SIGIR 2018 Workshop on eCommerce (ECOM18)

Jon Degenhardt
eBay Inc.
jdegenhardt@ebay.com

Pino Di Fabbrizio
Rakuten Institute of Technology
difabbrizio@gmail.com

Surya Kallumadi
Kansas State University
surya@ksu.edu

Mohit Kumar
Flipkart, India
k.mohit@flipkart.com

Yiu-Chang Lin
Rakuten Institute of Technology
yiuchang.lin@rakuten.com

Andrew Trotman
University of Otago
andrew@cs.otago.ac.nz

Huasha Zhao
Alibaba Group
huasha.zhao@alibaba-inc.com

ABSTRACT

eCommerce Information Retrieval has received little attention in the academic literature, yet it is an essential component of some of the largest web sites (such as eBay, Amazon, Airbnb, Alibaba, Taobao, Target, Facebook, and others). SIGIR has for several years seen sponsorship from these kinds of organisations, who clearly value the importance of research into Information Retrieval. The purpose of this workshop is to bring together researchers and practitioners of eCommerce IR to discuss topics unique to it, to set a research agenda, and to examine how to build datasets for research into this fascinating topic.

eCommerce IR is ripe for research and has a unique set of problems. For example, in eCommerce search there may be no hypertext links between documents (products); there is a click stream, but more importantly, there is often a buy stream. eCommerce problems are wide in scope and range from user interaction modalities (the kinds of search seen in when buying are different from those of web-page search (i.e. it is not clear how shopping and buying relate to the standard web-search interaction models)) through to dynamic updates of a rapidly changing collection on auction sites, and the experienceness of some products (such as Airbnb bookings).

This workshop is a follow up to the "SIGIR 2017 workshop on eCommerce (ECOM17)", which was organized at SIGIR 2017, Tokyo. In the 2018 workshop, in addition to a data challenge, we will be following up on multiple aspects that were discussed in the 2017 workshop.

CCS CONCEPTS

• **Information systems → Environment-specific retrieval**;

KEYWORDS

eCommerce, Product Search, Recommendation

1 INTRODUCTION

Search, ranking and recommendation have applications ranging from traditional web search to document databases to vertical search systems. In this workshop we will explore approaches for search and recommendations of products, and other related aspects of eCommerce Information Retrieval. Although the task is the same as web-page search (fulfill a user's information need), the way in which this is achieved is very much different. On product sites (such as eBay, Flipkart, Amazon, and Alibaba), the traditional web-page ranking features are either not present or are present in a different form.

The entities that need to be discovered (the information that fulfills the need) might be unstructured, associated with structure, semi-structured, or have facets such as: price, ratings, title, description, seller location, and so on. Domains with such facets raise interesting research challenges such as a) relevance and ranking functions that take into account the tradeoffs across various facets with respect to the input query b) recommendations based on entity similarity c) recommendations based on user location (e.g. shipping cost), and so on. In the case of eCommerce search and recommendation these challenges require inherent understanding of product attributes, user behavior, and the query context. Product sites are also characterized by the presence of a dynamic inventory with a high rate of change and turnover, and a long tail of query distribution.

Outside of search but still within Information Retrieval, the same feature in different domains can have radically different meaning. For example, in email filtering the presence of "Ray-Ban" along with a price is a strong indication of spam, but within an auction setting this likely indicates a valid product for sale. Another example is natural language translation; company names, product names, and even product descriptions do not translate well with existing tools. Similar problems exist with knowledge graphs that are not customised to match the product domain.

This workshop will bring together researchers and practitioners to identify a set of core research questions in eCommerce search and recommendation. This will bring the community together in a way that has never happened before ECOM17. It will help attract research funding to search in this domain. It will help attract researchers and postgraduate students to eCommerce and product search. Finally, it will help broaden the definition of Information Retrieval at conferences such as SIGIR.

The workshop will also examine the problem of data availability. As the purpose of a product site is to make data on entities available, the same security concerns that plague other search domains do not exist. However sales and seller information is private and proprietary and likely to be unavailable. We hope that the discussion on data will result in both a proposal to release data that can be put to an eCommerce site, as well as some tasks that can be examined on that data set – indeed the data challenge is a success in this area.

