CompetitiveBike: Competitive Analysis and Popularity Prediction of Bike-Sharing Apps Leveraging Multi-Source Data

Yi Ouyang[1], Bin Guo[1], Xinjiang Lu[1], Qi Han[2],
Tong Guo[1], and Zhiwen Yu[1]

[1] Northwestern Polytechnical University, Xi'an, Shaanxi, China
guob@nwpu.edu.cn
[2] Colorado School of Mines, Golden, CO, USA

Abstract. In recent years, bike-sharing systems have been widely deployed in many big cities, which provide an economical 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 for mobile apps development. Therefore, it is significant to predict and compare the popularity of different 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 CompetitiveBike, a system to predict the popularity contest among bike-sharing apps. Moreover, we conduct experiments on real-world datasets collected from 11 app stores and Sina Weibo, and the experiments demonstrate the effectiveness of our approach.

Keywords: Bike-sharing app, Mobile app, Competitive analysis, Popularity prediction

1 Introduction

In recent years, shared transportation has grown tremendously, which provides us an economical lifestyle. Among the various forms of shared transportation, public bike-sharing systems [1], [2], [3] have been widely deployed in many metropolitan areas (e.g. 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, such as bike demand prediction [5], [6], [7], bike rebalancing optimization [8], and bike lanes planning [9].

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

---

[1] https://en.wikipedia.org/wiki/Mobike

Ofo² 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. 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 bike-sharing 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 accordingly [10] for mobile app development and evolution [11]. Therefore, it is significant and necessary to predict and compare the future popularity of different bike-sharing apps.

When users download and install a mobile app, they may submit user experience to the app store [12], [13], [14]. 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. 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. 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 app store data and microblogging data.

To the best of our knowledge, the problem of predicting the competitiveness of mobile apps has not been well investigated in the literature. There are several challenging questions to be answered. How to estimate and forecast the popularity contest outcomes of bike-sharing apps? How to extract effective features to characterize the popularity of bike-sharing apps from multi-source data?

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. 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. Finally, we train a regression model to predict the outcomes of popularity contest. We make the following contributions.

(1) This work is the first to study the problem of competitive analysis 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 `https://en.wikipedia.org/wiki/Ofo\_(bike\_sharing)`

(2) To predict popularity contest between apps, we extract features from different perspectives including the descriptive information of apps, users' sentiment, and comparative opinions. Using the basic information, we further extract two novel features: coarse-grained and fine-grained competitive features, and choose Random Forest for prediction.

(3) To evaluate CompetitiveBike, we collect data about Mobike and Ofo from 11 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. The results demonstrate the effectiveness of our approach.

2   Related Work

2.1   App Popularity Prediction

Recently, a significant effort has been spent on predicting popularity of mobile app [15], [17], [16], [18]. Zhu et al. [15] proposed the Popularity-based Hidden Markov Model (PHMM) to model the popularity information of mobile apps. Wang et al. [16] proposed a hierarchical model to forecast the app downloads. Malmi [17] found that there existed connection between app popularity and the past popularity of other apps from the same publisher. Finkelstein et al. [18] found that there is a strong correlation between rating and the downloads.

Our work differs from and potentially outperforms the previous work in several aspects. First, we focus on the problem of competitive analysis and popularity prediction of bike-sharing apps, instead of the prediction of a single app. Second, we predict the popularity contest leveraging multi-source data (i.e., app store data and microblogging data) that are often complementary and can reflect mobile app popularity contest from different perspectives.

2.2   Competitive Analysis

Competitive analysis involves the early identification of potential risks and opportunities to help managers making strategic decisions for an enterprise [10]. Jin et al. [19] selected subjective sentences from reviews which discuss common features of competing products. He et al. [20] analyzed the textual content on the social media of the three largest pizza chains, and the results revealed the business value of comparing social media content. Maksim et al. [21] proposed a generative model for comparative sentences, jointly modeling two levels of comparative relations: the level of sentences and the level of entity pairs. Zhang et al. [22] proposed to scan reviews to update a product comparison network.

