# Presentation – available topics

Amazon.com recommendations: item-to-item collaborative filtering (2003)

Recommendation algorithms are best known for their use on e-commerce Web sites, where they use input about a customer's interests to generate a list of recommended items. Many applications use only the items that customers purchase and explicitly rate to represent their interests, but they can also use other attributes, including items viewed, demographic data, subject interests, and favorite artists. At Amazon.com, we use recommendation algorithms to personalize the online store for each customer. The store radically changes based on customer interests, showing programming titles to a software engineer and baby toys to a new mother. There are three common approaches to solving the recommendation problem: traditional collaborative filtering, cluster models, and search-based methods. Here, we compare these methods with our algorithm, which we call item-to-item collaborative filtering. Unlike traditional collaborative filtering, our algorithm's online computation scales independently of the number of customers and number of items in the product catalog. Our algorithm produces recommendations in real-time, scales to massive data sets, and generates high quality recommendations.

<http://ieeexplore.ieee.org/abstract/document/1167344/>

# Matrix Factorization Techniques for Recommender Systems (2009)

As the Netflix Prize competition has demonstrated, matrix factorization models are superior to classic nearest neighbor techniques for producing product recommendations, allowing the incorporation of additional information such as implicit feedback, temporal effects, and confidence levels.

<http://ieeexplore.ieee.org/abstract/document/5197422/>

# LARS: A Location-Aware Recommender System (2012)

This paper proposes LARS, a location-aware recommender system that uses location-based ratings to produce recommendations. Traditional recommender systems do not consider spatial properties of users nor items, LARS, on the other hand, supports a taxonomy of three novel classes of location-based ratings, namely, spatial ratings for non-spatial items, non-spatial ratings for spatial items, and spatial ratings for spatial items. LARS exploits user rating locations through user partitioning, a technique that influences recommendations with ratings spatially close to querying users in a manner that maximizes system scalability while not sacrificing recommendation quality.

<http://ieeexplore.ieee.org/abstract/document/6228105/>

# **Linked open data to support content-based recommender systems (2012)**

The World Wide Web is moving from a Web of hyper-linked Documents to a Web of linked Data. Thanks to the Semantic Web spread and to the more recent Linked Open Data (LOD) initiative, a vast amount of RDF data have been published in freely accessible datasets. These datasets are connected with each other to form the so called Linked Open Data cloud. As of today, there are tons of RDF data available in the Web of Data, but only few applications really exploit their potential power. In this paper we show how these data can successfully be used to develop a recommender system (RS) that relies exclusively on the information encoded in the Web of Data. We implemented a content-based RS that leverages the data available within Linked Open Data datasets (in particular DBpedia, Freebase and LinkedMDB) in order to recommend movies to the end users.

<http://dl.acm.org/citation.cfm?id=2362501>

# **Collaborative filtering with temporal dynamics (2009)**

Customer preferences for products are drifting over time. Product perception and popularity are constantly changing as new selection emerges. Similarly, customer inclinations are evolving, leading them to ever redefine their taste. Thus, modeling temporal dynamics is essential for designing recommender systems or general customer preference models. However, this raises unique challenges. Within the ecosystem intersecting multiple products and customers, many different characteristics are shifting simultaneously, while many of them influence each other and often those shifts are delicate and associated with a few data instances. This distinguishes the problem from concept drift explorations, where mostly a single concept is tracked. Classical time-window or instance decay approaches cannot work, as they lose too many signals when discarding data instances. A more sensitive approach is required, which can make better distinctions between transient effects and long-term patterns. We show how to model the time changing behavior throughout the life span of the data. Such a model allows us to exploit the relevant components of all data instances, while discarding only what is modeled as being irrelevant.

<http://dl.acm.org/citation.cfm?id=1721677>

# Similarity of users’ (content-based) preference models for Collaborative filtering in few ratings scenario (2012)

Collaborative filtering is an efficient way to find best objects to recommend. This technique is particularly useful when there is a lot of users that rated a lot of objects. In this paper, we propose a method that improve the Collaborative filtering in situations, where the number of ratings or users is small. The proposed approach is experimentally evaluated on real datasets with very convincing results.

<http://www.sciencedirect.com/science/article/pii/S0957417412002059>

Content-boosted Matrix Factorization Techniques for Recommender Systems (2013)

Many businesses are using recommender systems for marketing outreach. Recommendation algorithms can be either based on content or driven by collaborative filtering. We study different ways to incorporate content information directly into the matrix factorization approach of collaborative filtering. These content-boosted matrix factorization algorithms not only improve recommendation accuracy, but also provide useful insights about the contents, as well as make recommendations more easily interpretable.

<https://arxiv.org/pdf/1210.5631.pdf>

Exploiting Query Reformulations for Web Search Result Diversification (2010)

When a Web user’s underlying information need is not clearly specified from the initial query, an effective approach is to diversify the results retrieved for this query. In this paper, we introduce a novel probabilistic framework for Web search result diversification, which explicitly accounts for the various aspects associated to an underspecified query. In particular, we diversify a document ranking by estimating how well a given document satisfies each uncovered aspect and the extent to which different aspects are satisfied by the ranking as a whole.

<https://pdfs.semanticscholar.org/ded1/fa5432e066c3a9f998bd52c6ed736068c582.pdf>

# **Beyond clicks: dwell time for personalization (2014)**

Many internet companies, such as Yahoo, Facebook, Google and Twitter, rely on content recommendation systems to deliver the most relevant content items to individual users through personalization. Delivering such personalized user experiences is believed to increase the long term engagement of users. While there has been a lot of progress in designing effective personalized recommender systems, by exploiting user interests and historical interaction data through implicit (item click) or explicit (item rating) feedback, directly optimizing for users' satisfaction with the system remains challenging. In this paper, we explore the idea of using item-level dwell time as a proxy to quantify how likely a content item is relevant to a particular user. We describe a novel method to compute accurate dwell time based on client-side and server-side logging and demonstrate how to normalize dwell time across different devices and contexts.

<http://dl.acm.org/citation.cfm?doid=2645710.2645724>