WSDM Cup 2018: Music Recommendation and Churn Prediction

Conference Paper · February 2018

DOI: 10.1145/3159652.3160605

Yian Chen
KKBOX Inc.,

WSDM Cup 2018: Music Recommendation and Churn Prediction

Yian Chen
KKBOX Inc.,
Taipei, Taiwan
annchen@kkbox.com

Xing Xie
Microsoft Research Asia
Beijing, China
xingx@microsoft.com

Shou-De Lin
National Taiwan University
Taipei, Taiwan
sdlin@csie.ntu.edu.tw

Arden Chiu
KKBOX Inc.,
Taipei, Taiwan
ardenchiu@kkbox.com

ABSTRACT

Recommendation systems facilitate users retrieving contents they might like but not aware of yet. Furthermore, an effective recommendation system can potentially increase users' retention and conversion rate. One critical challenge for building a recommender system lies in the existence of *cold start* cases when we have sparse records for certain users or items: without enough rating data about a new song or a new user, it is necessary to rely on auxiliary information to perform effective recommendation. In the first task of WSDM Cup 2018, we challenge the participants to solve the abovementioned challenges in building a music recommendation system. The 2nd task of the Cup focuses on churn prediction. For a subscription business, accurately predicting churn is critical to its long-term success as even a slight variation in churn can significantly affect the profits. In this task, participants are asked to build an algorithm that predicts whether a user will churn after their subscription expires. The competition data and award are provided by KKBOX, a leading music streaming service company from Taiwan.

CCS CONCEPTS

•Information systems → Personalization; •Computing methodologies → Ranking; Supervised learning by classification; Supervised learning by regression;

KEYWORDS

Recommendation, Personalization, Churn Prediction

ACM Reference format:
Yian Chen, Xing Xie, Shou-De Lin, and Arden Chiu. 2018. WSDM Cup 2018: Music Recommendation and Churn Prediction. In *Proceedings of 11th ACM International Conf. on Web Search and Data Mining, Marina Del Rey, CA, USA, February 5–9, 2018 (WSDM 2018),* 2 pages.
DOI: 10.1145/3159652.3160605

DOI: 10.1145/3159652.3160605

1 TASK ON MUSIC RECOMMENDATION SYSTEM

Not many years ago, it was inconceivable that the same person would listen to the Beatles, Vivaldi, and Lady Gaga on their morning commute. However, the glory days of Radio DJs have passed, and musical gatekeepers have been replaced with personalizing algorithms and unlimited streaming services.

While the public's now listening to all kinds of music, recommender systems still struggle in key areas such as the cold start problem[9][4] and context aware recommendation [7][10]. This challenge focuses on address such concerns for a music recommendation task.

In this task, participants are asked to predict the chances of a user listening to a song repetitively after the first observable listening event within a time window was triggered. If there are recurring listening event(s) triggered within a month after the userfis very first observable listening event, its target is marked 1, and 0 otherwise in the dataset.

1.1 Data Description

KKBOX provides training data consisting information of the first observable listening event for each unique user-song pair within specific time duration. Moreover, we provide the metadata of our users and songs. What worth to mention is, participants are encouraged to use public external data in this data challenge.

The statistics of data are listed in Table 1 and file names represent the files we released on the competition platform, Kaggle.com[2][1]. The training data contains the contextual information, such as System tab, Screen name, and Source type, of listening events. What is worth to note is, within the training data, there are 17.5% of users and 83% of items have less then 10 observations, 1 illustrate cumulative probability of number of observations of each user and item, respectively. This phenomenon indicates we have cold start problem both for users and items and item cold start problem seems more severe then user cold start problem. The schema of training data is available on the competition platform [2].

Table 1: Overall metrics of the dataset of recommendation challenge

description	training data	metadata of songs	metadata of users
number of rows	7,377,418	2,296,320	34,403

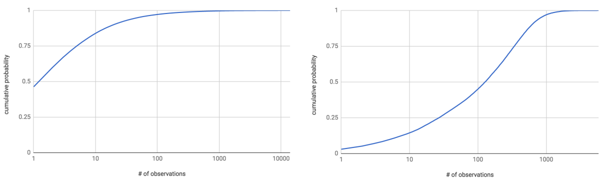

Figure 1: cumulative probablity of number of observations

1.2 Evaluation Metrics

We formulate the recommendation challenge as a ranking task. Thus, we employ Area under ROC curve as our evaluation metrics. It is a de-facto standard for highly imbalanced data.

2 TASK ON CHURN PREDICTION

For a subscription business, accurately predicting churn is critical to its long-term success. There have been several challenges associating with churn prediction in the past [12][6][5].

KKBOX, as the leading music streaming service in Asia, is supported by advertising and paid subscriptions. The success of such delicate business model relies heavily on accurately predicting churn of their paid users.

2.1 Data Description

In this task, participants will be predicting whether a user will churn after their subscription expires. Specifically, we want to see if a user make a new service subscription transaction within 30 days after their current membership expiration date.

