.6014**	0.5711
Lender3	**0.6338**	0.6055	**0.5958**	0.5901	**0.5880**	0.5761	0.5676	0.5799
Lender4	**0.6162**	0.5799	0.6039	0.6063	**0.5872**	0.5819	**0.5555**	0.5412

We compare the results obtained by directly using *loan because* and *loan description* vectors to train the model, as shown in Fig. 4 and Table 7. The results indicate that classifying lenders' statements into different motivation categories can improve recommendation performance. As shown in Fig. 4, we can see the improvement is more significant in inactive lender groups and unpopular loan groups. In addition, another classifier, the Gradient Boosting Tree algorithm, is used, and the results (Table 8) also show better performance is obtained using motivation than not using it. (Fig. 5)

5. Conclusion

Although big data is available for project recommendation in social lending, loan project recommendation faces unique challenges. Compared with traditional product recommendation, loan project recommendation cannot be based on some simple sets of straightforward features that are directly available from the projects, and does not have reusable items to recommend, leading to data sparsity and cold start problems. We propose to address these problems by incorporating big data analytics into loan project recommendation to generate business intelligence. In this paper, we present a motivation-based recommendation approach and conduct an experiment to apply the proposed approach in Kiva project recommendations. The experiment results indicate that, compared with prior work, the proposed approach has improved project recommendations in inactive lender groups and unpopular loan groups, which shows the superiority of the proposed approach in addressing the data sparsity and cold start problems in loan project recommendation.

Acknowledgments

The authors thank the editors and anonymous reviewers for helpful feedback. This work was partially supported by the National Natural Science Foundation of China (NSFC No. 71701091, 71701043, 71704078, 71622008, 71301071, 71471083, 71771118, 71390521) and the Chinese Ministry of Education Project of Humanities and Social Science (No. 17YJC870020), the Fundamental Research Funds for the Central Universities (14380002), Jiangsu Social Science Foundation (16TQC002), The China Ten Thousand Talent Program (Youth Talent Support Program), Jiangsu Planned Projects for Postdoctoral Research Funds (1601100C).

References

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

Bachmann, A., Becker, A., Buerckner, D., Hilker, M., Kock, F., Lehmann, M., et al. (2015). Online peer-to-peer lending: A literature review. *Journal of Internet Banking and Commerce, 16*, 1–18.

Belleflamme, P., Omrani, N., & Peitz, M. (2015). The economics of crowdfunding platforms. *Information Economics and Policy, 33*, 11–28.

Binucci, C., De Luca, F., Di Giacomo, E., Liotta, G., & Montecchiani, F. (2017). Designing the content analyzer of a travel recommender system. *Expert Systems with Applications, 87*, 199–208.

Bogan, V. L. (2012). Capital structure and sustainability: An empirical study of microfinance institutions. *Review of Economics and Statistics, 94*, 1045–1058.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5–32.

Cadsby, C. B., Du, N., Song, F., & Yao, L. (2015). Promise keeping, relational closeness, and identifiability: An experimental investigation in China. *Journal of Behavioral and Experimental Economics, 57*, 120–133.

Chen, D., Lai, F., & Lin, Z. (2014). A trust model for online peer-to-peer lending: A lender's perspective. *Information Technology Management, 15*, 239–254.

Chi, H.-H. (2011). Interactive digital advertising vs. virtual brand community: Exploratory study of user motivation and social media marketing responses in Taiwan. *Journal of Interactive Advertising, 12*, 44–61.

Choo, J., Lee, D., Dilkina, B., Zha, H., & Park, H. (2014). To gather together for a better world: Understanding and leveraging communities in micro-lending recommendation. In *Proceedings of the 23rd international conference on world wide web* (pp. 249–260).

Deci, E. L. (1975). Conceptualizations of intrinsic motivation. In *Intrinsic motivation* (pp. 23–63). Plenum.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*, 227–268.

Dorfleitner, G., Priberny, C., Schuster, S., Stoiber, J., Weber, M., de Castro, I., et al. (2016). Description-text related soft information in peer-to-peer lending: Evidence from two leading European platforms. *Journal of Banking & Finance, 64*, 169–187.

Duarte, J., Siegel, S., & Young, L. (2012). Trust and credit: The role of appearance in peer-to-peer lending. *Review of Financial Studies, 25*, 2455–2484.

Feng, Y., Fan, X., & Yoon, Y. (2015). Lenders and borrowers' strategies in online peer-to-peer lending market: An empirical analysis of PPDai.com. *Journal of Electronic Commerce Research, 16*, 242–260.

Greiner, W. E., & Wang, H. (2011). Building consumer-to-consumer trust in e-finance marketplaces: An empirical analysis. *International Journal of Electronic Commerce, 15*, 105–136.

