LCW: A Lightweight Recommendation Framework for Non-profit Crowdfunding Projects

JINXIN LIU, YINGYUAN XIAO*, and WENGUANG ZHEN, Engineering Research Center of Learning-Based Intelligent System, Ministry of Education, Tianjin University of Technology, China and Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology, China

In recent years, crowdfunding platform has gradually emerged as a new economic model, and more and more people choose crowdfunding to obtain the initial funds needed by the project. However, with the increasing number of projects, it is difficult for sponsors to find suitable projects they want to invest in. Therefore, it is very important to build an efficient recommendation system on the crowdfunding platform. Different from previous crowdfunding recommendation systems designed to improve returns, this paper designs a lightweight recommendation framework LCW for non-profit crowdfunding platforms. A large number of experiments on real kiva data sets have proved the effectiveness of this method and it obtains the SOTA results.

CCS Concepts: • **Information systems** → **Data analytics**; • **Computing methodologies** → *Artificial intelligence.*

Additional Key Words and Phrases: crowdfunding, neural networks, recommendation, non-profit

ACM Reference Format:
Jinxin Liu, Yingyuan Xiao, and Wenguang Zhen. 2021. LCW: A Lightweight Recommendation Framework for Non-profit Crowdfunding Projects. In *International Conference on Computer Science and Software Engineering (CSSE '21), Singapore.* ACM, New York, NY, USA, 8 pages. https://doi.org/-

1 INTRODUCTION

In recent years, crowdfunding platforms, as a new internet economic model, has developed rapidly. Crowdfunding platforms have three major advantages. First of all, with the help of a crowdfunding platform, the startup team can obtain initial capital, thus realizing innovative ideas. Second, through the early followers of crowdfunding platforms, the entrepreneurial team can obtain the first batch of users and build the user community foundation. Thirdly, with the help of services and support provided by crowdfunding platforms, the team can obtain professional marketing guidance and more opportunities to cooperate with sales channels. Because of these advantages, more and more project teams have chosen to launch start-up projects on crowdfunding platforms. At present, there are some well-known crowdfunding companies include kiva and indiegogo, among which kiva is dedicated to assisting public welfare projects in developing countries to alleviate local poverty. Since its establishment, kiva platform has helped people in more than 70 countries around the world. On the kiva platform, the startup team initiated the application for funds. If the project application is approved, the investor will lend the fund to the corresponding fund management institutions free of charge, and then the fund management institutions will allocate the fund to the applicants according to the plan. This article focuses on the non-profit crowdfunding project on kiva platform.

*corresponding author

Manuscript submitted to ACM

With the rise of crowdfunding platforms, many researchers turned their attention to the research of crowdfunding platforms. In the past research, Xiao et al. [20] developed a corresponding forecasting system by analyzing the information of borrowers and projects. The key to this system lies in predicting whether crowdfunding projects can raise enough funds. Zhang et al. [22] developed a recommendation system on the for-profit crowdfunding platform, aiming at improving the investor's return. Yan et al. [21] has collected a large amount of text data, which is developed by analyzing the motives of both borrowers and lenders. The purpose of this recommendation process is to match the investment preferences of projects and potential investors. These models are all profit-based crowdfunding platform, and the core of which is to improve the returns of investors.

On kiva, a non-profit crowdfunding platform that this paper focuses on, investors should be called sponsors, because they are not very concerned about the return on investment, and the funds are lent to applicants free of charge, and they are more concerned about whether there are any projects that they want to support. However, with the increasing scale of the platform, more and more projects are applied by the startup team on the platform, and the sponsors can not browse all the projects. The platform needs a recommendation system, which can recommend crowdfunding projects that can better meet the needs of potential sponsors. In such tasks, it is necessary to comprehensively consider the sponsor's preference, the applicant's attributes and historical repayment status, project information and so on. At present, the recommended programs for non-profit crowdfunding projects are mainly based on some traditional financial forecasting programs, but these programs have great limitations in extracting data features. Nowadays, with the improvement of machine computing ability, deep learning technology has been widely developed and applied. Because of its nonlinear combination ability, the neural network model has a good generalization effect, so if the scheme based on deep neural network is adopted, it will help to further improve the prediction accuracy.

Based on the above knowledge, this paper proposes a lightweight model LCW based on neural network to recommend crowdfunding projects to sponsors first. Our main contributions are as follows:

- Aiming at the non-profit crowdfunding platform, this paper proposes a lightweight recommendation model LCW based on neural network.
- The experimental results on the crowdfunding data set kiva show that our method exceeds the current SOTA.