These studies conduct competitive analysis simply via semantic analysis of users' opinion. In contrast, our work extracts features from different perspectives

including the descriptive information of apps, user's sentiment, and comparative opinions. Using the basic information, we further extract coarse-grained and fine-grained competitive features, and train a model to predict popularity contest.

3   Data Acquisition and Analysis

3.1   App Store Data

We collected data from 11 mainstream Android app stores[3] in China, including: Wandoujia, Huawei, 360, Meizu, OPPO, VIVO, Yingyongbao, Xiaomi, Baidu, Lenovo and Anzhi market. An overview of app store data is listed in Table 1.

Table 1. Basic Statistics of the App Store Data

Property	Statistics
App stores	11
Time span	04/22/2016 - 03/14/2017
Reviews of Mobike	69,228
Reviews of Ofo	13,928
Total downloads of Mobike	35,591,757
Total downloads of Ofo	30,423,077

We collected data between 04/22/2016 and 03/14/2017. At the beginning, these two apps were still relatively new and they are not as popular now, so there were not a lot of data. To ensure prediction accuracy, the actual time span of the app store data we use is from 06/20/2016 to 03/12/2017, exactly 38 weeks.

Figure 1 shows the weekly downloads of the two apps. We can observe that their downloads are all increasing, and for the recent months, Mobike and Ofo have comparable downloads.
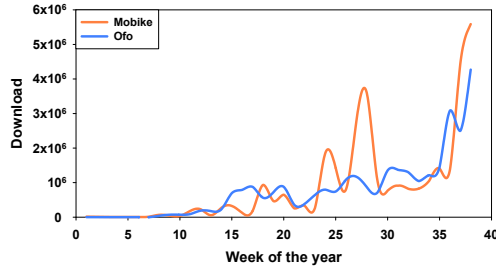

Fig. 1. The weekly downloads of Mobike and Ofo.

---

[3] Data from Google Play is more sparse than these app stores as Mobike and Ofo users are mainly from China, so we did not collect data from Google Play.

3.2 Microblogging Data

We crawled three microblogging datasets from Sina Weibo[4], the most popular microblogging service in China. The first dataset was crawled by using a combination of the two keywords "Mobike" and "Ofo", we refer it as the "Mobike & Ofo". The second one was crawled by using the keyword "Mobike", we refer it as the "Mobike". The third one was crawled by using the keyword "Ofo", we refer it as the "Ofo". An overview of three datasets is listed in Table 2.

Table 2. Basic Statistics of the Microblogging Data

Dataset	Time Span	Microblogs	Users	Reposts	Comments	Likes
Mobike & Ofo	06/21/2016 - 03/14/2017	11,176	8,725	34,801	35,646	31,295
Mobike	04/22/2016 - 03/14/2017	52,718	40,187	151,126	207,926	181,560
Ofo	05/30/2016 - 03/14/2017	43,746	35,752	145,882	181,815	170,644

4 Problem Statement and System Framework

4.1 Problem Statement

The problem can be stated as follows: given the app store data and microblogging data about Mobike and Ofo, we want to predict which app will be more popular in the future.

Definition 1 *Popularity Contest*. Inspired by [23], the popularity of Mobike (or Ofo) can be measured by the downloads, and the popularity contest (PC) between Mobike and Ofo can be defined by the difference in their downloads D_m and D_o:

$$PC = \frac{D_m - D_o}{D_m + D_o} \tag{1}$$

Definition 2 *Competitive Relationship*. The competitive relationship (CR) between Mobike and Ofo can be one of the two possbilities: 1) Mobike is more popular than Ofo, or 2) Ofo is more popular than Mobike. According to Formula (1), when $PC > 0$, Mobike is more popular; otherwise, Ofo is more popular.

Definition 3 *Competitive Intensity*. The competitive intensity (CI) between Mobike and Ofo is the absolute value of PC. The smaller the value, the higher the competitive intensity is.