As a music streaming service provider, KKBOX has members subscribe to their service. When the subscription is about to expire, the user can choose to renew, or cancel the service. They also are given the option to auto-renew, despite their membership can still be cancelled at any time.

The definition of churn/renewal can be tricky due to KKBOX's subscription model. Since the majority of the subscription length is 30 days, a lot of users re-subscribe every month. The key fields to determine churn/renewal are transaction date, membership expiration date, and is_cancel. Note that the is_cancel field indicates whether a user actively cancels a subscription. Note that a cancellation does not imply the user has churned. A user may cancel service subscription due to change of service plans or other reasons. Here we define "churn" as: no new valid service subscription within 30 days after the current membership expires. Concrete examples are listed on competition site [1]

Table 2 shows the statistics of the data released. The training and testing data are selected from users whose membership expire within a certain month. The training data consists of users whose subscription expire within the month of February 2017, and the testing data is with users whose subscription expire within the month of March 2017. This means we are looking at user churn or renewal roughly in the month of March 2017 for training set, and the user churn or renewal roughly in the month of April 2017. In this dataset, KKBOX has included a variety types of users behaviors. For example, a user could actively cancel the subscription, but renew within 30 days.

The schema of transaction log, training data, user behavioral log and user meta table is available on the competition platform [1].

Table 2: Overall metrics of the dataset of churn prediction

description	training data	user logs	transactions	members
number of rows	970961	18396363	1431010	6769474

2.2 Evaluation Metrics

The evaluation metric for this competition is Log Loss as follows

$$-\frac{1}{N}\sum_{i=1}^{N}[y_i \log p_i + (1-y_i)\log(1-p_i)]$$

where y_i is the actual label, i.e., is_churn, of sample i and p_i is the predicted probability of the sample's label equals to 1.

3 METRICS OF COMPETITION

Based on the data presented on competition platform, Kaggle [2][1], there are 1096 teams ,formed by 1284 individual competitors joined recommendation challenge, and 526 teams, formed by 780 individual competitors joined churn prediction on December 14, 2017.

REFERENCES

[1] WSDM - KKBox's Churn Prediction Challenge. https://www.kaggle.com/c/kkbox-churn-prediction-challenge. (????).
[2] WSDM - KKBox's Music Recommendation Challenge. https://www.kaggle.com/c/kkbox-music-recommendation-challenge. (????).
[3] Gediminas Adomavicius and Alexander Tuzhilin. 2015. *Context-Aware Recommender Systems.* Springer US, Boston, MA, 191–226. https://doi.org/10.1007/978-1-4899-7637-6_6
[4] Jesfls Bobadilla, Fernando Ortega, Antonio Hernando, and Jesfls Bernal. 2012. A collaborative filtering approach to mitigate the new user cold start problem. *Knowledge-Based Systems* 26, Supplement C (2012), 225 – 238. https://doi.org/10.1016/j.knosys.2011.07.021
[5] J. Burez and D. Van den Poel. 2009. Handling class imbalance in customer churn prediction. *Expert Systems with Applications* 36, 3, Part 1 (2009), 4626 – 4636. https://doi.org/10.1016/j.eswa.2008.05.027
[6] Kristof Coussement and Dirk Van den Poel. 2008. Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications* 34, 1 (2008), 313 – 327. https://doi.org/10.1016/j.eswa.2006.09.038
[7] Alexandros Karatzoglou, Xavier Amatriain, Linas Baltrunas, and Nuria Oliver. 2010. Multiverse Recommendation: N-dimensional Tensor Factorization for Context-aware Collaborative Filtering. In *Proceedings of the Fourth ACM Conference on Recommender Systems (RecSys '10).* ACM, New York, NY, USA, 79–86. https://doi.org/10.1145/1864708.1864727
[8] Xuan Nhat Lam, Thuc Vu, Trong Duc Le, and Anh Duc Duong. 2008. Addressing Cold-start Problem in Recommendation Systems. (2008), 208–211. https://doi.org/10.1145/1352793.1352837
[9] Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. 2014. Facing the cold start problem in recommender systems. *Expert Systems with Applications* 41, 4, Part 2 (2014), 2065 – 2073. https://doi.org/10.1016/j.eswa.2013.09.005
[10] Steffen Rendle, Zeno Gantner, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2011. Fast Context-aware Recommendations with Factorization Machines. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '11).* ACM, New York, NY, USA, 635–644. https://doi.org/10.1145/2009916.2010002
[11] Andrew I. Schein, Alexandrin Popescul, Lyle H. Ungar, and David M. Pennock. 2002. Methods and Metrics for Cold-start Recommendations. (2002), 253–260. https://doi.org/10.1145/564376.564421
[12] Chih-Ping Wei and I-Tang Chiu. 2002. Turning telecommunications call details to churn prediction: a data mining approach. *Expert Systems with Applications* 23, 2 (2002), 103 – 112. https://doi.org/10.1016/S0957-4174(02)00030-1