Real-time Recommendation of Streamed Data

Frank Hopfgartner
University of Glasgow, UK
frank.hopfgartner@glasgow.ac.uk

Benjamin Kille
TU Berlin, Germany
benjamin.kille@dai-labor.de

Tobias Heintz
plista GmbH, Germany
thz@plista.com

Roberto Turrin
ContentWise, Italy
roberto.turrin@contentwise.tv

ABSTRACT

This tutorial addressed two trending topics in the field of recommender systems research, namely A/B testing and real-time recommendations of streamed data. Focusing on the news domain, participants learned how to benchmark the performance of stream-based recommendation algorithms in a live recommender system and in a simulated environment.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Experimentation

Keywords

Stream Recommendation; A/B testing; Offline Evaluation

1. INTRODUCTION

1.1 Evaluation Paradigms

The release of the Netflix dataset and the associated challenge was a key event that resulted in *offline evaluation* [11] being the most commonly used methodology for the evaluation of recommender systems. Although it comes with clear advantages such as the ability to compare the performance in predicting user ratings with state-of-the-art techniques, it has been criticized for its inability to incorporate the user in the evaluation process. An alternative evaluation paradigm is *online evaluation*, also referred to as A/B testing [1]. Online evaluation aims to benchmark varieties of a recommender system by a larger group of users. It is increasingly adopted for the evaluation of commercial systems with a large user base as it provides the advantage of observing the efficiency of recommendation algorithms under real conditions. However, while online evaluation is the de-facto standard evaluation methodology in Industry, university-based researchers often do not have access to either infrastructure or user base to perform online evaluation on a larger scale. Addressing this deficit, participants learned in this tutorial how they can join a living lab on news recommendation that allows them to perform A/B testing. The living lab and its recommendation scenario are described further in Section 2.

RecSys '15 Vienna, Austria
ACM 978-1-4503-3692-5/15/09.
DOI: http://dx.doi.org/10.1145/2792838.2792839.

1.2 Recommendations of Streamed Data

Offline evaluation allows for the evaluation of research hypotheses that center around modeling recommendation as user-specific selection from static collections of items. While this might be suitable in some domains where the content does not change too often (e.g., in movie databases), it fails in more dynamic domains where items continuously emerge and extend collections, and where existing items become less and less relevant. Examples include news, microblog, or advertisement recommendations [3, 4, 5, 7, 6, 10] where content comes in the form of a constant stream of data.

Streamed data triggers specific challenges for recommender systems. For example, it challenges collaborative filtering as this imposes that the sets of users and items rapidly fluctuate. This tutorial focused on stream-based recommenders which reflect these dynamics. It concentrates on the domain of new recommendation as addressed NewsREEL. The scenario is described in the following section.

2. NEWS ARTICLE RECOMMENDATION

News content published on news portals represents a typical example for streamed data as new content is constantly added to the data corpus. At the same time, the more users visit a news portal, the more requests for providing news recommendations are triggered, resulting in a constant stream of requests. This recommender scenario is realized in the News REcommendation Evaluation Lab (NewsREEL)[1], a campaign style evaluation lab that focuses on benchmarking stream-based recommenders. The NewsREEL challenge consists of two subtasks: In the first subtask, the idea of *living laboratories* is implemented, i.e., researchers gain access to the resources of a company to evaluate different information access techniques using A/B testing. The infrastructure is provided by plista GmbH, a company that provides a recommendation service for online publishers. Whenever a user requests an article from one of their customers' web portals, plista recommends further articles that the user might be interested in. In NewsREEL, plista outsourced this rec-

[1] `http://clef-newsreel.org/`

ommendation task to interested researchers via their Open Recommendation Platform (ORP) [2]. As users visit selected news publishers, ORP randomly forwards recommendation requests to registered participants [8]. The tutorial will detail the idea, technologies, and existing SDK to use ORP. Hence, participants will be able to evaluate their recommendation algorithms in a live system. We argue that this use case represents a unique opportunity for researchers in the area of recommender systems.

The second subtask of NewsREEL focuses on simulating a constant data stream as provided by ORP. In contrast to the first scenario, performing an offline evaluation allows us to issue the same request to different algorithms and subsequently compare them. Additionally, it allows to measure factors such as time and space complexity. In NewsREEL, the Idomaar framework[2] is employed to simulate this data stream. Idomaar is a novel reference framework that was developed in the context of the European project CrowdRec [9]. It adopts open-source technologies widely known by the research community to allow handling of large-scale streams of data (e.g., Apache Kafka, Apache Spark, etc.). Focusing on the news scenario, participants learned how to use this technology to simulate constant data streams.