Formally, we extract feature set X from app store data and microblogging data, then we want to predict the popularity contest Y. Let $X = \{x_1, ...x_N\}$ and $Y = \{y_1, ...y_N\}$, given $X^{(1:t+1)}(= \{X^{(1)}, ..., X^{(t+1)}\})$ and $Y^{(1:t)}(= \{Y^{(1)}, ..., Y^{(t)}\})$, our objective is to predict $Y^{(t+1)}$.

---

[4] https://weibo.com/

4.2   System Framework

The overview of the framework is illustrated in Figure 2, which mainly consists of three layers: data preparation, feature extraction, and competitive analysis and popularity prediction.
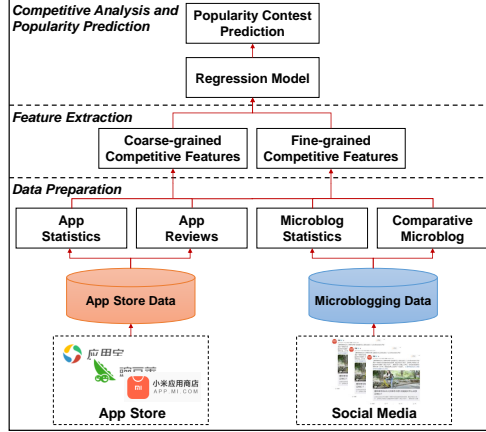

Fig. 2. System framework.

Data Preparation. We obtain app statistics and reviewers' sentiment from app store data, and microblogging statistics and comparative information from microblogging data.

Feature Extraction. To effectively extract and quantify the factors impacting mobile app popularity contest, we extract features from different perspectives including the inherent descriptive information of apps, users' sentiment, and comparative opinions. With this information, we further extract two novel sets of features: coarse-grained and fine-grained competitive features.

Competitive Analysis and Popularity Prediction. With these two extracted feature sets, we train a model to predict the popularity contest between Mobike and Ofo.

5   Popularity Contest Prediction

In this section, we first analyze the factors impacting the popularity contest between Mobike and Ofo, then extract coarse-grained and fine-grained competitive features from these factors to characterize popularity contest. Finally, we train a model to predict popularity contest.

5.1   Coarse-grained Competitive Features

Features from App Store. When users download and install a mobile app, they may submit reviews and ratings to the app store. For example, a user

wrote: "*The Mobike app cannot launch today, it was still okay yesterday, what's the matter? It's terrible!*" According to the review, we believe that app store data (e.g. reviews, ratings) can reflect users' online experience with the app. Typically, users may upload their requirements (e.g. functional requirements), preferences (e.g. UI preferences), or sentiment (e.g. positive, negative) through reviews, and they may also rate the app based on their overall satisfaction. Therefore, we extract features from reviews and ratings to characterize popularity contest.

App Statistics. Generally, the numerical statistics of reviews and ratings in each time window can reflect the popularity of the app. In other words, a bigger number of reviews and a higher rating score may indicate that the app is more popular. We use the difference between app's review number DN (and rating scores DS) to characterize popularity contest. A small value of DN (and DS) indicates that they have similar number of reviews (and rating score), thus their competition is more intense.

Sentiment Similarity. Besides numerical statistics, app reviews can express users' sentiment. We use a Chinese sentiment analyzer called SnowNLP[5] to analyze the sentiment of reviews. We calculate the sentiment value s_i of each review at time instant t_i, then we obtain the sentiment distribution vector $\mathbf{v_i} = (p_1, p_2, p_3)$ at time t_i, where p_1, p_2, p_3 is corresponds to negative, neutral and positive sentiment proportion respectively.

The extracted sentiment sequences are only for a single app, when we consider the competition between two apps, we compute sentiment similarity to capture the difference of users' sentiment about these apps, and the similarity can be measured by calculating the cosine similarity [24]. The higher similarity means that users' opinions about them are more similar, and the competition between them is more intense.

Features from Microblogging. When users ride the bike of different apps, they may share their riding experience on social media. An example of a microblog is like this: "*This is my first ride of Mobike, it is so cool!*" We believe online social media is another way to express users' riding experience. Therefore, we extract features from microblogging data to help understand the popularity contest of different apps.

