IPAM RIPS 2013

Shoah Foundation Institute Project Description

Introduction
The USC Shoah Foundation Institute for Visual History and Education has nearly 52,000 videotaped interviews with survivors and other witnesses of the Holocaust in their archives. These testimonies have not been transcribed. Instead, the videos are indexed by keywords, personal names, interview codes, and geolocation. To realize the full potential of the archive, we would like to design and develop a new tool that help researchers in their study of these testimonies.

Project objective
The goal of the RIPS summer project is to leverage users' activity to build a recommendation system. Over the past decade, thousands of researchers have used the VHA (subscription-based) and VHAOnline (public) interface to access testimonies. As a result, the interface keeps logs of users' activity, from free-form searches to video views by segments. Nearly 39,000 unique terms have been searched for more than 110,000 times. There are almost 1.2 million video views resulting from these searches. This information provides us a useful bridge to find the connections between searched terms and segments of testimonies. The goal of this project is to utilize this information along with the metadata of each interview to build a recommendation system.

Project description
Given the massive records of user activities, we can build a recommendation system for several applications including the following: Given a search term, can we provide related terms that users may also be interested in? Given a segment of testimony, can we retrieve similar segments that users may also be interested in?

In order to establish recommendations, collaborative filtering systems need to compare fundamentally different objects: items (we refer to search terms and testimonies as items here) versus users. There are two primary approaches to facilitate such a comparison: the neighborhood approach and latent factor models. Neighborhood methods are centered on computing the relationships between items. Latent factor models, such as Singular Value Decomposition (SVD), comprise an alternative approach by transforming both items and users to the same latent factor space, thus making them directly comparable. We will research available literature, brainstorm ideas, and come up with methods, with a final goal of successful implementation of these methods.

References
Su, Xiaoyuan, and Taghi M. Khoshgoftaar. "A survey of collaborative filtering techniques." Advances in Artificial Intelligence 2009 (2009): 4.

Koren, Yehuda, Robert Bell, and Chris Volinsky. "Matrix factorization techniques for recommender systems." Computer 42.8 (2009): 30-37.