ECOM'20: The SIGIR 2020 Workshop on eCommerce

Dietmar Jannach
University of Klagenfurt
dietmar.jannach@aau.at

Surya Kallumadi
The Home Depot
surya@ksu.edu

Tracy Holloway King
Adobe
tking@adobe.com

Weihua Luo
Alibaba
weihua.luowh@alibaba-inc.com

Shervin Malmasi
Amazon
malmasi@amazon.com

ABSTRACT

eCommerce Information Retrieval (IR) is receiving increasing attention in the academic literature and is an essential component of some of the largest web sites (e.g. Amazon, Alibaba, Taobao, eBay, Airbnb, Target, Facebook). eCommerce organisations consistently sponsor SIGIR, reflecting the importance of IR research to them. This workshop (1) brings together researchers and practitioners of eCommerce IR to discuss topics unique to it, (2) determines how to use eCommerce's unique combination of free text, structured data, and customer behavioral data to improve search relevance, and (3) examines how to build data sets and evaluate algorithms in this domain. Since eCommerce customers often do not know exactly what they want to buy, recommendations are valuable for inspiration, serendipitous discovery and basket building. The theme of this year's eCommerce IR workshop is integrating recommendations into search for eCommerce. In addition to the focus on recommender systems in eCommerce search, Rakuten France is sponsoring a data challenge on taxonomy classification using multimodal (image, text and structured data) input. The data challenge reflects themes from the 2017–2019 SIGIR workshops.

1 MOTIVATION

Search and recommendation have applications ranging from traditional web search to document databases to vertical search systems. In this workshop, we explore approaches to search and recommendation of products in eCommerce IR. Although the basic search task (i.e. to fulfill a user's information need) is the same for web-page search and eCommerce search, the way in which this is achieved is different. On eCommerce sites such as Alibaba, Amazon, eBay, Flipkart, and Home Depot, the data available for retrieval and ranking are different as are the signals of success (e.g. adding items to a cart, purchasing).

The entities that need to be discovered are documents which are combinations of unstructured text (e.g. titles, descriptions, reviews), images, and structured data (e.g. price, brand, ratings, seller location). This complex combination of data raises interesting research

challenges including recall (matching) and ranking functions that take into account the trade-offs across various facets with respect to the input query. The features available for building the click models used in ranking are different, and often stronger, in eCommerce than in web search. As well as queries, hover time, clicks, and browse time, eCommerce sites also have signals from add-to-cart, purchase, remove-from-cart, return of goods, etc. When incorporating promotions and personalization such as individual pricing, the click models are more complex than seen for web search. eCommerce is also characterized by dynamic inventory with a high rate of change and turnover and a classic long tail query distribution.

Since eCommerce customers often do not know exactly what they want to buy, recommendations are valuable for inspiration, serendipitous discovery, and basket building. The theme of this year's workshop is integrating recommendations into search for eCommerce. Recommendations can be based on item (i.e. content) similarity, collaborative filtering and other customer behavior similarity, or a combination of both. In the case of eCommerce search and recommendation, providing high quality results requires inherent understanding of product attributes, customer behavior, and the query context.

This workshop brings together researchers and practitioners to identify and discuss core research problems in eCommerce search and recommendation. The workshop aims to foster collaboration by bringing the community together, to attract research funding to this increasing important domain and to introduce researchers and postgraduate students to eCommerce and product search. Finally, it helps broaden the definition of Information Retrieval (IR) at research venues such as SIGIR.

The workshop supports data availability for eCommerce IR research. In 2018 and 2019 ECOM released data from Rakuten and eBay respectively. This year Rakuten France is releasing data for a large scale multi-modal product classification task. As the purpose of an eCommerce site is to make product data available to potential consumers, the privacy concerns that plague certain search domains do not apply to eCommerce product data, although such concerns still apply to query and customer behavior data.