Microblogging Statistics. In the "Mobike & Ofo" dataset, the number of microblogs, users, reposts, comments, and likes can reflect the attention about Mobike and Ofo on microblogging, the bigger value indicates more intense competition between Mobike and Ofo.

In the "Mobike" dataset, more microblogs that contain the keyword "Ofo" imply that Ofo is more frequently mentioned in the "Mobike" dataset. We use the ratio (R_{om}) of "Ofo" and "Mobike" to characterize the competition. Formally, $R_{om} = \frac{MN_o}{MN_m}$, where MN_o and MN_m represent the number of microblog that contains "Ofo" and "Mobike", respectively. Similarly, in the "Ofo" dataset, we use the ratio (R_{mo}) of "Mobike" and "Ofo" to characterize the competition. The higher ratios, the more intense competition.

---

[5] `https://github.com/isnowfy/snownlp`

Comparative Analysis. In addition to the numerical statistics, the textual information in microblog content is also valuable. The "Mobike & Ofo" dataset often contains the comparison between Mobike and Ofo. Let us consider a microblog: "*Mobike is too heavy, and it is uncomfortable to ride. It is also slightly expensive. Of course, there are some aspects where Mobike is better than Ofo, such as: Mobike is more solid than Ofo, and its bell is also better.*" According to this post, we observe that (1) there exists comparison between Mobike and Ofo; (2) a single microblog may compare the apps many times on different aspects (e.g. price, quality); (3) each comparison can discuss the advantages and disadvantages of the bike. Therefore, we need to address three issues in comparative analysis: (1) how to identify comparison between Mobike and Ofo; (2) how to calculate the comparison count; (3) how to determine the *comparison direction*, which means whether Mobike is better than Ofo, or Ofo is better than Mibike. We next describe our methods to address these issues.

First, the occurrences of comparative words such as "better" often indicate comparison and these comparative words are usually adjective or adverb. Therefore, to identify the comparison, we try to determine whether there exist comparative words in microblogs. Specifically, we use a Chinese lexical analyzer called Jieba[6] to annotate part of speech, and extract adjectives and adverbs to build a dictionary. We then determine whether there exist comparative words by querying the dictionary and filtering out microblogs without comparative words. After this, all the remaining microblogs contain comparison between Mobike and Ofo.

Next, when calculating the comparison count, we do not need to differentiate which aspects are in comparison. We can count the number of comparative word to determine the comparison count.

Last, the sentiment of the comparative words can be used to infer comparison direction. In the example above, "Mobike is more solid than Ofo" implies that Mobike is better than Ofo. We divide the dictionary into two sub-dictionaries: positive and negative. With a positive comparative word, 1 is added to its own score; with a negative comparative word, 1 is added to the score of the competitor. This way, we can obtain the comparison direction scores for Mobike and Ofo. We use the scores to characterize popularity contest.

5.2   Fine-grained Competitive Features

Each coarse-grained competitive feature is a time series with time window of one week. In each time window, we extract the temporal dynamics of the coarse-grained competitive features as the fine-grained competitive features to characterize the trend of the sequence [25].

Overall Descriptive Statistics describe the basic properties of the coarse-grained competitive features from multiple aspects. We extract the mean, standard deviation, median, minimum and maximum as features.

---

[6] https://github.com/fxsjy/jieba

Hopping Counts can effectively describe the "pulse" of sequence and is calculated as the number of elements whose values are greater than their next element. This feature is used to characterize the fluctuation of the sequences.

Lengths of Longest Monotonous Subsequences describe the size of gradient descent or ascent patterns in a sequence. We examine the longest monotone (including increasing and decreasing) subsequences, and use the lengths of these two subsequences to describe the tendency of the sequence.

5.3   Popularity Contest Prediction

With these two extracted feature sets, we want to predict the popularity contest in the future, we use regression-based methods. Since the extracted features are sequences, and the time window is one week, we treat successive several weeks as the training set, then compare the state-of-the-art regression models. Section 6 has the details on the models we compared and the one we eventually use.

