

Personalization Approaches for Ranking: A Review and Research Experiments

Madhuri A. Potey, Savitribai Phule Pune University, Pune, India

Pradeep K. Sinha, College of Engineering, Pune, India

ABSTRACT

Search engine technologies are evolving to satisfy the user's ever increasing information need; but are yet to achieve perfection especially in ranking. With the exponential growth in the available information on the internet; ranking has become vital for satisfactory search experience. User satisfaction can be ensured to some extent by personalizing the search results based on user preferences which can be explicitly stated or learned from user's search behavior. Machine learning algorithms which predict user preference from the available information related to the user are extensively experimented for personalization. Among several studies undertaken for re-ranking the documents, many focus on the user. Such approaches create user model to capture the search context and behavior. This study attempts to analyze the research trends in user model based personalization and discuss experimental results in personalized information retrieval area. The authors experimented to extend the state of the art in the specific areas of personalization.

KEYWORDS

Adaptive User Model, Concept Hierarchy, Context Modeling, Implicit Relevance, Personalization, Query Expansion, Query Intent, Ranking, Re-ranking, Social Bookmarking, User Model

INTRODUCTION

Information overload is widely recognized problem addressed by the Information Retrieval research community. Personalization has been actively studied to reduce the information overheads to leverage user-specific retrieval and existing works in the web search domain. Traditional retrieval system may train and apply a single ranking model across all users, however user's information needs are often diverse; hence single generic ranking model is ineffective for personalization. Personalized Information Retrieval (PIR) focuses on user satisfaction by making use of the diverse information about user such as user interests, queries, weblogs etc. to learn specific model of user interests and uses it to re-rank the results retrieved using the generic model. Some form of user model is inevitable in the personalization process. User model can be explicitly learned by capturing user's interests and feedback (Chirita, 2007). Alternatively, the user's search behavior can be used to implicitly learn about user preferences and create user models. The research trends in user model-based personalization in information retrieval focus on the retrieval process at three distinct stages (Micarelli, et al. 2007); as a part of actual retrieval process, information need specification, and re-ranking of retrieved results. The first technique may provide quick query response as traditional ranking system is directly adapted for personalization. However, the major drawback of applying personalization as a part of retrieval takes a longer time compared to traditional IR techniques. Since commercial search engines are bound by time constraints that mandate the results to be presented quickly; such constraints cannot be met

for all users. Hence we have focused on later two phases where the user model based personalization is experimented at querying stage or at re-ranking stage after retrieval and the results obtained are encouraging.

PERSONALIZATION APPROACHES AT QUERYING STAGE

Query logs record user's search queries and related actions; which is an important source to learn user's search behavior implicitly. Web Search technology may collect information about user in general and then identify groups of user to improvise ranking. Users may explicitly specify interests, demographics or cognitive characteristics. One disadvantage of such explicit feedback is that it is difficult to collect and maintain it for temporal personalization. The user models can also be built implicitly based on content the user has clicked on or the web history of interaction. These models are suitable for personalization using query adaptations or query modification where query is improvised to better represent the information need based on the user model. Short queries can be augmented with additional words to minimize the vocabulary problem, such as polysemy and synonymy, which are prevalent in keyword-based search. Alternatively, if the query retrieves a smaller number of resources than expected, it is possible to expand it using words or phrases with a similar meaning or applying some other statistical relations to the set of relevant documents [Micarelli et al.2007]. The major advantage of query modification approach is that the amount of work required to retrieve the results is the same as in the non-personalized search. On the other hand, query modification approach is less likely to affect the result lists as it does not affect all stages of ranking process.

Query Modifications and Augmentation Experiments

Query logs; often in a form of click-through data which capture user search actions are rich source of information about the user's search intent. The challenge is to effectively extract useful knowledge about user from very large query logs which are generally in raw format. Approaches are proposed by different researchers to retrieve the significant features from query log effectively and derive a user model. We reviewed literature to identify different approaches proposed by researchers in order to extract essential features from query log (Dayong et al., 2010; Chirita, 2007) Techniques for query log processing have been explored by previous works in this area (Hofmann et al., 2009; Radlinski et al., 2005; Radlinski et al., 2010). It was analyzed that identification of query intent of the user can improve search results. We improved upon the query intent technique by combining it with other features identified for improving search experience. Query log processing involves 1) Query Reformulation and Suggestion; 2) Log analysis and 3) Query intent identification. Query can help user to state the information need explicitly. However implicitly capturing the search goal using only query is difficult. Although query is explicitly stated, the underlying search goals of a user are often hidden. Personalization systems need to capture the search intent. Query chain is the successive reformulation of web queries done by user in same session. The query log helps to detect query chains and by traversing those chains, it is possible to identify correct underlying intention while submitting the query to search engine. We experimented with intent detection using query chains in combination with bigram features, term features, content features. The goal of the experiment was to improve classification of web query intent and relevant result retrieval using identified intent. The query log exploits query text related features to classify query intention. We adopted Broder's taxonomy (Broder, 2002) for classifying web query intent as informational (searching information on web related to particular topic), navigational (user wants to reach a particular website) or transactional (user wants to perform web-mediated task). The query chains are formed when a user may reword query to get more relevant results by correcting a typographic error or by increased or decreased specificity. Hence it is the useful resource to extract user's behavior or intent of finding information.

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

www.igi-global.com/article/personalization-approaches-for-ranking/165376?camid=4v1

This title is available in InfoSci-Journals, InfoSci-Journal Disciplines Library Science, Information Studies, and Education, InfoSci-Knowledge Discovery, Information Management, and Storage eJournal Collection. Recommend this product to your librarian:

www.igi-global.com/e-resources/library-recommendation/?id=2

Related Content

Innovative Product Design using Metaontology with Semantic TRIZ

Koswatte R. C. Koswatte, Incheon Paik, Wonhee Park and Banage T. G. S. Kumara (2015). *International Journal of Information Retrieval Research* (pp. 43-65).

www.igi-global.com/article/innovative-product-design-using-metaontology-with-semantic-triz/130007?camid=4v1a

Hadoop Map Only Job for Enciphering Patient-Generated Health Data

Arushi Jain and Vishal Bhatnagar (2017). *International Journal of Information Retrieval Research* (pp. 72-86).

www.igi-global.com/article/hadoop-map-only-job-for-enciphering-patient-generated-health-data/186826?camid=4v1a

Retrieval of Web Pages on Real-World Events related to Physical Objects

Takeshi Okadome, Hajime Funai, Sho Ito, Junya Nakajima and Koh Kakusho (2012). *International Journal of Information Retrieval Research* (pp. 65-80).

www.igi-global.com/article/retrieval-web-pages-real-world/72706?camid=4v1a

Effectiveness of Visualization for Information Retrieval through Ontologies with Entity Evolution: The Impact of Ontology Modeling

Akrivi Katifori, Costas Vassilakis, George Lepouras and Elena Torou (2015).

International Journal of Information Retrieval Research (pp. 66-91).

www.igi-global.com/article/effectiveness-of-visualization-for-information-retrieval-through-ontologies-with-entity-evolution/130008?camid=4v1a