The rest of our paper is organized as follows. In the second section introduces the background of research. The third section describes the proposed method. The fourth section is the experimental results, and fifth section concludes this paper.

2 RELATED WORK

In this section, we will introduce the work related to recommendation system and crowdfunding.

Recommendation system is a kind of decision-making assistant system for users [9]. In the face of massive Internet data, users need to get exactly what they want, instead of all kinds of low-quality messy data. In addition, users may not really know what they want, but just search aimlessly on the internet. Under this requirement, recommendation systems has become an important means to filter data and recommend effective information to users. The classic recommendation systems includes collaborative filtering [18] and matrix decomposition. Subsequently, with the development of deep learning technology, the recommendation system ushered in a new explosion, including recommendation system based on high-dimensional feature combination [5, 24], recommendation system based on graph neural network [1, 6], recommendation system based on probability generation model [13, 19], etc.

Crowdfunding is a new internet economic model. At present, the mainstream crowdfunding platform can be divided into four modes: donation, reward, rights and interests, and borrowing [23]. The research on crowdfunding can be divided into three perspectives: investor perspective, platform perspective and borrower perspective. From the investor's point of view, it is necessary to choose the projects that meet the wishes, and at the same time, it is necessary to make decisions between the benefits and risks [4]. From the perspective of the platform, it is necessary to recommend suitable projects to suitable investors, and constantly expand the daily activities, traffic and turnover of the platform [14]. From the borrower's point of view, it is necessary to find investors who may invest in themselves, so that their projects can be supported with sufficient funds [12]. The research in this paper is mainly based on the platform perspective, matching investors and project parties to increase transaction volume.

3 PROPOSED METHOD

3.1 Problem definition

With the rise of crowdfunding platforms, more and more project parties choose to launch their projects on crowdfunding platforms. The general operation process of the mainstream crowdfunding platform is that borrower can initiate project application on the platform, and the investor can choose project that meet their investment intentions. A project can have multiple investors, and one investor can invest in multiple projects. When the project initiated by the borrower obtains sufficient investment, the platform will allocate the investor's funds to its subordinate fund management institutions, and then the borrower will obtain support funds as planned. Kiva crowdfunding platform operates in a similar way.

In the process of recommending projects to investors by kiva crowdfunding platform, we mainly pay attention to three aspects of data, namely, investor attribute S, user attribute B, and project content and characteristics P. Follow the research of [] [] , the recommendation task of crowdfunding project is regarded as a dichotomous problem, that is, to predict whether the investor will vote for this project or not.

3.2 Model detail

An overview of our crowdfunding recommendation is shown in Fig 1.

Referring to Figure 1, we can see that the input of the overall model is various types of data, and the final output is the prediction of whether the sponsor will invest in the borrower's crowdfunding project. There are two types of input data, the first is category data and the second is numeric data. For category data (including ID type data), the method adopted in this paper is to encode them with one-hot encoding. Because the length of the field obtained by this method is too long, an embedding coding layer is connected to it. The numerical data can be divided into 0, 1 data and ordinary numerical data. 0, 1 data itself remains as it is, and other data are processed according to formula 3.2.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

At this point, we have obtained the basic pre-processing data. Next, the processed vectors are spliced. Then, the model is processed in two directions. In the first part, the spliced vectors are input to the fully connected layer. Then, it connects a Dropout layer [17] to prevent the model from over-fitting. Next to this part will be two fully connected layers. Here, every fully connected layer has L2 regular terms, and the activation function is Selu [11]. The second part is the inner product operation between the splice vector and itself. Then, the results of the two parts are spliced again, and input into the fully connected layer. Finally, the final output is obtained through a Sigmoid function.

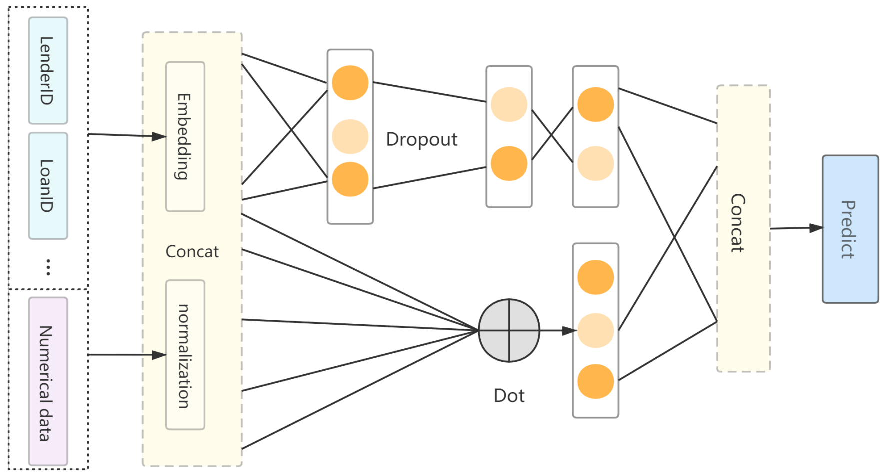

Fig. 1. Framework of LCW.