6   Performance Evaluation

6.1   Experimental Setup

Comparison Settings. To demonstrate the effectiveness of different types of features, we divide the extracted features into two categories: (1) *coarse-grained competitive features* (CF); (2) *fine-grained competitive features* (FF).

To demonstrate the effectiveness of multi-source data, we divide the extracted features into another two categories according to the data source: (1) *features from app store data* (AF) and (2) *features from microblogging data* (MF).

Regarding algorithm comparison, in the phase of *competitive relationship* prediction, we evaluate three state-of-the-art classification algorithms: Decision Tree (DT), Adaboost and Random Forest (RF). In the phase of *competitive intensity* prediction, we evaluate two state-of-the-art regression algorithms: Support Vector Regression (SVR) and Random Forest (RF).

To conduct popularity contest prediction, we use the following setup: we use ten successive weeks as the training set and the next one week as the test set.

Baseline Algorithms. For popularity contest prediction, we use the following methods as the baselines:

- *Last_predcition*: it predicts the popularity contest using the last one week, i.e. $Y^{(t+1)} = Y^t$. We refer it as "Last".
- *CF*: it predicts the popularity contest using the coarse-grained competitive features alone.
- *FF*: it predicts the popularity contest using the fine-grained competitive features alone.
- *AF*: it predicts the popularity contest using the features from app store alone.
- *MF*: it predicts the popularity contest using the features from microblogging platform alone.

Evaluation Metrics. For features evaluation, we measure the effectiveness of the extracted features using the following metric:

– In the phase of *competitive relationship* prediction, we use *Accuracy, Precision, Recall, F-measure* as the evaluation metrics. Higher values of these metrics means the better performance in *competitive relationship* prediction.
– In the phase of *competitive intensity* prediction, we use *RMSE* as the evaluation metric. A smaller RMSE means the better performance in *competitive intensity* prediction.

6.2   Experimental Results

Comparison of Different Algorithms. We want to compare the effectiveness of different algorithms in popularity contest: *competitive relationship* and *competitive intensity.*

Regarding the *competitive relationship* prediction, Figure 3 shows the Accuracy, Precision, Recall and F-measure of DT, Adaboost and RF. We observe that RF outperforms the other algorithms, with the Accuracy of 71.4%, and the state-of-the-art classification algorithms outperforms the baselines.

Regarding the *competitive intensity* prediction, Table 3 shows the RMSE of Last, SVR, and RF. We observe that RF again outperforms other algorithms, and the RMSE of the baseline is much larger than RF regression algorithm.

In summary, the state-of-the-art machine learning algorithms can train a better learning model by using the proposed features. RF performs well on *competitive relationship* prediction as well as *competitive intensity* prediction. Therefore, we choose RF as the default predictor for predicting popularity contest.
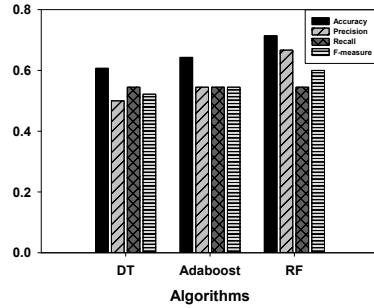

Table 3. RMSE of Algorithms

Last	SVR	RF
0.2280	0.2185	0.1886

Fig. 3. Comparison of Algorithms.

Comparison of Different Features. We try to determine whether the combination of the coarse-grained and fine-grained competitive features can improve the performance of prediction. Therefore, we compare the CF, FF, and CF+FF, respectively.

Figure 4 shows the Accuracy, Precision, Recall and F-measure of CF, FF and CF+FF. We observe that FF outperforms CF, with the Accuracy of 67.9%, while

CF is 60.7%. This is because FF is generated based on CF, and it can reflect the fine-grained tendency of CF. Furthermore, the combination of the coarse-grained and fine-grained competitive features (CF+FF) improves the performance in *competitive relationship* prediction, compared with CF and FF alone.

Table 4 shows the RMSE of CF, FF and CF+FF. We can observe that FF outperforms CF, and can reflect the temporal dynamics of the CF. Furthermore, the combination of the coarse-grained and fine-grained competitive features (CF+FF) improves the performance in *competitive intensity* prediction, compared with CF and FF alone.

