Islands of Interest: Mining Concentrations of User Search Intent over E-commerce Product Categories

Neeraj Pradhan, Vinay Deolalikar, Deependra Singh
Groupon Data Mining
Palo Alto, CA 94040

Abstract. E-commerce is burgeoning. Within the last few years, both in the US and the UK, E-commerce has overtaken brick and mortar stores for retail. E-commerce comes with its own set of core problems and challenges: constructing recommendations for users; designing product taxonomies so that inventories can be classified accurately, and in a way that the user finds easy to navigate to; facility allocation for inventory, so that shipping costs are minimized, to name a few. The first of these problems has received much research attention. Accordingly, there is abundant literature on building such recommendations using user's browsing data.
Let us now consider E-commerce *data*. Search as a component of E-commerce shows a steady rise, and has, in many cases, overtaken browse as the primary means of user interaction with E-commerce sites. Accordingly, E-commerce generates more search data than browse data. However, search data is inherently "local" to a query. Therefore, it is not immediately obvious whether it can be used to build "global" (i.e., where no query is involved) knowledge. On the other hand, each of the three problems mentioned earlier requires such global knowledge.
In this paper, we introduce a global structure—that we call *islands of interest*—that is mined from local search data. We show that islands of interest are highly relevant to each of the E-commerce problems mentioned earlier. We introduce two algorithms—one based on community detection, and the other on clustering—that can identify islands of interest. We build a framework that can compare the islands identified using these two approaches, as being efficacious in solving our motivating E-commerce problems.
We believe that in addition to insights into user behavior, islands of interest will provide further impetus to work on hitherto lesser known problems of E-commerce such as design of product taxonomies and facility allocation for inventory.