Bike-Sharing Apps Using Multi-Source Data on Analysis and Popularity Prediction

N.MD.FIRDOSH
Computer Science Engineering, M.Tech
SVR Engineering College,Nandyal,India

Abstract—In recent years, bike-sharing systems have been widely deployed in many big cities, which provide an economical and healthy lifestyle. With the prevalence of bike-sharing systems, a lot of companies join the bike-sharing market, leading to increasingly fierce competition. To be competitive, bike-sharing companies and app developers need to make strategic decisions and predict the popularity of bike-sharing apps. However, existing works mostly focus on predicting the popularity of a single app, the popularity contest among different apps has not been explored yet. In this paper, we aim to forecast the popularity contest between Mobike and Ofo, two most popular bike-sharing apps in China. We develop, a system to predict the popularity contest among bike-sharing apps leveraging multi-source data. We extract two novel types of features: coarse-grained and fine-grained competitive features, and utilize Random Forest model to forecast the future competitiveness. In addition, we view mobile apps competition as a long-term event and generate the event storyline to enrich our competitive analysis. We collect data about two bike-sharing apps and two food ordering & delivery apps from 11 app stores and Sina Weibo, implement extensive experimental studies, and the results demonstrate the effectiveness and generality of our approach.

Index Terms—Bike-sharing app, mobile app, popularity prediction, competitive analysis, event storyline.

1 Introduction

In recent years, shared transportation has grown tremendously, which provides us an economical and healthy lifestyle. Among the various forms of shared transportation, public bike-sharing systems [1], [2], [3] have been widely deployed in many metropolitan areas such as New York City in the US and Beijing in China. A bike-sharing system provides short-term bike rental service with many bicycle stations distributed in a city [4]. A user can rent a bike at a nearby bike station, and return it at another bike station near his/her destination. The worldwide prevalence of bike-sharing systems has inspired lots of active research, addressing interesting topics such as bike demand prediction [5], [6], [7], [8], bike rebalancing optimization [4], [9] and bike lanes planning [10].

More recently, station-less bicycle-sharing systems are becoming the mainstream in many big cities in China such as Beijing and Shanghai. Mobike[1] and Ofo[2] are two most popular station-less bicycle-sharing systems. Unlike traditional public bike-sharing systems, station-less bike sharing systems aim to solve "the last one mile" issue for users. Using the Mobike/Ofo mobile app, users can search and unlock nearby bikes from Mobike/Ofo. When users arrive at their destinations, they do not have to return the bikes to the designated bike station. Instead, they can park the bicycles at a location more convenient for them. Therefore, it is easier for users to rent and return bikes than traditional bike-sharing systems.

As bike-sharing apps become increasingly popular, a lot of companies join the market, leading to fierce competition. To thrive in this competitive market, it is vital for bike-sharing companies and app developers to understand their competitors, and then make strategic decisions [11] for mobile app development and evolution [12], [13], [14]. Therefore, it is significant and necessary to predict and compare the future popularity of different bike-sharing apps.

With the rapid development of mobile social media, more and more users can contribute valuable data [15], [16], [17], [18], and data from multiple sources can bring multi-dimensional and rich information about bike-sharing apps. When users download and install a mobile app, they may submit user experience to the app store [19], [20], [21]. Specifically, users may upload their requirements (e.g. functional requirements), preferences (e.g. UI preferences) or sentiment (e.g. positive, negative) through reviews, as well as their satisfaction level through ratings. Therefore, the app store data can reflect users' online experience with the app. Online social media is another way to share the user experience of a mobile app. When users actually use the bike, they may share the ride experience on social media. Specifically, users may record the feeling of the ride, the advantages and disadvantages of the bike/system, or the comparison with other bikes/systems. Therefore, the microblogging data can reflect users' offline experiences with

- Y. Ouyang, B. Guo, T. Guo and Z. Yu are with the Northwestern Polytechnical University, Xi'an, Shaanxi, 710129, China.
E-mail: {ouyangyi, tongg}@mail.nwpu.edu.cn, {guob, zhiwenyu}@nwpu.edu.cn.
- X. Lu is with Baidu Inc., Beijing, China.
E-mail: luxinjiang@baidu.com.
- Q. Han is with Colorado School of Mines, Golden, CO, 80401.
E-mail: qhan@mines.edu.

Manuscript received December 09, 2017; revised June 10, 2018.

