Knowledge based content suggestion on the social web

Billions of users daily use Web-based social networks. On those platforms, people share thoughts, opinions, documents, or events. In the process, large amounts of data are created. An important part of this social data on the Web is accessible. We have the opportunity to use this data to help us expand our knowledge on individual or global questions. Imagine Stevie is part of a social network and fan of jazz pianists, given his social context and preferences, who should be the next person he contacts?

This vast amount of human-generated data has the potential to open the way to many advances in artificial intelligence. However, it is still very difficult to process intelligently and at large-scale. Most state-of-the-art techniques do not take into account the subtlety of the social context or rely on keyword-based information retrieval. There has also been important progress in representing human knowledge in a way that is understandable by machines. We propose to create systems able to access data from the social Web and make intelligent content suggestions given a user, an entity from a knowledge base, or both. It requires us to work on ambitious challenges in many research areas, in particular linked-based ranking, community detection, and recommendation. Moreover, scalability, data acquisition, and data heterogeneity are, in our context, important issues.

There are many exciting research opportunities in integrating social data with knowledge bases. Let us take a concrete example. Maurice, a user, is very much interested in the entity Herbie Hancock. Our aim is to suggest him some more content he would like. We could start looking for users related to him and interested in topics close to Herbie Hancock. This implies having a good user-centric ranking that would combine an idea of social proximity and interestingness relatively to an entity. Then, we want to make relevant recommendations among those users and pick different possibilities from different social or topical context.

Ranking, recommendation, and clustering are fairly mature research areas, but data heterogeneity, data size, the use of knowledge bases and a clever way to combine all these techniques make it possible to innovate on many points and achieve unattained pertinence of results. Given the amount of data we expect to deal with, it is also important to develop scalable systems. To this end, we shall focus on efficient sequential algorithms, distributed algorithms, intelligent indexing, and approximation techniques.