2 SIGIR 2017 WORKSHOP ON ECOMMERCE (ECOM17)

The SIGIR 2017 Workshop on eCommerce (ECOM17)[1] was the first time this workshop was held. ECOM17 was a full day workshop, and about 60 people registered to attend [1]. ECOM17 had a rich and diverse set of contributions and discussions ranging from eCommerce systems architecture to product search and recommendations to fraud detection. The workshop received twenty one submissions from both industry and academia, and accepted thirteen submissions (62% acceptance rate). SIGIR was an excellent venue for the workshop. This workshop was a first and a much needed forum for bringing together practitioners from industry and academia working in the eCommerce domain, and acted as a venue for eCommerce research.

The need to have an active community to discuss problems and potential opportunities in the eCommerce domain came to the forefront in the group discussion and panel sessions. Availability of high quality data sets coupled with well defined problem statements is another aspect identified that would move the field forward. We intend to address this in ECOM18 in the form of a data challenge. A number of participants from industry were willing to contribute data sets and identify interesting research areas. The workshop attendees were enthusiastic in their agreement that there should be future workshops on this topic. A report on ECOM17 was recently published in SIGIR Forum [2].

3 THEME AND PURPOSE

The primary theme of the workshop is eCommerce search and recommendation.

The purpose of the workshop is to provide a venue for publication and discussion of Information Retrieval research and ideas as they pertain to products and eCommerce. We will be bringing together practitioners and researchers from academia and industry to discuss the challenges and approaches to search and recommendation. In particular, how to get data.

The goal is to foster collaboration, discussion in the broader IR community. We are happy to state that we have an agenda to raise awareness within the academic community of the problems faced by this domain.

3.1 Scope

The workshop relates to all aspects of eCommerce search and recommendations. Research topics and challenges that are usually encountered in this domain include:

- Machine learning techniques such as online learning and deep learning for eCommerce applications
- Semantic representation for users, products and services & Semantic understanding of queries
- Structured data and faceted search
- The use of domain specific facets in search and other IR tasks
- Temporal dynamics for Search and Recommendation
- Models for relevance and ranking for multi-faceted entities
- Deterministic (and other) sorting of results lists (e.g. price low to high including postage)
- Personalized search and recommendations
- Session aware and session oriented search and recommendation
- Inventory display issues (example: legal, ethical, and spam)
- Cold start issues
- Personalization and the use of personal facets such as age, gender, location etc.
- Indexing efficiency incorporating structured product data and catalog information
- Indexing and search in a rapidly changing environment (for example, an auction site)
- Scalability
- Diversity in product search and recommendations
- Strategies for resolving extremely low (or no) recall queries
- Query intent
- Semantic understanding of queries
- The use of external features such as reviews and ratings in ranking
- User interfaces and personalization
- Reviews and sentiment analysis
- The use of social signals in ranking and beyond
- The balance between business requirements and user requirements (revenue vs relevance)
- Trust
- Live experimentation
- Desktop and mobile issues
- Questions and answering, chatbots for eCommerce
- Conversational commerce: shopping using voice assistants such as Amazon Alexa and Google Now
- Fashion eCommerce

3.2 Data Challenge

The workshop is hosting a machine learning data challenge targeting taxonomy classification for eCommerce-scale product catalogs[2]. The data set contributed by Rakuten[3] contains 1 Million titles spread across more than 3000 leaf categories, with unbalanced class sizes.

---

[1] ECOM17 workshop website: http://sigir-ecom.weebly.com/

[2] Rakuten Data Challenge: https://sigir-ecom.github.io/data-task.html
[3] The data is provided by Rakuten Institute of Technology, Boston (RIT-Boston), a dedicated R&D organization for the Rakuten group

This challenge focuses on the topic of large-scale taxonomy classification where the goal is to predict each product's category as defined in the taxonomy tree given a product's title. For example, in the Rakuten.com catalog, "Dr. Martens Air Wair 1460 Mens Leather Ankle Boots" is categorized under the "Clothing, Shoes & Accessories -> Shoes -> Men -> Boots" leaf. Manual and rule based approaches to categorization are not scalable since commercial product taxonomies are organized in tree structures with three to ten levels of depth and thousands of leaf nodes. The cataloging of product listings through taxonomy categorization is a fundamental problem for any eCommerce marketplace, with applications ranging from personalized search & recommendation to query understanding.

