E-commerce Product Search: Personalization, Diversification, and beyond

Atish Das Sarma
eBay Research Labs
atish.dassarma@gmail.com

Nish Parikh
eBay Data Labs
nparikh@ebay.com

Neel Sundaresan
eBay Data Labs
nsundaresan@ebay.com

ABSTRACT

In this tutorial we discuss challenges, techniques and analytics in search ranking particularly applied to product search in e-commerce. Several challenges appear in this context, both from a research as well as an application standpoint. We present various approaches adopted in the industry, review well-known research techniques developed over the last decade, draw parallels to traditional web search highlighting the new challenges in this setting, and dig deep into some of the algorithmic and technical approaches developed. A specific approach that advances theoretical techniques and illustrates practical impact considered here is of identifying most suited results quickly from a large database. Settings span cold start users and advanced users for whom personalization is possible. In this context, top-k and skylines are discussed as they form a key approach that spans the web, data mining, and database communities. These present powerful tools for search across multi-dimensional items with clear preferences within each attribute, like product search as opposed to regular web search.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Storage and Retrieval - Information Search and Retrieval

Keywords

Search; e-commerce; top-k; skylines; algorithms

1. MAIN TOPICS

Problems in search and ranking ranges from traditional relational databases to document systems to web search and, more recently, vertical applications like product search in e-commerce or vertical search applications in insurance, finance, or health. This tutorial addresses the intersection of web, data mining, and the database research and also touches upon user interface issues and interaction experiences. A specific focus in e-commerce involves returning a small set of results to users, as well as requiring minimal intervention or input from them. These two constraints are main concerns with skylines and top-k, respectively, and therefore have drawn a

great deal of attention in the recent years with several interesting ideas being proposed in the research community. Our work describes several of these approaches. We talk about the research and application of some of these techniques to e-commerce websites like eBay.

We have developed a series of techniques that introduce new approaches in this realm and have shown formal theoretical guarantees as well as direct applications to real world settings. The majority of the tutorial is based on the following papers.

- **Identifying representative subset of the skyline set:** We focus on our work on efficient skyline algorithms [1] and regret-minimizing representative databases [2]. For the former, we consider a streaming model and present theoretically provable bounds. The latter involves modeling users with unknown utility functions. Given a list of k tuples, we say that a user is $x\%$ satisfied with the list if the utility she obtains from the best tuple in this list is at least $x\%$ of the utility she obtains from the best tuple in the whole database. We show an algorithm exists that outputs a small set of k tuples that makes *every* user at least $x\%$ satisfied.

- **Interaction based approaches:** Our work in this technique builds on [2] and presents an interactive regret minimization framework [3]. In this paper, we adopt the notion of *maximum regret ratio when users have linear utility functions* proposed in [2]. Here we study how interactions through help further improve the guarantees on user happiness.

- **Web Search Applications:** Web search ranking and relevance, and the importance of Top-k has been well studied. User attention and response to search results decay non-linearly by rank [4]. Measures like DCG and NDCG are well studied in the IR community and have been applied in web search systems, advertising modeling systems [5, 6]. This tutorial delves in to some of these techniques originally designed for generic web search and contrasts them with approaches specifically designed for e-commerce product search.

- **e-commerce product search:** More specifically in commerce search top-k plays an important role as product placement at the right rank has revenue implications. Relevance, diversity, and revenue all play a role in this. One of the primary works we focus on this topic is our industry application paper that goes beyond relevance in marketplace search [7]. In this paper we study diversity and its relations to search relevance in the context of an online marketplace. We conduct a large-scale log-based study using click-stream data and introduce three main metrics: selection (diversity), trust, and value. In our analysis we also show how these interact with

relevance in different ways. We also present a user-tunable approach to marketplace search demonstrated in [8]. Further, we show data mining techniques for commerce search queries that have low recall [9].

- **Social meets e-commerce:** There has been considerable effort in recent years to integrate signals revealed in the context of social behavior towards improving product recommendations. We will present some learnings from such investigations, touching upon work in [10, 11].

2. REFERENCES

[1] A. Das Sarma, A. Lall, D. Nanongkai, and J. J. Xu, "Randomized multi-pass streaming skyline algorithms," *PVLDB*, vol. 2, no. 1, pp. 85–96, 2009.

[2] D. Nanongkai, A. Das Sarma, A. Lall, R. J. Lipton, and J. J. Xu, "Regret-minimizing representative databases," *PVLDB*, vol. 3, no. 1, pp. 1114–1124, 2010.

[3] D. Nanongkai, A. Lall, A. Das Sarma, and K. Makino, "Interactive regret minimization," in *SIGMOD Conference*, 2012, pp. 109–120.

[4] T. Joachims, "Optimizing search engines using clickthrough data," in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2002, pp. 133–142.

[5] W. B. Croft, D. Metzler, and T. Strohman, *Search engines: Information retrieval in practice*. Addison-Wesley Reading, 2010.

[6] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of ir techniques," *ACM Transactions on Information Systems (TOIS)*, vol. 20, no. 4, pp. 422–446, 2002.

[7] N. Parikh and N. Sundaresan, "Beyond relevance in marketplace search," in *CIKM*, 2011, pp. 2109–2112.

[8] ——, "A user-tunable approach to marketplace search," in *WWW (Companion Volume)*, 2011, pp. 245–248.

[9] G. Singh, N. Parikh, and N. Sundaresan, "Rewriting null e-commerce queries to recommend products," in *WWW (Companion Volume)*, 2012, pp. 73–82.

[10] A. Das Sarma, S. Si, E. F. Churchill, and N. Sundaresan, "The "expression gap": Do you like what you share?" in *WWW (Companion Volume)*, 2014.

[11] S. Si, A. Das Sarma, E. F. Churchill, and N. Sundaresan, "Beyond modeling private actions: Predicting social shares," in *WWW (Companion Volume)*, 2014.