3.3 Model training

In the training of the model, we use cross entropy as the loss function, and the formula is as follows:

$$Loss = -\left[y log\hat{y} + (1-y)\, log\,(1-\hat{y}) \right]$$

The y here is the true label of a record, and the $\hat{y}$ is the value we predict. On the selection of the optimizer, we chose RMSprop [8], and the initial learning rate was set to 0.001 and the batch-size was set to 128.

In this part, we introduce the basic structure and training methods of the model. The overall structure of the model is lightweight and has high computational efficiency. However, it exceeds many complex models in performance. It will be shown in the next experiment part.

4 EXPERIMENTAL EVALUATION

This section consists of three parts, including experimental setup, results comparison and analysis and hyper-parameters verification.

4.1 Experimental Setup

4.1.1 The kiva dataset. In this paper, we use the data crawled from kiva by [15]. In order to better evaluate the performance of the model, we adopted the same preprocessing scheme as the predecessors, removed the wrong values and abnormal values in the data set, and then screened out the sponsors with at least four investment records. The final data set includes 4,560 different sponsors and 5,487 projects(sponsors and projects are a many-to-many model).The data pairs of sponsor-borrower and other attributes are input into the model every time. The fields contained in the dataset are shown in Table 1.

4.1.2 Basic settings. In order to better evaluate the performance of the current recommendation framework, we randomly divide the historical data set into training sets and test set according to the ratio of 8: 2. The model is trained

Table 1. Fields in the dataset.

	Field Name	Description
Sponsor	LenderID	ID of the person who made the funds.
	Country-code	Country code of Sponsor.
	Loan-count	Number of sponsored projects.
	Invitee-count	Number of responses to a specific request.
Borrower and Project	LoanID	ID of the person who made the funds.
	Funded-amount	The amount of the project already funded.
	Sector	Category to which the project belongs.
	Loc-Country-Code	Location code of borrower.
	Lender-count	Number of sponsors of the current project.
	bonus-credit-eligibility	Historical credit score.

with training set, and then the test set is used as input to predict whether the current sponsor will invest in the current project, and compare the deviation between the predicted result and the actual value. Here, we choose to use cross entropy and AUC, which are two classical evaluation indexes.

In this paper, binary cross entropy is used as the loss function, which is used to judge the prediction results of a binary classification model. Generally speaking, if the predicted value $p(y \mid x)$ is close to 1, the loss function value should be close to 0. On the contrary, if the predicted value p (y) is close to 0 at this time, the value of loss function should be very large, which is very consistent with the properties of the logarithmic function. The lower this indicator, the better.

AUC (Area Under Curve) refers to the area enclosed with the coordinate axis under the ROC curve, and the value of this area will not be greater than 1. In addition, since the ROC curve is generally above the straight line y = x, the AUC is usually between 0.5 and 1. The higher the AUC value, the more accurate the model is, while the closer it is to 0.5, the less discriminating the model is.

4.2 Comparison and analysis of experimental results

4.2.1 Baselines. We consider to compare with the following methods.

- RF [2] Random forest is a classical algorithm, which is used as binary classification here.
- SVM [10] Support vector machine is another classical algorithm, which has been widely used.
- GBDT+LR [7] GBDT+LR is a recommendation algorithm from Facebook advertising team, which plays an important role in large-scale recommendation system. Because of its stability and reliability, it is still the mainstream algorithm in many financial fields.
- FM [16] Factorization machine combines the advantages of SVM and factorization models, and it has unique advantages in feature cross.
- DeepFM [5] DeepFM can be regarded as an algorithm derived from FM, which combines Deep network with FM, uses FM as low-order combination between features, and then uses deep network as high-order combination between features.
- FLEN [3] FLEN divides attributes into domains, and then uses attribute domains to interact with features within and between domains.

- CFILM-P [20] CFILM-P is a SOTA method for recommendation on crowdfunding platform at present. It assists recommendation by calculating the success rate of projects in advance.

Our baseline has good coverage, which includes the classic general classification model in the financial field, the mainstream model in the direction of recommendation system and the SOTA method in the field of crowdfunding recommendation.

4.2.2 Performance. The results of comparison between our method and baseline methods are shown in table 2.

Table 2. Results on cross entropy and AUC.