In summary, FF outperforms CF in both *competitive relationship* and *competitive intensity* prediction, and the combination of the coarse-grained and fine-grained competitive features (CF+FF) can further improve the performance in competition prediction.
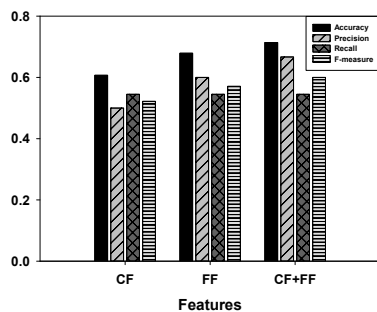

Table 4. RMSE of Features

CF	FF	CF+FF
0.2059	0.1980	0.1886

Fig. 4. Comparison of Features.

Comparison of Different Data Sources. We aim to determine whether the combination of app store data and microblogging data can improve the performance of prediction. Therefore, we compare the AF, MF, and AF+MF, respectively.

Figure 5 shows the Accuracy, Precision, Recall and F-measure of AF, MF and AF+MF. We can observe that AF outperforms MF, with the Accuracy of 64.3%, while MF is 60.7%. This is because that AF constitutes reviews and scores which can reflect users' online experience with the app. Users may report their sentiment or requirement through reviews, and their satisfaction degree through rating scores. It will directly affect the popularity of the app, therefore will affect the popularity contest. In contrast, MF reflects the popularity contest indirectly. Furthermore, the combination of features from app store and microblogging (AF+MF) improves the performance in *competitive relationship* prediction, compared with AF and MF alone.

Table 5 shows the RMSE of AF, MF and AF+MF. We can observe that AF outperforms MF, because AF will directly affect the popularity of the mobile app, while MF reflects the competition indirectly. Furthermore, the combination of features from app store and microblogging (AF+MF) improves the performance in *competitive intensity* prediction, compared with AF and MF alone.

In summary, AF outperforms MF in both *competitive relationship* and *competitive intensity* prediction, and the combination of features from app store and microblogging (AF+MF) further improve the performance in competition prediction.
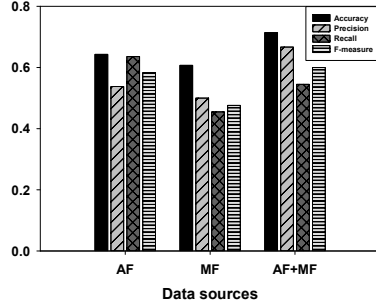

Table 5. RMSE of Data Sources

AF	MF	AF+MF
0.1965	0.2062	0.1886

Fig. 5. Comparison of Data Sources.

7   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 choose the Random Forest algorithm to predict the popularity contest. Moreover, we collect data about two bike-sharing apps from 11 online mobile app stores and Sina Weibo, implement extensive experimental studies, and the results demonstrate the effectiveness of our approach.

In the future work, we will enrich our problem statement and system framework by learning from the classical economic theories on competitive analysis [26], [27]. In order to provide competitive analysis for mobile apps, we will view the mobile apps competition as a long-term event, and generate the event storyline [28] and present descriptive information regarding popularity contest to enrich the competitive analysis. Besides, we will improve the prediction model by analyzing the couplings [29], [30] among features and determining their mutual influence. Moever, we will collect more categories of apps to enrich our datasets, and extend the generality of our approach to other apps.

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, 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.

8. 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.

9. 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, pp. 1377–1386.

10. K. Xu, S. S. Liao, J. Li, and Y. Song, "Mining comparative opinions from customer reviews for competitive intelligence," *Decision support systems*, vol. 50, no. 4, pp. 743–754, 2011.

11. A. Di Sorbo, S. Panichella, C. V. Alexandru, J. Shimagaki, C. A. Visaggio, G. Canfora, and H. C. Gall, "What would users change in my app? summarizing app reviews for recommending software changes," in *Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering*. ACM, 2016, pp. 499–510.