3. TARGETED AUDIENCE

This tutorial was designed for researchers and practitioners in the field of recommender algorithms. Researchers learned the steps required to setup a recommendation server, connect to ORP and evaluate recommendation algorithms. Practitioners learned about the special characteristics of streamed data for recommender systems. Further, they got to know recommendation algorithms able to deal with such conditions. Participants learned how to use ORP and the reference framework Idomaar to evaluate their recommendation algorithms in an authentic and a simulated setting, respectively.

4. IMPORTANCE OF THIS TOPIC

In recent years, a significant number of papers focused on improving the accuracy when predicting ratings. Typically, rating prediction evaluation starts by partitioning available data into training and test sets. However, commercial recommender systems face dynamics disregarded by this methodology as they often meet an endless stream of interactions between two fluctuating sets.

This tutorial supported researchers to start focusing on streams as recommender systems' input. We showed how participants can approximate an authentic way to evaluate their recommendation algorithms. This can help the community to better distribute their attention to problems urging industrial recommender systems.

5. OBJECTIVES

Having attended the tutorial, participants learned how to tackle the specific challenges of stream-based recommendations, in particular news recommendation. The tutorial highlighted differences between living lab scenarios and stream simulation. The following aspects were covered:

- recommenders operating on streamed data as opposed to static sets of users, items, and interactions
- living laboratories and A/B testing as evaluation methodologies
- specificities of news challenging recommender systems
- comparing recommendation algorithms on streamed data
- frameworks and tools to evaluate news recommenders, in particular ORP and Idomaar
- current trends in the field of real-time recommender systems

As a result, users learned how to implement, evaluate, and continuously improve recommenders for streamed data in general, and news in particular.

[2] `http://rf.crowdrec.eu/`

Acknoledgments

The second and fourth authors are funded by European Union's Seventh Framework Programme (FP7/2007-2013) under CrowdRec Grant Agreement n.610594.

6. REFERENCES

[1] X. Amatriain. Beyond data: from user information to business value through personalized recommendations and consumer science. In *CIKM'13*, pages 2201–2208, 2013.

[2] T. Brodt and F. Hopfgartner. Shedding light on a living lab: the CLEF NEWSREEL open recommendation platform. In *IIiX'14*, pages 223–226, 2014.

[3] J. Chen, R. Nairn, L. Nelson, M. Bernstein, and E. Chi. Short and tweet: Experiments on recommending content from information streams. In *CHI'10*, pages 1185–1194. ACM, 2010.

[4] Y. Chen, P. Berkhin, B. Anderson, and N. R. Devanur. Real-time bidding algorithms for performance-based display ad allocation. In *KDD '11*, pages 1307–1315. ACM, 2011.

[5] E. Diaz-Aviles, L. Drumond, L. Schmidt-Thieme, and W. Nejdl. Real-time top-n recommendation in social streams. In *RecSys'12*, pages 59–66, 2012.

[6] D. Doychev, R. Rafter, A. Lawlor, and B. Smyth. News recommenders: Real-time, real-life experiences. In *UMAP'15*, pages 337–342, 2015.

[7] F. Garcin, B. Faltings, O. Donatsch, A. Alazzawi, C. Bruttin, and A. Huber. Offline and online evaluation of news recommender systems at swissinfo.ch. In *RecSys'14*, pages 169–176. ACM, 2014.

[8] F. Hopfgartner, B. Kille, A. Lommatzsch, T. Plumbaum, T. Brodt, and T. Heintz. Benchmarking news recommendations in a living lab. In *CLEF'14*, pages 250–267, 2014.

[9] B. Kille, A. Lommatzsch, R. Turrin, A. Sereny, M. Larson, T. Brodt, J. Seiler, and F. Hopfgartner. Stream-based recommendations: Online and offline evaluation as a service. In *CLEF'15*, 2015.

[10] A. Lommatzsch. Real-time recommendations for user-item streams. In *SAC'15*, pages 1039–1046, 2015.

[11] G. Shani and A. Gunawardana. Evaluating recommendation systems. In *Recommender Systems Handbook*, pages 257–297. 2011.