1. https://en.wikipedia.org/wiki/Mobike
2. https://en.wikipedia.org/wiki/Ofo_(bike_sharing)

ISSN: 2278-4632
Vol-10 Issue-7 No. 4 July 2020

respect to the bikes/system. Both users' online and offline experience will affect the popularity of the apps, thereby affecting their popularity contest outcome. Therefore, app store data and microblogging data are complementary, and can describe a mobile app from different perspectives. In this paper, we study the problem of competitive analysis and popularity prediction of bike-sharing apps using both app store and microblogging data.

To the best of our knowledge, the problem of predict- ing the competitiveness of mobile apps has not been well investigated in the literature. There are several challenging questions to be answered. First, how to forecast the popularity contest of bike-sharing apps? Second, how to extract effective features to characterize the contest of bike-sharing apps from multi-source data? Last, how to generate the event storyline to provide competitive analysis and present the results of the popularity contest of bike-sharing apps?

To answer these questions, we propose CompetitiveBike, a system that predicts the outcomes of the popularity contest among bike-sharing apps leveraging app store data and microblogging data, and then generates the event storyline of the contest. We first obtain app descriptive statistics and sentiment information from app store data, and descriptive statistics and comparative information from microblogging data. Using these data, we extract both coarse-grained and fine-grained competitive features, we then train a regression model to predict the outcomes of popularity contest. We finally generate the event storyline to provide competitive analysis and present the popularity contest. In summary, we make the following contributions.

(1) This work is the first to study the problem of competitive analysis and popularity contest of bike-sharing apps. We use two indicators for the comparison: i) competitive relationship to indicate which app is more popular; and ii) competitive intensity to measure the popularity gap between the two apps/systems.

(2) To predict popularity contest, we extract features from different aspects including inherent descriptive information of apps, users' sentiment, and comparative opinions. With this information, we further extract two novel features: coarse-grained and fine-grained competitive features, and choose Random Forest algorithm for prediction.

(3) To provide competitive analysis, we utilize topic model to analyze the topics in competing apps, and apply the minimum-weight dominating algorithm to select representative microblogs. We also generate event storyline to present and visualize the popularity contest.

(4) To evaluate CompetitiveBike, we collect data about Mobike and Ofo from 11 mobile app stores and Sina Weibo. With the data collected, we conduct extensive experiments from different perspectives. We find that the Random Forest model performs well on *competitive relationship* prediction (the Accuracy is 71.4%) as well as *competitive intensity* prediction (the RMSE is 0.1886). A combination of the coarse-grained and fine-grained competitive features improves performance in popularity contest prediction, and a combination of data from app store and microblogging also improves performance in popularity contest prediction. Besides, we collect data about two mobile food ordering & delivery apps, the results with extensive experiments demonstrate the effectiveness and generality of our method.

2 Event Storyl ine Generation

In order to provide analysis for mobile apps, we view the mobile apps competition as a long-term event, and generate the event storyline and present descriptive information regarding popularity contest to enrich our com- petitive analysis.

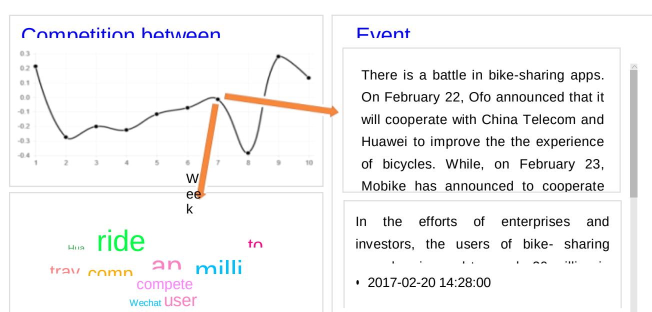
2.1 Event Summarization

In general, the textual content of microblogs can describe the competition between Mobike and Ofo. To explore fine-grained event storyline of the mobile apps competition, we divide the timeline into several time windows with one week as a time window. In each time window, there may be several topics involved with each topic reporting a different aspect of the event. It is hard to distinguish different topics and track the development of the event. In order to clearly understand the event, microblogs referring to different topics of the event should be separated. Therefore, we utilize LDA (Latent Dirichlet Allocation) [38] to extract topics in each time window.

Generally, there are redundant textual contents in each topic. To reduce redundancy, we select representative microblogs to facilitate storyline presentation and easy reading, and use the dominating set algorithm (DS) [39] to select representative microblogs. A subset of vertices in a graph is called a dominating set if every vertex in the graph is contained in the set or has a neighbor in it. The minimum-weight dominating set is the dominating set with minimum weight. In our problem, the microblogs weighted by TF-IDF are connected as a graph based on textual similarity α, and the dominating set represents the set of important microblogs in the graph. Selecting important microblogs is an NP-hard problem and we employ a greedy method [39] to attain a near-optimal solution. Finally, we select the representative microblogs to generate the event storyline.

