Using internet search queries to predict human mobility in social events

While our transport systems are generally designed for habitual behavior, the dynamics of large and mega cities systematically push it to its limits. Particularly, transport planning and operations in large events are well known to be a challenge. Not only they imply stress to the system on an irregular basis, their associated mobility behavior is also difficult to predict. Previous studies have shown a strong correlation between number of public transport arrivals with the semi-structured data mined from online announcement websites. However, these models tend to be complex in form and demand substantial information retrieval, extraction and data cleaning work, and so they are difficult to generalize from city to city. In contrast, this paper focuses on enriching previously mined information about special events using automated web search queries. Since this context data comes in unstructured natural language form, we employ supervised topic model to correlate it with real measurements of transport usage. In this way, the proposed approach is more generic and a transit agency can start planning ahead as early as the event is announced on the web. The results show that using information mined from the web search not only shows high prediction accuracy of public transport demand, but also potentially provides interesting insights about popular event categories based on extracted topics.