Data summaries for data exploration in recommender systems

Big data are difficult to process due to its size and complexity; typically, big data does not fit even in highly distributed infrastructures. Ongoing work in this area suggests to provide meaningful summaries so that we use these summaries wherever possible instead of processing the full datasets. On the other hand, recommender systems become more and more important in various applications since they develop the ability to accurately understand and predict users, or systems, wishes, and improve their choices. In that respect, recommender systems so far are exploring the entirety of the (large) datasets, which is becoming more time consuming as the datasets become bigger. On the other hand, users are expecting instant recommendations.

In this work, we propose to use graph-based data summaries to explore and analyze big streaming data, i.e., big data coming with high velocity. This is a highly complex problem because we need to develop algorithms that "understand" the correct semantics of a dataset and provide summaries that preserve these semantics, so query engines can query datasets based on the summaries and provide subsequently correct recommendations without having to process the full dataset. The problem involves work in the area of graph theory, recommender systems and machine learning and exploratory data analytics.