2.2 Popularity Contest Visualization

The selected representative microblogs can provide competitive analysis about the popularity contest, and we generate the event storyline [40] to present the popularity contest. Figure 3 presents the event storyline of the competition between Mobike and Ofo. The top left figure shows the curve of competition over Mobike and Ofo, which is weekly recorded on the timeline. The word cloud in the bottom left corner presents the hot topics at the particular time instant shown in the top left figure. The right figure presents the event storyline, which provides the content, released timestamp and url for users to explore the eve



ISSN: 2278-4632
Vol-10 Issue-7 No. 4 July 2020

3 Performance Evaluation

3.1 Experimental Setup

The performance of CompetitiveBike is evaluated from the following aspects: 1) *popularity contest prediction*, whether the multi-source data and the extracted features can result in better performance on popularity contest prediction in comparison with baseline methods; 2) *storyline generation*, whether the generated event storyline can present the competition evolution and support competitive analysis.

3.2 Discussion

We next discuss the research findings from this work and potential future directions to improve this work.

Competitive Analysis and Popularity Prediction. This work measures competitiveness among mobile apps based on the number of downloads. However, competitiveness is related to many different aspects. The classical economic theories on competitive analysis [43], [44] can be leveraged to enrich our system framework. Furthermore, we may also consider other competitiveness-related factors that are specific to mobile apps, such as app ranking in app stores.

Multi-Source Data Fusion. We use a combination of app store data and microblog data for competitive analysis and prediction of bike-sharing apps. In fact, the bike sharing sys- tem can be regarded as a cyber-physical-social system, so it is also important to consider more spatial-temporal features in learning [45], and the real-world operation data (e.g., daily usage, profitability). Besides the download records of bike-sharing apps, real-world operation data can also characterize the popularity of the bike-sharing systems from different aspects, and all of these indicators are critical to the

TABLE 9
An example of representative microblogs

Week	Microblogs
36	2017-02-24 09:54:12, There is a battle in bike-sharing apps. On February 22, Ofo announced that it will cooperate with China Telecom and Huawei to improve the the experience of bicycles. While, on February 23, Mobike has announced to cooperate with WeChat.
	2017-02-20 14:28:00, In the efforts of enterprises and investors, the users of bike-sharing apps has jumped to nearly 20 million in 2016, especially of Ofo and Mobike. Currently, Ofo and Mobike are the two leading companies in bike-sharing system.

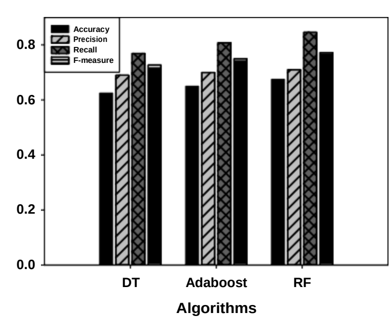


Fig. 9. Comparison of algorithms.

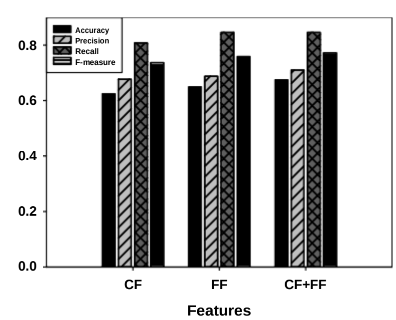


Fig. 10. Comparison of features.

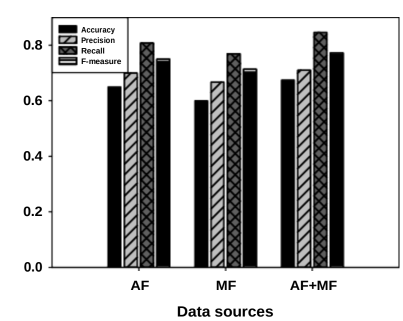


Fig. 11. Comparison of data sources.

TABLE 10
RMSE of Different Algorithms

Last	SVR	RF
0.1653	0.1400	0.1331

TABLE 11
RMSE of Different Features

CF	FF	CF+FF
0.1421	0.1373	0.1331

TABLE 12
RMSE of Different Data Sources

AF	MF	AF+MF
0.1354	0.1429	0.1331

real-world operation data

***Prediction Model**. As the competition among mobile apps is a dynamic and evolutionary process, the popularity contest prediction of mobile apps can be regarded as a se- quence prediction problem.*

Taking the Mobike and Ofo for example, their own popularity can be represented as a se- quence. There may exist complex interactions (or couplings) between the two sequences [46], [47], i.e., competitive cou- plings, which will be affected by both current and historical success of the bike-sharing systems. Though existing bike-sharing apps do not provide public interfaces to retrieve bike trajectory data, we may refer to geo-tagged microblog posts to understand real-world human interaction with shared bikes. In the future work, we will seek collaboration with the bike-sharing companies to obtain more real-world operation data (e.g., daily usage, bike trajectory), and improve the design and evaluation of our approach.

