RecSys Challenge 2015 and the YOOCHOOSE Dataset

David Ben-Shimon
YOOCHOOSE Labs Ltd
Omer, Israel
david.ben-shimon@yoochoose.com

Alexander Tsikinovsky
YOOCHOOSE Labs Ltd
Omer, Israel
alexander.tsikinovsky@yoochoose.com

Michael Friedmann
YOOCHOOSE GmbH
Cologne, Germany
michael.friedmann@yoochoose.com

Bracha Shapira
Ben-Gurion University of the Negev
Beer-Sheva, Israel
bshapira@bgu.ac.il

Lior Rokach
Ben-Gurion University of the Negev
Beer-Sheva, Israel
liorrk@bgu.ac.il

Johannes Hoerle
YOOCHOOSE GmbH
Cologne, Germany
johannes.hoerle@yoochoose.com

ABSTRACT

The 2015 ACM Recommender Systems Challenge offered the opportunity to work on a large-scale e-commerce dataset from a big retailer in Europe which is accepting recommender system as a service from YOOCHOOSE. Participants tackled the problem of predicting what items a user intends to purchase, if any, given a click sequence performed during an activity session on the e-commerce website.

The challenge ran for seven months and was very successful, attracting 850 teams from 49 countries which submitted a total of 5,437 solutions. The winners of the challenge scored approximately 50% of the maximum score, which we considered as an impressive achievement. In this paper we provide a brief overview of the challenge and its results.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Data mining

General Terms

Algorithms, Performance, Experimentation

Keywords

Recommender Systems, RecSys Challenge 2015, E-Commerce, YOOCHOOSE

1. INTRODUCTION

Traditionally, the annual ACM Conference on Recommender Systems features a related challenge designed to test participants' skills and algorithms [2][3]. The 2015 RecSys Challenge presented here is associated with ACM RecSys 2015 and organized by the recommender system (RC) provider, YOOCHOOSE[1].

Many small and mid-sized e-commerce businesses outsource the implementation and operation of their RSs. As part of the services offered, RS providers record users' activities and compute statistical models in order to deliver top-N recommendations. In e-businesses the task of delivering the top-N recommendations is more pertinent than the task of rating items, because e-businesses strive to present interesting items to the user. YOOCHOOSE is a RS provider [1] specializing in the calculation of top quality (best matching) top-N recommendations for a variety of purposes, such as generating cross or up-sell, exploiting long tail items, etc., while keeping the user engaged and satisfied.

In this challenge[2], YOOCHOOSE provides a collection of sequence of click events (click sessions) as a training set. Some of the sessions also include buying events. Given a new sequence of clicks, the goal is to predict what items the user will buy, if any. Such information is highly valuable to e-businesses, since it can indicate not only what items to suggest to the user but also how to encourage the user to become a buyer. For example, if the recommender system is able to determine that the user doesn't intend to make a purchase during the current session, it may try to change this behavior by offering the user dedicated promotions, discounts, and the like.

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. Copyright is held by the owner/author(s). RecSys '15, September 16-20, 2015, Vienna, Austria ACM 978-1-4503-3692-5/15/09.
http://dx.doi.org/10.1145/2792838.2798723

2. DATASET & TASK DESCRIPTION

The data represents six months of user activities for a large European e-commerce business that sells a variety of consumer goods including garden tools, toys, clothes, electronics, and more. The training data is comprised of two different files: yoochoose-clicks.dat and yoochoose-buys.dat. Each record in yoochoose-clicks.dat represents a click event and includes the following four fields: Session ID, Time stamp, Item ID, and Item category. Each record in yoochoose-buys.dat represents a purchase and includes the following five fields: Session ID, Time Stamp, Item ID, Item Price, and Quantity. There were 33,040,175 records in the clicks file and 1,177,769 records in the buys file. Together these 9,512,786 unique sessions form the training set. Participants are required to run algorithms on the training data in order to predict which items will be bought during each session in a testing data. The yoochoose-test.dat file consists of 2,312,432 click sessions and is structured identically to the training data's yoochoose-clicks.dat file.

For every item in each of the test file's click sessions the task is to predict whether an item will be purchased during a given session or not. Upon completion of this task, participants must upload their solution/predictions to the challenge website, which in turn calculates a score to this solution immediately and updates the leaderboard online. The solution file to be submitted is comprised of records that have exactly two fields: 1) Session ID and, 2) Comma separated list of item IDs that have been purchased during this session. If a Session ID exists in the test file but does not exist

[1] www.yoochoose.com

[2] http://recsys.yoochoose.net/

in the uploaded solution file, it means that the challenger predicts that the session ends without a buying event. It should be noted that approximately 95% of the sessions in the training set end without a buying event, a distribution was also maintained in the test set.

3. EVALUTION

The task of the challenge can be divided into two sub-goals, first to determine whether a session ends with buying event or not, and if so, to identify what items will be bought. Consequently, we developed a unique measure to emphasize the two sub-goals, as well as to maintain the balance of the importance between the two. Let's note the following:

- Sl – number of sessions in the submitted solution file
- S – number of sessions in the test set
- s – a session in test set
- Sb – set of sessions in test set which end with buy
- As – set of predicted bought items in session s
- Bs – set of actual bought items in session s

Then, equation (1) represents the score of a solution.

$$Score(Sl) = \sum_{\forall s \in Sl} \begin{cases} if\ s \in Sb \rightarrow \frac{Sb}{S} + \frac{|As \cap Bs|}{|As \cup Bs|} \\ else \rightarrow -\frac{Sb}{S} \end{cases} \quad (1)$$

For each session included in the solution file, we add the value of $|Sb|/|S|$ to the overall score in cases which actually includes a buying event. Additionally, the Jaccard score $|As \cap Bs|/|As \cup Bs|$ / is computed between the predicted and actual set of bought items and is later added to the overall score.