Methods	Cross entropy	AUC
RF	0.5923	0.6843
SVM	0.5958	0.6788
GBDT+LR	0.5853	0.6901
FM	0.5780	0.7174
DeepFM	0.5495	0.7631
FLEN	0.5443	0.7680
CFILM-P	0.5196	0.8137
LCW	**0.5095**	**0.8255**

By comparing the above experimental results, we have the following findings:

- Some classical effective classification methods, such as RF and SVM, do not perform well in crowdfunding forecasting tasks.
- While other mainstream recommendation algorithms such as GBDT+LR and FM perform better than SVM and RF. Although these algorithms have good ability of feature combination, their results are still not very good due to the dimension of features.
- DeepFM and FLEN are feature interaction methods based on deep learning, which have amazing performance in the field of general recommendation. In the experiment of crowdfunding recommendation, it also significantly surpasses the previous algorithm.
- CFILM-P, as the latest SOTA in the field of crowdfunding recommendation, is significantly better in AUC and cross entropy than the previous methods.
- The proposed LCW method is 1.45% higher than CFILM-P in AUC index and 1.98% better than CFILM-P in cross entropy index. Our method became the new SOTA in the field of crowdfunding recommendation.

4.2.3 Hyper-parameter verification. In order to verify the stability of the proposed model, we will verify the main hyper-parameters in the model in this subsubsection.Here, we mainly consider the influence of two main parameters, embedding size and dropout rate, on the performance of the model.

The figure 2 shows the performance results after adjusting the embedding size of various ID class features. Considering the data size and vector length, we compare the results with sizes of 5, 10, 20 and 100 respectively. Compared with the experimental results, it can be found that the embedding size has a certain impact on the performance of the model. When embedding size is set to very small, the performance of the model on both indicators will drop sharply. Because if embedding size is set too small, too much information will be lost in the encoding and compression process. But, when the embedding size is increased to a certain range, the performance of the model remains relatively stable. At

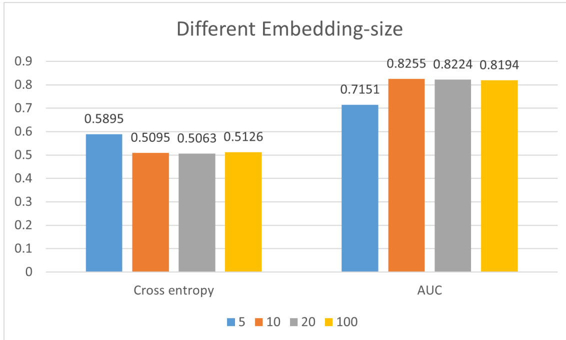

Fig. 2. Hyper-parameter verification on embedding size.

this time, if the size continues to increase, the performance will slightly decrease again, because if the size increases indefinitely, the module will lose the ability to extract features.

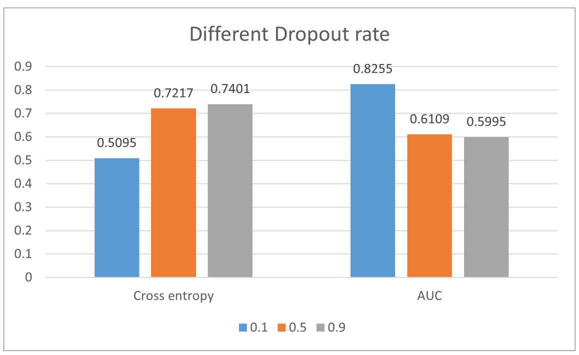

Fig. 3. Hyper-parameter verification on dropout rate.

The figure 3 shows the performance results after adjusting the dropout rate. Here, we compare the results with dropout rate of 0.1, 0.3 and 0.9 respectively. Unlike setting the dropout rate near 0.5, which is common in other deep learning tasks, setting a smaller dropout rate in this paper has better results. The reason may be that the LCW model is more suitable, the risk of over-fitting is less, and setting a higher dropout rate will make the model unable to converge.

4.3 Conclusion

In this paper, we put forward a lightweight LCW model to recommend non-profit crowdfunding projects. The LCW model has simple structure and high computational efficiency. Experiments on kiva data set have proved its effectiveness, and surpass some complicated recommendation models at present, reaching the present SOTA in AUC and cross entropy. For the next step, we consider introducing the concept of disentangled representation, decoupling the user preferences, and further improving the recommendation performance.

ACKNOWLEDGMENTS

This work is supported by Tianjin "Project + Team" Key Training Project under Grant No. XC202022.

REFERENCES

[1] Rianne van den Berg, Thomas N Kipf, and Max Welling. 2017. Graph convolutional matrix completion. *arXiv preprint arXiv:1706.02263* (2017).