COMPETITIVE ***Feature Extraction***. Different categories of features are defined for app contest prediction and they are proved useful in the experiments. In the future work, we plan to explore new features that are related to competitive analysis and prediction. For example, the information in the app stores (e.g., app ranking, topics and opinions extracted from reviews), the inherent attributes of apps (e.g., categories, permissions, functions, quality), the influence and reputation of the app developers, and the spatial-temporal features (e.g., daily usage, bike trajectory) extracted from observations (i.e., the competitive features mentioned in this paper). To capture the complex interactions between the popularity sequences, in the future work, we will consider the following possible approaches.

The Hidden Markov Model (HMM) [48] is often used for sequence modeling. However, the HMM can only be used for single sequence modeling, it cannot model the complex interactions (i.e., competition) between the above two popularity sequences [46]. To capture the complex interactions, the Coupled Hidden Markov Model (CHMM) [46], [49] can be applied to model the competitive coupling between the two popularity sequences. The CHMM consists of two coupled chains of HMMs representing different popularity sequences, in which the state of any chain of the HMM at time t depends not only on the state of its own chain, but also on the states of other chain of the HMM at time t 1, namely the interaction between two popularity sequences.

4 Conclusion

In this paper, we focus on the problem of competitive analysis and popularity prediction over Mobike and Ofo. We propose CompetitiveBike to predict the popularity contest between Mobike and Ofo leveraging app store data and microblogging data. Specifically, we first extract features from different perspectives including the inherent descriptive information of apps, users' sentiment, and comparative opinions. With this information, we further extract two sets of novel features: coarse-grained and fine-grained competitive features. Finally, we generate the event storyline to provide competitive analysis and present the popularity contest. We collect data about two bike-sharing apps and two food ordering & delivery apps from 11 mobile app stores and Sina Weibo, implement extensive experimental studies, and the results demonstrate the effectiveness and generality of our approach.

Acknowledgments

This work was partially supported by the National Key R&D Program of China (No. 2017YFB1001800), and the Na- tional Natural Science Foundation of China (No. 61332005, 61772428, 61725205).

References

[1] P. DeMaio, "Bike-sharing: History, impacts, models of provision, and future," *Journal of Public Transportation*, vol. 12, no. 4, p. 3, 2009.
[2] S. Shaheen, S. Guzman, and H. Zhang, "Bikesharing in europe, the americas, and asia: past, present, and future," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2143, pp. 159–167, 2010.
[3] J. Pucher, J. Dill, and S. Handy, "Infrastructure, programs, and policies to increase bicycling: an international review," *Preventive medicine*, vol. 50, pp. S106–S125, 2010.
[4] J. Liu, L. Sun, W. Chen, and H. Xiong, "Rebalancing bike sharing systems: A multi-source data smart optimization," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 1005–1014.
[5] L. Chen, D. Zhang, G. Pan, X. Ma, D. Yang, K. Kushlev, W. Zhang, and S. Li, "Bike sharing station placement leveraging heterogeneous urban open data," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015, pp. 571–575.
[6] L. Chen, D. Zhang, L. Wang, D. Yang, X. Ma, S. Li, Z. Wu, G. Pan, T.-M.-T. Nguyen, and J. Jakubowicz, "Dynamic cluster-based over-demand prediction in bike sharing systems," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016, pp. 841–852.
[7] J. Liu, Q. Li, M. Qu, W. Chen, J. Yang, H. Xiong, H. Zhong, and Y. Fu, "Station site optimization in bike sharing systems," in *Data Mining (ICDM), 2015 IEEE International Conference on*. IEEE, 2015, pp. 883–888.
[8] J. Liu, L. Sun, Q. Li, J. Ming, Y. Liu, and H. Xiong, "Functional zone based hierarchical demand prediction for bike system expansion," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2017, pp. 957–966.
[9] A. Singla, M. Santoni, G. Bartók, P. Mukerji, M. Meenen, and A. Krause, "Incentivizing users for balancing bike sharing systems." in *AAAI*, 2015, pp. 723–729.
[10] J. Bao, T. He, S. Ruan, Y. Li, and Y. Zheng, "Planning bike lanes based on sharing-bikes' trajectories," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM.,2017.