The challenge presents several interesting research aspects due to the intrinsic noisy nature of the product labels, the size of modern eCommerce catalogs, and the typical unbalanced data distribution. Advances in this area of research have been limited due to the lack of real data from actual commercial catalogs. Making the data available to the participants will attract research institutions and practitioners who did not have the opportunity to contribute their ideas due to the previous lack of data.

3.3 Publish open datasets to facilitate research in eCommerce IR

During the group and panel discussions of ECOM17, a number of participants from industry were willing to identify interesting research problems, and provide data sets tailored for these research problems. In order to encourage this, we have a data track where participants can publish open data sets along with details regarding necessary metrics and replicable baselines. In our opinion the availability of such open datasets and baselines would help researchers advance the state of the field in this area.

4 WORKSHOP FORMAT

Similar to ECOM17, ECOM18 will be a full day workshop. The workshop will start with an invited talk from a well respected practitioner (or academic) who is tackling eCommerce search problems.

The invited talk will then be followed by the presentation of selected papers. A call asking not only for research papers, but also for position and opinion papers, posters, and dataset papers has been circulated. All submitted papers and posters will be single-blind, peer reviewed by an international program committee of researchers of high repute. Accepted papers will be presented with ample time for discussion; there will also be a poster session fostering further discussion.

After the thought provoking invited talk and presentations we will break-out into small groups to identify key areas for future research in eCommerce search. Each group will identify one key research problem, endeavour to identify the challenges and opportunities (which will likely include the issue of data availability), and report back to the workshop on their findings. The ideas that come from this break-out session will form the basis of a research agenda.

The final session of the day will be a general discussion session. The primary topic for discussion will be data availability. Specifically, what guarantees need to be in place for an organisation like eBay to make a dump available, and can we meet those guarantees?

The workshop schedule and activities are structured to contain substantial time for discussion and engagement by all participants.

5 PARTICIPATION

This workshop is open to all interested parties.

6 WORKSHOP OUTCOMES

We believe that the most important outcome of the workshop is the discussion between individuals at the workshop. It is these discussions that lead to collaboration and future research - unarguably the ultimate goal of any workshop. This, however, is hard to directly capture. We will capture what we can in the form of a SIGIR Forum workshop report. We will produce a proceedings of the workshop and work with CEUR Workshop Proceedings to ensure they are appropriately archived.

We expect that there will be lively discussion on data availability. Since there will be representatives from eCommerce sites present, we will engage them in discussions on the requirements for data release. Although the release of data is not a specific goal of this workshop, we expect it to pave the way for such an event in the future. We hope that the challenge data will serve as a model to openness.

An additional goal is to raise awareness of the fascinating problems that litter the eCommerce search battlefield. We hope that through discussion at the workshop, and SIGIR 2018, we can help steer the research community towards these problems and in doing so find solutions to some very difficult problems.

REFERENCES

[1] Jon Degenhardt, Surya Kallumadi, Maarten de Rijke, Luo Si, Andrew Trotman, and Yinghui Xu. 2017. SIGIR 2017 Workshop on eCommerce (ECOM17). In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*. ACM, New York, NY, USA, 1425–1426. https://doi.org/10.1145/3077136.3084367

[2] Jon Degenhardt, Surya Kallumadi, Yiu-Chang Lin, Maarten de Rijke, Luo Si, Andrew Trotman, Sindhuja Venkatesh, and Xu Yinghui. 2018. Report on the SIGIR 2017 Workshop on eCommerce (ECOM17). *SIGIR Forum* 51, 3 (Feb. 2018), 128–138. https://doi.org/10.1145/3190580.3190600