

Recommender Systems Challenge 2014

Alan Said¹, Simon Doods², Babak Loni¹, Domonkos Tikk³

¹TU-Delft, The Netherlands

²Ghent University, Belgium

³Gravity R&D, Hungary

alansaid@acm.org, simon.doods@intec.ugent.be,
b.loni@tudelft.nl, domonkos.tikk@gravityrd.com

ABSTRACT

The 2014 ACM Recommender Systems Challenge invited researchers and practitioners to work towards a common goal, this goal being the prediction of users engagement in movie ratings expressed on Twitter. More than 200 participants sought to join the challenge and work on the new dataset released in its scope. The participants were asked to develop new algorithms to predict user engagement and evaluate them in a common setting, ensuring that the comparison was objective and unbiased in the setting of the challenge.

Categories and Subject Descriptors

D.2.8 [Software Engineering]: Metrics - complexity measures, performance measures; H.3.3 [Information Storage and Retrieval]: Information search and retrieval - information filtering, relevance feedback; H.3.4 [Information Technology and Systems Applications]: Decision support; H.3.5 [Online Information Services]: Data Sharing; H.5.1 [Multimedia Information Systems]: Evaluation/methodology

General Terms

Algorithms; Design; Experimentation; Human Factors; Measurement

Keywords

Recommender Systems; Dataset; Challenge; Competition; Context-aware; Benchmarking

1. INTRODUCTION

The recommender systems community has a long standing tradition of organized benchmarking events, the most prominent one being the Netflix Prize¹. Benchmarking events serve the purpose of creating a focused attempt at solving

¹<http://www.netflixprize.com>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

RecSys'14, October 6–10, 2014, Foster City, Silicon Valley, CA, USA.

ACM 978-1-4503-2668-1/14/10.

<http://dx.doi.org/10.1145/2645710.2645779>.

a predefined problem. By providing a dataset, a certain setting and the means of evaluation, participants only have to focus on algorithmic performance as everything else is already provided.

The 2014 edition of the RecSys Challenge was the fifth annual ACM RecSys competition, and the first of the RecSys Challenges to focus on *user engagement*. Previous challenges had focused on movie rating prediction [1,6], scientific paper recommendation and user targeting [4], as well as point-of-interest recommendation [2]. In recent years, the recommender systems community has put considerable focus on the users of these systems, which is also reflected in the context of the Challenge.

For the 2014 edition, an extended version of the MovieTweets dataset [3] was used. MovieTweets contains movie ratings expressed by Twitter users who have tied their IMDb accounts to their Twitter account. The dataset contains the expected information found in a movie rating dataset, i.e. the user-movie rating matrix, timestamps, etc. This extended dataset also includes the interactions given to tweets that express ratings, i.e. the number of favorites and retweets each of the rating tweets got from other users. This engagement could be considered as an additional signal reflecting e.g., the importance of a given rating with respect to other provided ratings.

Participants were asked to rank a given number of ratings within tweets by their predicted engagement for which we used the sum of the number of retweets and favorite counts as a proxy. The goal of the challenge is to provide insight on user interaction behavior with publicly posted ratings and to learn how being able to predict user engagement may be used for the benefit of recommender systems and its users.

2. DATASET & EVALUATION

For the purpose of the challenge, the MovieTweets dataset was extended with additional meta-data from Twitter API. Therefore, besides the typical rating information consisting of user ids, item ids, ratings and timestamps, also many other types of data such as location and user profile information were available. The dataset contained 212,857 tweets collected over a period of 13 months and was chronologically split up in a training set (80% of the data), a test set (10% of the data) and an evaluation set (the remaining 10%). While the training set contained all data, the *retweet_count* and *favorite_count* data fields were removed in the test and evaluation set. Participants were given the training and test set to allow them to train their models and evaluate their

