ABSTRACT

One of the most popular applications of *Location-Based Services* (LBS) is to recommend proximal *Points of Interest* (PoIs) – e.g., nearby restaurants and museums, police stations, hospitals, etc. – or a sequence of PoIs to visit. An important recently addressed variant of the problem not only considers the preference of distance/proximity, but also desires that the returned proximal objects satisfy certain semantic constraints. For instance, rather than picking several close-by attractions with similar features – e.g., restaurants with similar menus; museums with similar art exhibitions – a tourist may be more interested in a result set that could potentially provide more diverse types of experiences, as long as they are within an acceptable distance from a given (current) location. In our research so far, we introduced a topic modeling approach based on the *Latent Dirichlet Allocation* (LDA), a generative statistical model, to effectively model and exploit a fine-grained notion of diversity, namely based on sets of keywords and/or reviews instead of a coarser user-given category. More importantly, for efficiency purpose, we devised two novel indexing structures – *Diversity Map* and *Diversity Aggregated R-tree*. In turn, each of these enabled us to develop efficient algorithms to generate the answer-set for two novel categories of queries. While both queries are focusing on determining the recommended locations among a set of given PoIs that will maximize the semantic diversity within distance limits along a given road network, they each tackle a different variant. The first type of query is k**DRQ**, which finds k such PoIs with respect to a given user's location. The second query k**DPQ** generates a path to be used to visit a sequence of k such locations (i.e., with max diversity), starting at the user's current location, and its extended variant k_T**DPQ** allows user to specify not only starting location but also destination. Moreover, we tackled another query **PaDOC** of jointly determining: (a) a sequence of PoIs that satisfies both traveling budget and the user-desired semantic diversity; (b) recommend the best origin (e.g., hotel) for a given user. Based on a novel indexing structure k-*Closest Category Matrix*, we propose two

efficient algorithms for approximate **PaDOC** query processing based on both conservative and progressive distance estimations. Our experimental evaluations conducted on real datasets demonstrate the benefits of each of the proposed methodologies over different baseline approaches.