RecSys Challenge 2016: Job Recommendations

Fabian Abel
XING AG
Hamburg, Germany
fabian.abel@xing.com

András Benczúr
Hungarian Academy of Sciences
Budapest, Hungary
benczur@sztaki.hu

Daniel Kohlsdorf
XING AG
Hamburg, Germany
daniel.kohlsdorf@xing.com

Martha Larson
TU Delft
Radboud University Nijmegen
Netherlands
m.a.larson@tudelft.nl

Róbert Pálovics
Hungarian Academy of Sciences
Budapest, Hungary
palovics@sztaki.hu

ABSTRACT

The 2016 ACM Recommender Systems Challenge focused on the problem of job recommendations. Given a large dataset from XING that consisted of anonymized user profiles, job postings, and interactions between them, the participating teams had to predict postings that a user will interact with.

The challenge ran for four months with 366 registered teams. 119 of those teams actively participated and submitted together 4,232 solutions yielding in an impressive neck-and-neck race that was decided within the last days of the challenge.

CCS Concepts

•**Information systems** → **Recommender systems;** *Data mining; Test collections;*

Keywords

Recommender Systems; Data Mining Challenge; XING

1. INTRODUCTION

The ACM RecSys Challenge allows researchers and engineers around the world to jointly work on real-world recommender system problems in various domains such as social media [3] or product recommendations [2]. In 2016 the challenge was dedicated to the problem of job recommendation[1] [1]. We released a dataset from XING[2]—a business-oriented social network with around 18 Million users worldwide and more than 10 Million users in Germany.

XING supports people in discovering career opportunities. Job recommendations are therefore an essential part of the XING platform and its mobile apps. Those recommendations aim to satisfy the demands of both the job seekers who have certain preferences concerning their next career step and the recruiters who aim to hire the most appropriate candidate for a given job. In this challenge, we focused on the demands of the job seekers by defining the following task: *Given a XING user, the recommender had to predict those job postings that a user will positively interact with by clicking on it or bookmarking it.*

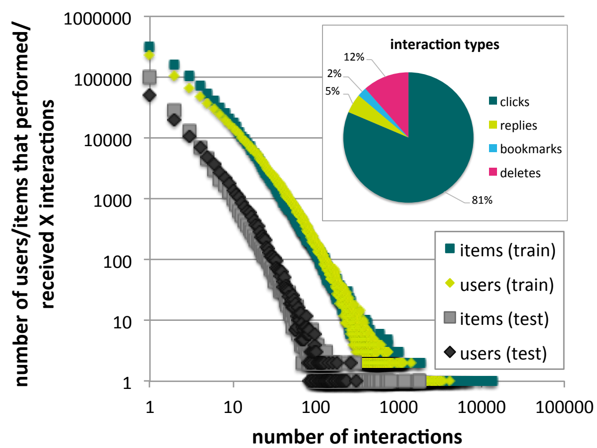


Figure 1: Number of interactions per user and item. 328,618 items (24.2%) and 582,370 users (42.9%) remained without interactions during the training period. More than 80% of the interactions are clicks, followed by deletes, replies and bookmarks. The distributions for training data and test data (ground truth) follow similar characteristics.

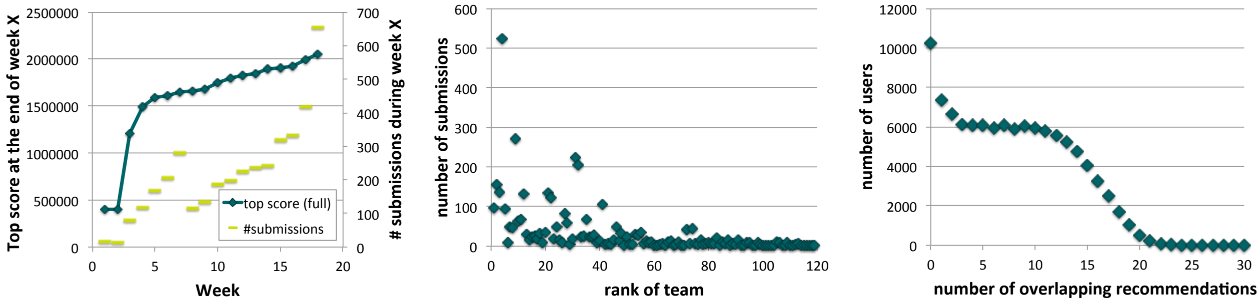
2. DATASET

We provided a training dataset featuring user profiles, job postings, and interactions that users performed on job posts. The dataset also incorporated user-item *impressions*, i.e. information about job postings that were shown to users. The training data[3] contained 1,367,057 users and 1,358,098 items. Users and items were described by several similar attributes such as job categories, career level, industry, lo-

[1]http://2016.recsyschallenge.com
[2]http://xing.com
[3]https://github.com/recsyschallenge/2016/#data

RecSys '16 September 15-19, 2016, Boston , MA, USA

ACM ISBN 978-1-4503-4035-9/16/09.
DOI: http://dx.doi.org/10.1145/2959100.2959207



(a) Top score over time and number of submissions during each week of the challenge.