Herzenstein, M., Dholakia, U. M., & Andrews, R. L. (2011). Strategic herding behavior in peer-to-peer loan auctions. *Journal of Interactive Marketing, 25*, 27–36.

Herzenstein, M., Sonenshein, S., & Dholakia, U. M. (2011). Tell me a good story and I may lend you money: The role of narratives in peer-to-peer lending decisions. *Journal of Marketing Research, 48*, S138–S149.

Jenq, C., Pan, J., & Theseira, W. (2015). Beauty, weight, and skin color in charitable giving. *Journal of Economic Behavior & Organization, 119*, 234–253.

Kiva. *What is microfinance?* Retrieved from *Last accessed on 13 August 2017* https://www.kiva.org/microfinance .

Kumar, N., & Benbasat, I. (2006). Research note: The influence of recommendations and consumer reviews on evaluations of websites. *Information Systems Research, 17*, 425–439.

Larrimore, L., Jiang, L., Larrimore, J., Markowitz, D., & Gorski, S. (2011). Peer to peer lending: The relationship between language features, trustworthiness, and persuasion success. *Journal of Applied Communication Research, 39*, 19–37.

Lee, E., & Lee, B. (2012). Herding behavior in online P2P lending: An empirical investigation. *Electronic Commerce Research and Applications, 11*, 495–503.

Lee, E. L., Lou, J.-K., Chen, W.-M., Chen, Y.-C., Lin, S.-D., Chiang, Y.-S., et al. (2014). Fairness-aware loan recommendation for microfinance services. In *Proceedings of the 2014 international conference on social computing* (p. 3).

Lin, M., Prabhala, N. R., & Viswanathan, S. (2013). Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. *Management Science, 59*, 17–35.

Liu, Y., Chen, R., Chen, Y., Mei, Q., & Salib, S. (2012). "I loan because...": Understanding motivations for pro-social lending. In *Proceedings of the fifth international conference on web search and data mining* (pp. 503–512).

Malekipirbazari, M., & Aksakalli, V. (2015). Risk assessment in social lending via random forests. *Expert Systems with Applications, 42*, 4621–4631.

McKinnon, S. L., Dickinson, E., Carr, J. N., & Chávez, K. R. (2013). Kiva.org, person--to-person lending, and the conditions of intercultural contact. *Howard Journal of Communications, 24*, 327–347.

Mittelman, R., & Rojas-Mendez, J. I. (2013). Exploring consumer's needs and motivations in online social lending for development. *Journal of Nonprofit & Public Sector Marketing, 25*, 309–333.

Rakesh, V., Choo, J., & Reddy, C. K. (2015). Project recommendation using heterogeneous traits in crowdfunding. In *Proceedings of ICWSM* (pp. 337–346).

Rakesh, V., Lee, W.-C., & Reddy, C. K. (2016). Probabilistic group recommendation model for crowdfunding domains. In *Proceedings of the ninth ACM international conference on web search and data mining* (pp. 257–266).

Serrano-Cinca, C., Gutiérrez-Nieto, B., & López-Palacios, L. (2015). Determinants of default in P2P lending. *Plos One, 10*.

Tseng, F.-C. (2011). Segmenting online gamers by motivation. *Expert Systems with Applications, 38*, 7693–7697.

Vallerand, R. J. (1997). Toward a hierarchical model of intrinsic and extrinsic motivation. *Advances in Experimental Social Psychology, 29*, 271–360.

Venkatesh, V. (1999). Creation of favorable user perceptions: Exploring the role of intrinsic motivation. *MIS Quarterly*, 239–260.

Wangdaizhijia Retrieved from *Last accessed on 13 August 2017* http://shuju.wdzj.com/industry-list.html .

Xu, Y., Guo, X., Hao, J., Ma, J., Lau, R. Y. K., & Xu, W. (2012). Combining social network and semantic concept analysis for personalized academic researcher recommendation. *Decision Support Systems, 54*, 564–573.

Yan, J., Yu, W., & Zhao, J. L. (2015). How signaling and search costs affect information asymmetry in P2P lending: The economics of big data. *Financial Innovation, 1*, 19.

Zhang, F., Liu, Q., & Zeng, A. (2017). Timeliness in recommender systems. *Expert Systems with Applications, 85*, 270–278.

Zhao, H., Liu, Q., Wang, G., Ge, Y., & Chen, E. (2016). Portfolio selections in P2P lending: A multi-objective perspective. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 2075–2084).

Zhao, H., Wu, L., Liu, Q., Ge, Y., & Chen, E. (2014). Investment recommendation in P2P lending: A portfolio perspective with risk management. In *Proceedings of IEEE international conference on data mining* (pp. 1109–1114).