Dynamic Query Intent Prediction from a Search Log Stream

Wael K. Hanna, Mansoura University, Mansoura, Egypt
Aziza Saad Asem, Mansoura University, Mansoura, Egypt
M. B. Senousy, Sadat Academy for Management Sciences, Cairo, Egypt

ABSTRACT
The users that used search engines are obligated to express their goals in few words (queries). Sometimes search queries are ambiguous. Moreover, the users’ intents are dynamically evolving. This paper analyzes the user’s query logs to classify the related queries, the related intent topic categories and the related intent types and use this classification to dynamically predict the users’ future queries, its intent topic and its intent type. AOL Search Query Log is taken as an experimental data set. Then use evaluation metrics to evaluate the prediction results.

KEYWORDS
Intent Mining, Intent Prediction, Query Logs, Query, Search Engine

INTRODUCTION
Understanding different possible intents of a given input query is a central step for effective search. It has been noticed that the intents of a given query may dynamically evolve, because: 1) New intents may formulate to represent new events that are related to the query. 2) An existing query term may later have a new meaning (e.g. many people today refer to “apple” as a company). 3) Some intents are naturally time-dependent. It noticed that there are two types of intents: those that are constant along the timeline, and those that are bursty. Most current methods on query intent mining try to determine intents in a batch fashion. Hence, they may not be able to capture bursty intents in a timely manner, especially when the bursts happen within a very short time range. This means that applications such as query suggestion and diversified search may not be able to quickly respond to these bursts. [Cong et al. 2013]

Understanding what users are doing in the current search, and predict what they will do in the future search have many possible applications such as query recommendation, web page re-ranking [Hsin et al. 2011]. Caching of query results to enhance retrieval process [Ida, 2012]. And user intent discovery during the web search. In this paper, we dynamically predict users’ future actions (i.e., queries and clicked URLs) based on dynamic mining of users’ query logs. A user submits a sequence of queries associate with URL clicks [Cong et al. 2013]. After each query submission or URL click in a specific time, we predict the queries and intent types and intent topic categories next period of time. Correct predictions can facilitate users’ search processes and ordering of search results.
PROBLEM DEFINITION

This problem is defined formally as follows. Let $s = (a_1, a_2, a_3, \ldots, a_n)$ be a query log stream of a user (search stream of a single user as a document) [Hsin et al. 2011]. [Di et al. 2014]. But the user query intent change dynamically, so let’s divide the query log stream to stream bins $(b_1, b_2, \ldots, b_n)$. Query log stream bins are interactions between user and the search engine in specific times or depend on number of queries like in our experiment for simplicity.

Each stream bin contains several actions where each action $a_i$ $(1 \leq i \leq n)$ is either a query submitted by the user or a URL clicked by the user. The actions $a_1, a_2, \ldots, a_n$ are ordered by the time of their occurrences, with $a_1$ having the earliest occurrence time [Hsin et al., 2011].

Stream bins in query stream can be divided into different pairs $(H_1, F_1), (H_2, F_2), \ldots, (H_{n-1}, F_{n-1})$ where $H_j$ and $F_j$ are two action sequences [Hsin, et al., 2011]. For each stream bin, it is possible to view $H_j$ as a history of the actions that a user has performed during time $t_n$ to $t_{n+1}$ and $F_j$ as the future action that the user will perform during the time $t_n$ to $t_{n+1}$. The goal is to dynamically predict $F_j$ through the stream while given that $H_j$ is known.

RELATED WORK

In Cong et al. (2013) proposed a method for mining dynamic query intents from query logs. By regarding the query logs as a data stream, they determined stable intents while capturing new bursty intents. We add a step for their work to dynamically predict the user intents terms, types and topics through the stream. In addition to analyzing the current events and its relation to the prediction of the user intents.

In Hsin et al. (2011) predicted the next search actions by analyzing the actions in the existent search and search log. In our work we regarding the query logs as a data stream to dynamically predict not only the next actions, but also the user intents terms, types and topics through the stream bins.

In Cristina and Ricardo (2011) presented a study of the feasibility of the query intent prediction. An analysis was made from two aspects: the type of intent of the user’s queries Informational, Not Informational and Ambiguous and 18 topic categories in which the queries within these intents can be placed. But in our work, we regarding the query logs as a data stream to dynamically predict not only the user intents types and topics but also the intent terms through the stream.

Dongug and Seung (2013) identified the bursty intent time of a query from its search log. We can use this method to detect a time where the search intent of a given query changes or evolves in very short times or small numbers of queries.

Chieh-Jen and Hsin-His (2011) determined a system depend on the search-query-log to predict the change of users’ intents. We will detect a time where the search intent of a given query changes or evolves in but very short times or small numbers of queries.

Claudio et al. (2013), determined a two-step technique to detect tasks that users want to do during their search.

Claudio and Giovanni (2012), showed a survey of different approaches to query expansion.

METHOD

Our method takes the search log stream as input, where each log entry is a triple $<t, Q, U>$: $t$ is a timestamp, $Q$ is the search query and $U$ is a set of clicked URL.