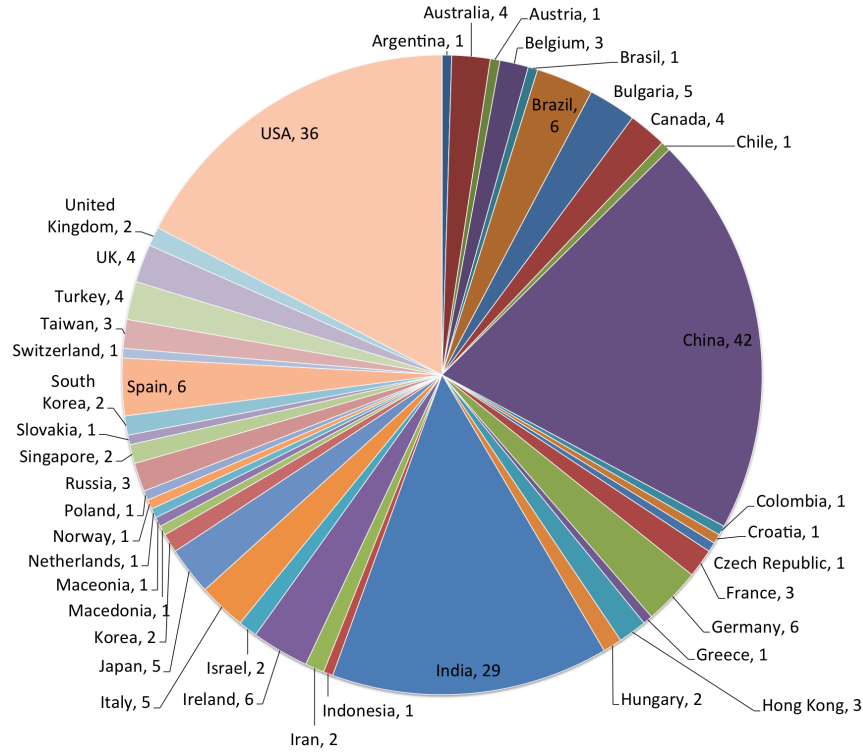


Figure 1: The distribution of participants by country. Note that some team did not state their country and have been omitted in the figure.

results. Final evaluation was however performed by the organizers using the private evaluation set.

For the evaluation the *NDCG@10* metric – as available in the Rival evaluation framework² – was chosen [5]. The metric expresses how well the challenge participants were able to capture the amount of interaction and engagement that each of the tweets in the test set obtained. This is done by comparing the ranked list generated by participant’s algorithm to the ground truth that we provided in the dataset. For each user the true ranked list is considered to be a ranked list of tweets ordered by their *engagement* in descending order.

3. OVERVIEW

The challenge attracted more than 200 teams of international participants, Figure 1 shows the distribution of teams by country. All participating teams were invited to update the challenge leaderboard in order to allow the other teams to compare their current results. The challenge workshop took place on October 10, 2014. At the workshop, participants presented their approaches and discussed methods and possibilities of predicting user engagement.

Acknowledgments

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n°610594.

²<http://rival.recommenders.net>

4. REFERENCES

- [1] Gediminas Adomavicius, Alexander Tuzhilin, Shlomo Berkovsky, Ernesto W. De Luca, and Alan Said, ‘Context-awareness in recommender systems: Research workshop and movie recommendation challenge’, in *Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys ’10*, pp. 385–386, New York, NY, USA, (2010). ACM.
- [2] Jim Blomo, Martin Ester, and Marty Field, ‘Recsys challenge 2013’, in *Proceedings of the 7th ACM Conference on Recommender Systems, RecSys ’13*, pp. 489–490, New York, NY, USA, (2013). ACM.
- [3] Simon Doods, Toon De Pessemier, and Luc Martens, ‘Movietweetings: a movie rating dataset collected from twitter’, in *Workshop on Crowdsourcing and Human Computation for Recommender Systems, CrowdRec at RecSys 2013*, (2013).
- [4] Nikos Manouselis, Alan Said, Domonkos Tikk, Jannis Hermanns, Benjamin Kille, Hendrik Drachsler, Katrien Verbert, and Kris Jack, ‘Recommender systems challenge 2012’, in *Proceedings of the Sixth ACM Conference on Recommender Systems, RecSys ’12*, pp. 353–354, New York, NY, USA, (2012). ACM.
- [5] Alan Said and Alejandro Bellogin, ‘Rival – a toolkit to foster reproducibility in recommender system evaluation’, in *RecSys*, (2014).
- [6] Alan Said, Shlomo Berkovsky, Ernesto William De Luca, and Jannis Hermanns, ‘Challenge on context-aware movie recommendation: Camra2011’, in *Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys ’11*, pp. 385–386, New York, NY, USA, (2011). ACM.