12. W. Martin, F. Sarro, Y. Jia, Y. Zhang, and M. Harman, "A survey of app store analysis for software engineering," *IEEE transactions on software engineering*, vol. 43, no. 9, pp. 817–847, 2017.

13. B. Fu, J. Lin, L. Li, C. Faloutsos, J. Hong, and N. Sadeh, "Why people hate your app: Making sense of user feedback in a mobile app store," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2013, pp. 1276–1284.

14. X. Gu and S. Kim, "" what parts of your apps are loved by users?"(t)," in *Automated Software Engineering (ASE), 2015 30th IEEE/ACM International Conference on*. IEEE, 2015, pp. 760–770.

15. H. Zhu, C. Liu, Y. Ge, H. Xiong, and E. Chen, "Popularity modeling for mobile apps: A sequential approach," *IEEE transactions on cybernetics*, vol. 45, no. 7, pp. 1303–1314, 2015.
16. Y. Wang, N. J. Yuan, Y. Sun, C. Qin, and X. Xie, "App download forecasting: An evolutionary hierarchical competition approach," in *Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2017, pp. 2978–2984.
17. E. Malmi, "Quality matters: Usage-based app popularity prediction," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*.   ACM, 2014, pp. 391–396.
18. A. Finkelstein, M. Harman, Y. Jia, F. Sarro, and Y. Zhang, "Mining app stores: Extracting technical, business and customer rating information for analysis and prediction," *RN*, vol. 13, p. 21, 2013.
19. J. Jin, P. Ji, and R. Gu, "Identifying comparative customer requirements from product online reviews for competitor analysis," *Engineering Applications of Artificial Intelligence*, vol. 49, pp. 61–73, 2016.
20. W. He, S. Zha, and L. Li, "Social media competitive analysis and text mining: A case study in the pizza industry," *International Journal of Information Management*, vol. 33, no. 3, pp. 464–472, 2013.
21. M. Tkachenko and H. Lauw, "Comparative relation generative model," *IEEE Transactions on Knowledge and Data Engineering*, 2016.
22. Z. Zhang, C. Guo, and P. Goes, "Product comparison networks for competitive analysis of online word-of-mouth," *ACM Transactions on Management Information Systems (TMIS)*, vol. 3, no. 4, p. 20, 2013.
23. W. S. DeSarbo, R. Grewal, and J. Wind, "Who competes with whom? a demand-based perspective for identifying and representing asymmetric competition," *Strategic Management Journal*, vol. 27, no. 2, pp. 101–129, 2006.
24. Y. Ouyang, B. Guo, J. Zhang, Z. Yu, and X. Zhou, "Sentistory: multi-grained sentiment analysis and event summarization with crowdsourced social media data," *Personal and Ubiquitous Computing*, pp. 1–15, 2016.
25. X. Lu, Z. Yu, L. Sun, C. Liu, H. Xiong, and C. Guan, "Characterizing the life cycle of point of interests using human mobility patterns," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*.   ACM, 2016, pp. 1052–1063.
26. M. Bergen and M. A. Peteraf, "Competitor identification and competitor analysis: a broad-based managerial approach," *Managerial and decision economics*, vol. 23, no. 4-5, pp. 157–169, 2002.
27. A. Borodin and R. El-Yaniv, *Online computation and competitive analysis*.   cambridge university press, 2005.
28. B. Guo, Y. Ouyang, C. Zhang, J. Zhang, Z. Yu, D. Wu, and Y. Wang, "Crowdstory: Fine-grained event storyline generation by fusion of multi-modal crowdsourced data," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, p. 55, 2017.
29. L. Cao, Y. Ou, and S. Y. Philip, "Coupled behavior analysis with applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 8, pp. 1378–1392, 2012.
30. L. Cao, T. Joachims, C. Wang, E. Gaussier, J. Li, Y. Ou, D. Luo, R. Zafarani, H. Liu, G. Xu *et al.*, "Behavior informatics: A new perspective," *IEEE Intelligent Systems*, vol. 29, no. 4, pp. 62–80, 2014.