(b) Number of submissions per team (ordered by final rank of the team).

(c) Overlap between top solution and recommendations generated by XING's job recommender.

Figure 2: Submission statistics.

cation, etc. In addition, the educational background and details about work experience were given for the users.

Around 12 weeks of interaction data between the users and the items was released for training including including (1) *clicks* on job postings, (2) *bookmarks*, (3) *replies* indicating that users intended to apply for the job and (4) *deletes* which corresponds to removing an item that the user no longer wants to see. Figure 1 describes the characteristics of the different interactions. Two weeks of interactions for a set of 150,000 *target users* were used as test data in order to evaluate the submitted solutions. Figure 1 also shows the distribution of that ground truth dataset.

2.1 Anonymization

Anonymization of the training dataset was carried out with an iterative protector/attacker procedure. We took a simple and straightforward approach to threat modeling: The attacker profile was implicit the choice of a highly experienced Data Science team that attempted to de-anonymize the data. At each step the dataset was further bleached, and additional synthetic users were added. Then, tests were carried out to check if the dataset could be de-anonymized by the attacker and also if the dataset supported training a recommender system algorithm effective on a plain-text test set. This procedure ensured that enough information was left in the data to make this to be a useful data set for solving the problem on the actual plain-text data, while at the same time protecting the privacy of XING users.

The bleaching procedure involved replacing named entities with IDs, removing a selection of user attribute values, and removing a selection of interactions. The relative ordering of the interactions was maintained. Synthetic users were created by clustering real users. Further protections included the obvious measure of including only a fraction of XING's users and job postings in the dataset, and also protecting the dataset legally with a user agreement that explicitly prohibits attempts to de-anonymize the dataset, share it, or use it for non-academic purposes.

3. CHALLENGE SETUP AND RESULTS

We used a scoring function as evaluation measure that combines precision@k and recall@k[4]. This evaluation measure was based on the key performance indicators that XING is using to monitor the quality of the job recommender system. Each team was allowed to submit 5 solutions per day and solution files were allowed to feature at most 30 recommendations per user. A public leaderboard that was based on 30% of the ground truth data immediately informed the teams about the performance of their algorithms after their submissions.

The participating teams came from more than 30 different countries such as USA (25%), Germany (11%), China (9%), France (7%) or Hungary (4%). Teams came both from academia (∼25%) and industry (∼75%, most common: *Internet & IT*), e.g. from larger companies such as Yandex or Amazon as well as from smaller start-up companies.

The evolution of the top full score—based on the entire ground truth—is depicted in Figure 2a together with the number of submissions that were performed during that week. The top score thus increased from week to week. In fact, the highest score of the challenge was achieved just 30 minutes before the official submission deadline. Figure 2b reveals that not only the top teams very highly active in submitting solutions, but also many teams that finally ended up at rank 30 and above submitted a high number of solutions.

The top solution achieved a score of 2,052,185 points that is 24.1% of the best possible score. Figure 2c also shows that the top solution managed to include items that XING's system does not recommended. Hence, algorithms developed during the challenge seem to complement XING's recommender ensemble and are likely to have a significant impact on XING's future job recommendations.

[4]https://github.com/recsyschallenge/2016/#evaluation

Acknowledgments. The work leading to these results has received funding from the EU's Seventh Framework Programme (FP7/2007-2013) under CrowdRec Grant Agreement no. 610594.

4. REFERENCES

[1] F. Abel. We know where you should work next summer: Job recommendations. In *Proceedings of ACM RecSys 2015*, pages 230–230, 2015. ACM.

[2] D. Ben-Shimon, A. Tsikinovsky, M. Friedmann, B. Shapira, L. Rokach, and J. Hoerle. RecSys Challenge 2015 and the YOOCHOOSE Dataset. In *Proceedings of ACM RecSys 2015*, pages 357–358, 2015. ACM.

[3] A. Said, S. Dooms, B. Loni, and D. Tikk. Recommender Systems Challenge 2014. In *Proceedings of ACM RecSys 2014*, pages 387–388, 2014. ACM.