Predictive Spatial Search

Mohamed Ali , Abdeltawab M. Hendawi, Ankur Teredesai
Center for Data Science, Institute of Technology, University of Washington, Tacoma, WA, USA
{mhali, hendawi, mdecock, ankurt}@uw.edu

Abstract

In a typical spatial search problem, (mobile) users search for (stationary or mobile) entities that have spatial attributes. The user's current location and/or the entities' locations are considered to assess the relevance of the search result. On one hand, we believe that the user's future location is more relevant to the search result than the current location. Hence, we study spatial search queries under predictive models of user locations. On the other hand and with the ubiquity of hand held devices, most users do not utilize the full power of spatial search and they do not know what to search for. Hence, we introduce a framework to answer the question: "given a user's current and predicted locations, what would the user be interested in searching for and seeing as a query result?" More specifically, we propose a *predictive* spatial search approach that continuously monitors the user's current location to: (1) predict the user's future location and integrate this prediction efficiently in the spatial search query processing pipeline, and (2) predict the search keywords that are of relevance to the user, given the user's location and context. The second type of prediction leverages the knowledge of existing search engines about the behavior of a global set of spatial search users and social media users. Such two-phase prediction capability will enable search engines to "pre-search" on behalf of their users and, thereby, leading to gains in user experience, search accuracy, and communication costs.

Introduction

Commercial search engines recognize the value of tracking the users' location to provide a better service. The typical straightforward approach is to upload the user's location every time a search query is issued. However, the vision presented in [1] proposes a novel approach that utilizes the power of data streaming systems to track the users' locations based on a *route-selection preference policy*. In this approach, the system accepts (voluntarily) the user's declared route preferences along with other context-based information. This information helps the system "*predict*" the user's location even under the absence of the real-time location feed, survive offline periods, reduce communication cost and predict the user's future location with acceptable accuracy. The next set of challenges is to identify the issues that can help integrate spatial search capabilities with such predictive data streaming systems. This integration is the theme of our vision.

Consider the scenario that a user is driving on the highway and searches for a restaurant. It may be desirable and more relevant to direct the user to his favorite restaurant chain (or similar such chains that are still several minutes ahead from his current location) than to simply identify a nearby less favorite restaurant chain. It is even better to predict that this user is, for example, heading to a football game. Consequently, we direct the user to the parade, events and activities next to the stadium that the user is *not* aware of. In this scenario, we notice two main characteristics: first, the search engine is geared up to perform a spatial search around a non-deterministic (yet to be predicted) future location ahead of time. Second, the search engine aims at discovering and recommending personalized search topics on a per user basis. Ranking of the search result

needs to be performed spatiotemporally taking into consideration the probability of the user's expected location along the future timeline.

Currently, users rely on multiple data sources such as live feeds from social media augmented with organic search to discover related events/facilities at the destination. An advantage of the proposed predictive spatial search frameworks is to help consolidate search results better across various faceted search channels. We believe that future geospatial search engine will inevitably adopt the "we know where you will be and *we search before you search*" principle. It is the proactiveness of search engines that will shape the future of spatial search. The search engine is the user's agent that continuously (1) adjusts the search result according to the user's predicted destination and (2) interacts/socializes with neighboring people/agents to highlight interesting topics that the user is totally unaware of.

Predictive Trees: An Index for Predictive Queries on Road Networks

Practical experience tells that it is absolutely a myth to assume that commercial search engines know everything about the user's past locations. Lots of research utilize manufactured databases of historical trajectories and apply machine learning techniques to predict the user's future location. These trajectory databases are usually collected by researchers or volunteers for research purposes. Yet, from a practical perspective, and due to privacy concerns [2], the user location is revealed on a session basis such that each session is no longer than few minutes. Commercial search engines care about users' privacy. Consequently, techniques that assume full knowledge of the user's behavior over extended periods of time are not considered practical in our approach. We propose a new index structure, the predictive tree [3, 4], that enables the evaluation of predictive queries [5] in the absence of the objects' historical trajectories. Based solely on the connectivity of the road network graph and assuming that the object follows the shortest route to destination, the predictive tree determines the reachable nodes of a moving object within a specified time window T in the future. Moreover, predictive trees utilize every additional piece of information and enhance the probability assignment of the predicted location as more trajectory data becomes available on the user.

The predictive tree: (1) provides a generic infrastructure for answering the common types of predictive queries including predictive point, range, KNN, and aggregate queries, (2) updates the probabilistic prediction of the object's future locations dynamically and incrementally as the object moves around on the road network, and (3) provides an extensible mechanism to customize the probability assignments of the object's expected future locations, with the help of user defined functions. In our ongoing effort, we leverage predictive trees to support spatial search and integrate this work with predictive data streaming systems.

References

[1] Mohamed Ali, Badrish Chandramouli, Balan Raman, and Ed Katibah, Spatio-Temporal Stream Processing in Microsoft StreamInsight. IEEE Data Eng. Bull. 33(2): 69-74 (2010).

[2] From GPS and Virtual Globes to Spatial Computing - 2020: The Next Transformative Technology. A Community Whitepaper resulting from the 2012 CCC Spatial Computing 2020 Workshop.

[3] A. M. Hendawi, J. Bao, and M. F. Mokbel. Predictive Tree Source Code and Sample Data. URL:http://www-users.cs.umn.edu/~hendawi/PredictiveTree/, Aug. 2014.

[4] A. M. Hendawi and M. F. Mokbel. Panda: A Predictive Spatio-Temporal Query Processor. In ACM SIGSPATIAL GIS, 2012.

[5] A. M. Hendawi and M. F. Mokbel. Predictive Spatio-Temporal Queries: A Comprehensive Survey and Future Directions. In MobiGIS, California, USA, Nov. 2012.