2 SCOPE

To support the eCommerce search and recommendation theme and the goal to provide a venue for discussion, collaboration and publication of IR research and ideas as they pertain to eCommerce, the workshop relates to all aspects of eCommerce search and recommendations. Research topics and challenges that are frequently encountered in this domain include:

- **Ranking and Whole Page Relevance:** Diversity in product search and recommendations; Relevance models for multi-faceted entities; The balance between business requirements and customer requirements (revenue vs. relevance); Ranking features (textual, image, structured data, customer behavior, reviews, ratings, social signals, etc.); Deterministic sorts (e.g. price low to high); Temporal dynamics and seasonality
- **Query Understanding:** Query intent, query suggestions, and auto-completion; Strategies for resolving low or zero recall queries; Converting across modalities (e.g. text, structured data, images)
- **Document Understanding:** Categorization and facets; Reviews and sentiment analysis
- **Recommendation and Personalization:** Personalization and contextualization, including the use of personal facets such as age, gender, location; Blending recommendations and search results
- **Representations and Data:** Semantic representation of products, queries, and customers; Construction and use of knowledge graphs for eCommerce
- **IR Fundamentals for eCommerce:** Cross-lingual search and machine translation; Machine learning techniques for eCommerce applications; Indexing and search in rapidly changing environments (e.g. auction sites); Experimentation techniques including AB testing and Multi-armed bandits
- **Other Challenges:** UX for eCommerce; The role of search in trust and security for marketplaces; Question answering and chat bots for eCommerce

3 DATA CHALLENGE

At the 2018 workshop, we collaborated with Rakuten to run a data challenge "Taxonomy Classification for eCommerce-scale Product Catalogs"[1] which addressed the problem of taking a product listing and choosing which category (from a taxonomy) that listing should belong. Academic, industrial, and independent participation was high, with 28 teams submitting runs. We worked with Rakuten to release the data from the challenge so that others can continue research on this topic.

For the 2019 workshop, eBay Inc. released item data as well as queries and assessments for a "High Accuracy Recall Task".[2] This task examined relevance of the recall set for deterministic sort orders (e.g. price low to high) and other tasks not seen in web search.

The 2020 workshop hosts a machine learning data challenge targeting taxonomy classification for eCommerce-scale multi-modal product catalogs. ECOM'20 provides data including several million product titles, images and descriptions from the French catalog of Rakuten.com, featuring thousands of taxonomy-structured labels. The cataloging of product listings through taxonomy categorization is a fundamental problem for eCommerce marketplaces, with applications ranging from query understanding to personalized search and recommendation. The challenge reflects multiple research aspects due to the intrinsic noisy nature of the product labels, the size of modern eCommerce catalogs, the multi-modal nature of the

data, and the typical unbalanced data distribution. Advances in this area of research have been limited due to the lack of real data from commercial catalogs. Making the data available to the participants attracts research institutions and practitioners who have not had the opportunity to contribute their ideas due to the previous lack of data.

The ECOM'20 data challenge will have an online leader board and maintain a Slack channel as in 2018 and 2019.

4 WORKSHOP FORMAT

This is the first virtual ECOM workshop. The workshop will be anchored by a panel discussion on the interplay of search and recommendations in eCommerce from a scientific, technical, and product perspective. There will be several invited talks by academic and industry eCommerce experts to provide a broad perspective on the current and future state of the field. The winners of the data challenge and the best two accepted papers will present their research. In place of the traditional poster session, all other accepted papers will have short pre-recorded video of their paper as a poster "teaser" and then lead discussion breakouts on their work.

5 WORKSHOP OUTCOMES

The most important outcome of the workshop is the interactions between delegates. These interactions lead to collaboration and future research — unarguably the ultimate goal of any workshop. The workshop format is designed to support this outcome despite the limitations of a virtual workshop. An additional goal is to raise awareness of the fascinating problems that litter the eCommerce search battlefield. We hope that through the workshop and SIGIR 2020, we can help steer the research community towards these unique IR problems.

---

[1]https://sigir-ecom.github.io/ecom2018/data-task.html
[2]https://sigir-ecom.github.io/data-task.html