

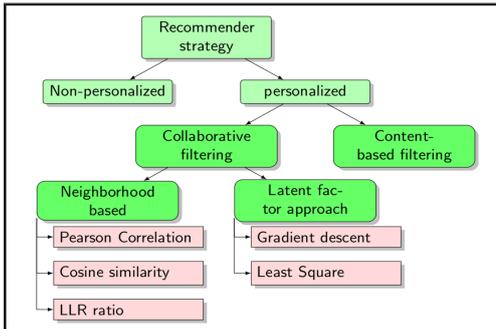


Lukas Hofmaier

Studenten/-innen	Lukas Hofmaier
Dozenten/-innen	Prof. Hansjörg Huser
Co-Betreuer/-innen	Prof. Hansjörg Huser
Topic	Software and Systems

Building and evaluating a multimodal co-occurrence-based recommender engine with Apache Mahout and Apache Solr

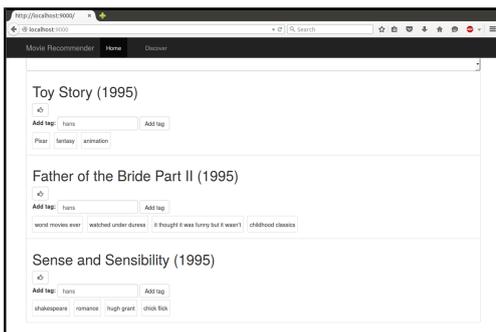
A practical approach to creating recommendations



Overview of some common recommender strategies

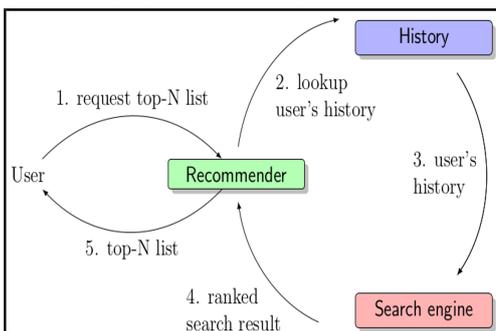
Building a recommendation engine can be a difficult and expensive task. Understanding all possible techniques and choosing the most suitable algorithm for the job at hand presupposes that you have a strong mathematical background. There are a lot of different algorithms and strategies and researchers are constantly coming up with new ones. Not all companies can afford to employ an army of data scientists and mathematicians to build a recommender system. They have to trade-off the accuracy of a recommender against the development costs.

This report describes a simple practical approach to build a recommender system that is more approachable for small development teams. In order to produce a list of recommendable items for a user, it uses the user's history of recent actions to query the search engine Apache Solr for similar items. Similarity among items is based on the log-likelihood ratio of co-occurrence of items with respect to a variety of action types (e.g. purchase, view). This approach exploits existing and proved technologies to save development costs. Furthermore, the system is scalable for big data because Solr is optimized for searching large volumes of data and co-occurrence can be computed at scale with Apache Mahout's spark-item similarity job.



In the demo web application the user can like and tag movies. The recommender will produce a top-N recommendations list based on these actions.

Furthermore, this report explains the core concepts of this practical approach on the basis of a small demo web application. It explains the similarities between the weighting of indicator scores of a memory-based model and the process of ranked retrieval. The evaluation of the described approach shows that the accuracy of the top-N recommendations Task matches the accuracy of a sophisticated item-based algorithm.



The process of producing top-N recommendations list can be divided into five steps.