[2] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.

[3] Wenqiang Chen, Lizhang Zhan, Yuanlong Ci, Minghua Yang, Chen Lin, and Dugang Liu. 2019. FLEN: leveraging field for scalable CTR prediction. *arXiv preprint arXiv:1911.04690* (2019).

[4] Zhabiz Gharibshah, Xingquan Zhu, Arthur Hainline, and Michael Conway. 2020. Deep learning for user interest and response prediction in online display advertising. *Data Science and Engineering* 5, 1 (2020), 12–26.

[5] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. DeepFM: a factorization-machine based neural network for CTR prediction. *arXiv preprint arXiv:1703.04247* (2017).

[6] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. 2020. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*. 639–648.

[7] Xinran He, Junfeng Pan, Ou Jin, Tianbing Xu, Bo Liu, Tao Xu, Yanxin Shi, Antoine Atallah, Ralf Herbrich, Stuart Bowers, et al. 2014. Practical lessons from predicting clicks on ads at facebook. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*. 1–9.

[8] Geoffrey Hinton, Nitish Srivastava, and Kevin Swersky. 2012. Neural networks for machine learning lecture 6a overview of mini-batch gradient descent. *Cited on* 14, 8 (2012), 2.

[9] Xu Jiao, Yingyuan Xiao, Wenguang Zheng, Hongya Wang, and Ching-Hsien Hsu. 2019. A novel next new point-of-interest recommendation system based on simulated user travel decision-making process. *Future generation computer systems* 100 (2019), 982–993.

[10] Thorsten Joachims. 1998. *Making large-scale SVM learning practical*. Technical Report. Technical report.

[11] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. 2017. Self-normalizing neural networks. In *Proceedings of the 31st international conference on neural information processing systems*. 972–981.

[12] Yung-Ming Li, Jhih-Dong Wu, Chin-Yu Hsieh, and Jyh-Hwa Liou. 2020. A social fundraising mechanism for charity crowdfunding. *Decision Support Systems* 129 (2020), 113170.

[13] Dawen Liang, Rahul G Krishnan, Matthew D Hoffman, and Tony Jebara. 2018. Variational autoencoders for collaborative filtering. In *Proceedings of the 2018 world wide web conference*. 689–698.

[14] Vineeth Rakesh, Jaegul Choo, and Chandan K Reddy. 2015. Project recommendation using heterogeneous traits in crowdfunding. In *Ninth International AAAI Conference on Web and Social Media*.

[15] Vineeth Rakesh, Wang-Chien Lee, and Chandan K Reddy. 2016. Probabilistic group recommendation model for crowdfunding domains. In *Proceedings of the ninth ACM international conference on web search and data mining*. 257–266.

[16] Steffen Rendle. 2010. Factorization machines. In *2010 IEEE International conference on data mining*. IEEE, 995–1000.

[17] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* 15, 1 (2014), 1929–1958.

[18] Xiaoyuan Su and Taghi M Khoshgoftaar. 2009. A survey of collaborative filtering techniques. *Advances in artificial intelligence* 2009 (2009).

[19] Jan Van Balen and Mark Levy. 2019. PQ-VAE: Efficient Recommendation Using Quantized Embeddings.. In *RecSys (Late-Breaking Results)*. 46–50.

[20] Yingyuan Xiao, Chichang Liu, Wenguang Zheng, Hongya Wang, and Ching-Hsien Hsu. 2021. A feature interaction learning approach for crowdfunding project recommendation. *Applied Soft Computing* 112 (2021), 107777.

[21] Jiaqi Yan, Kaixin Wang, Yi Liu, Kaiquan Xu, Lele Kang, Xi Chen, and Hong Zhu. 2018. Mining social lending motivations for loan project recommendations. *Expert Systems with Applications* 111 (2018), 100–106.

[22] Lei Zhang, Xin Zhang, Fan Cheng, Xiaoyan Sun, and Hongke Zhao. 2019. Personalized recommendation for crowdfunding platform: A multi-objective approach. In *2019 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 3316–3324.

[23] Hongke Zhao, Yong Ge, Qi Liu, Guifeng Wang, Enhong Chen, and Hefu Zhang. 2017. P2P lending survey: platforms, recent advances and prospects. *ACM Transactions on Intelligent Systems and Technology (TIST)* 8, 6 (2017), 1–28.

[24] Guorui Zhou, Na Mou, Ying Fan, Qi Pi, Weijie Bian, Chang Zhou, Xiaoqiang Zhu, and Kun Gai. 2019. Deep interest evolution network for click-through rate prediction. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 5941–5948.