If the challenger predicts that the session will end with buying events but lists the wrong items, he still gets a positive score. However, if the participant incorrectly predicts that a session ends with a buying event there is a penalty of $-|Sb|/|S|$ which is subtracted from the overall score for each wrong prediction. The score for a perfect and optimal solution on the test set is 135,176. The winning team achieved a score of 63,102.

4. SCHEDULE, PRIZES, AND PAPERS

The challenge was launched on November 15, 2014, with a final submission deadline of June 15, 2015. The challenge was announced during the 2014 ACM Conference on Recommender Systems, along with monetary prizes in the amounts of 3000€, 1500€, and 500€ for first, second, and third teams respectively.

The winners were determined based on the final ranking of the scores at the end of the challenge. Furthermore, in order to receive a monetary prize, the winners were required to have an accepted workshop paper detailing the applied algorithms, and attend the conference on the day of the challenge workshop. The papers went through a review process taking into account performance, as well as various other aspects such as simplicity, novelty, clarity, sophistication, and the like. Seven papers, out of twenty one submissions, accepted and published in a special volume of ACM SIG proceedings, dedicated to the challenge.

5. CHALLENGE RESULTS

The challenge was very popular primarily because it focused on tackling a current and real e-commerce problem, and also due to its early announcement, incentives, publication of the submitted papers, and the online evaluation system that enabled participants to receive their score upon submission. Figure 1 present distribution by country, figure 2 includes the number of submissions by month throughout the challenge, and figure 3 indicates the maximum score achieved by month.

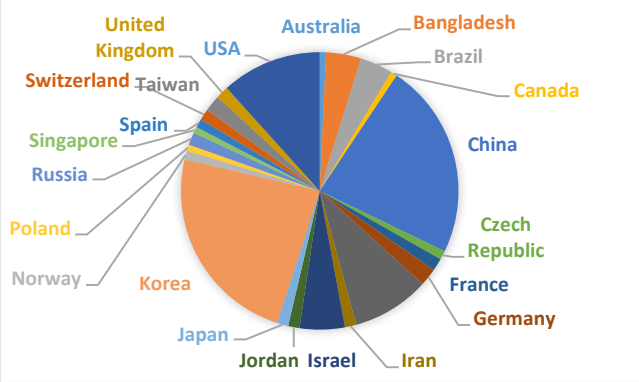

Figure 1. Distribution by Country (for countries with more than 5 teams). 850 teams from 49 countries participated.

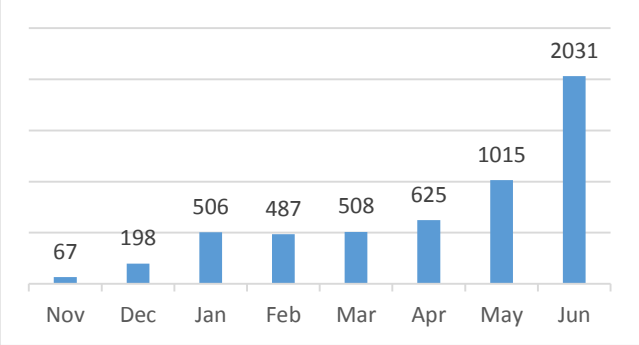

Figure 2. Number of submissions by month. Total of 5,437 solutions submitted during the 7 month challenge.

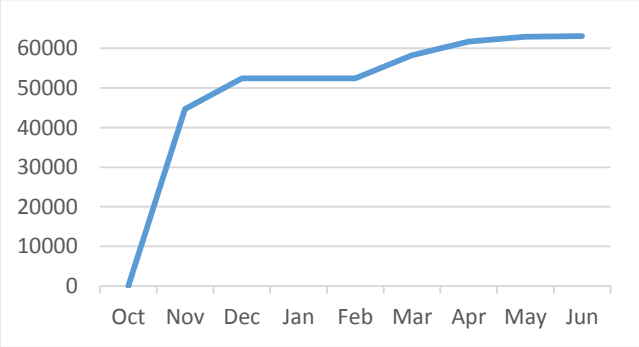

Figure 3. Maximum score achieved by month.

6. REFERENCES

[1] Ben-Shimon D., Friedman M., Hoerle J., Tsikinovsky A., Gude R., Aluchanov R. (2014). "Deploying recommender system for the masses." In Proceedings of the companion publication of the 19th international conference on Intelligent User Interfaces, pp. 1-4. ACM IUI 2014, Haifa, Israel

[2] Blomo, J., Ester, M., Field, M., 'Recsys challenge 2013', in Proceedings of the 7th ACM Conference on Recommender Systems, RecSys '13, pp. 489–490, New York, NY, USA, (2013). ACM..d

[3] Said, A., Dooms, S., Loni, B., & Tikk, D. (2014, October). Recommender systems challenge 2014. In Proceedings of the 8th ACM Conference on Recommender Systems. ACM