**Student:** Oleksandr Romanchenko
**ID:** 83459

**Essay on topic:** **"The art and science of recommendation algorithms. Example of the world's biggest retail company: Amazon".**

Recommendation algorithms belong to the field of affinity analysis. Definition: **Affinity analysis** is a data analysis and data mining technique that discovers co-occurrence relationships among activities performed by (or recorded about) specific individuals or groups. In general, this can be applied to any process where agents can be uniquely identified and information about their activities can be recorded. In retail, affinity analysis is used to perform **market basket analysis**, in which retailers seek to understand the purchase behavior of customers. This information can then be used for purposes of cross-selling and up-selling, in addition to influencing sales promotions, loyalty programs, store design, and discount plans.

To be able to say more about recommendation algorithms let's have a look at such example of greatest world leaders as… Amazon!



From Amazon's article about their algorithm:

*"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
- search-based methods.

I'm going to briefly go through these methods and Amazon's algorithm, which they call item-to-item collaborative filtering. Unlike traditional collaborative filtering, their algorithm's online computation scales independently of the number of customers and number of items in the product catalog. The algorithm produces recommendations in realtime, scales to massive data sets, and generates highquality recommendations.

Most recommendation algorithms start by finding a set of customers whose purchased and rated items overlap the user's purchased and rated items. The algorithm aggregates items from these similar customers, eliminates items the user has already purchased or rated, and recommends the remaining items to the user. Two popular versions of these algorithms are collaborative filtering and cluster models. Other algorithms — including search-based methods and Amazon's item-to-item collaborative filtering — focus on finding similar items, not similar customers. For each of the user's purchased and rated items, the algorithm attempts to find similar items. It then aggregates the similar items and recommends them.

Traditional Collaborative Filtering
A traditional collaborative filtering algorithm represents a customer as an N-dimensional vector of items, where N is the number of distinct catalog items. **The components of the vector are positive for purchased or positively rated items and negative for negatively rated items.** It can measure the similarity of two customers, A and B, in various ways; a common method is to measure the cosine of the angle between the two vectors:

$$similarity(\vec{A}, \vec{B}) = \cos(\vec{A}, \vec{B}) = \frac{\vec{A} * \vec{B}}{|\vec{A}| * |\vec{B}|}$$

The algorithm can select recommendations from the similar customers' items using various methods as well, a common technique is to rank each item according to how many similar customers purchased it.

## Cluster Models

To find customers who are similar to the user, cluster models divide the customer base into many segments and treat the task as a classification problem. The algorithm's goal is to assign the user to the segment containing the most similar customers. It then uses the purchases and ratings of the customers in the segment to generate recommendations. The segments typically are created using a clustering or other unsupervised learning algorithm, although some applications use manually determined segments. Using a similarity metric, a clustering algorithm groups the most similar customers together to form clusters or segments.

## Search-Based Methods

Search-or content-based methods treat the recommendations problem as a search for related items. Given the user's purchased and rated items, the algorithm constructs a search query to find other popular items by the same author, artist, or director, or with similar keywords or subjects. If a customer buys the Godfather DVD Collection, for example, the system might recommend other crime drama titles, other titles starring Marlon Brando, or other movies directed by Francis Ford Coppola.

## Item-to-Item Collaborative Filtering

Amazon.com uses recommendations as a targeted marketing tool in many email campaigns and on most of its Web sites' pages, including the hightraffic Amazon.com homepage. Clicking on the "Your Recommendations" link leads customers to an area where they can filter their recommendations by product line and subject area, rate the recommended products, rate their previous purchases, and see why items are recommended (see Figure 1). As Figure 2 shows, Amazon's shopping cart recommendations, which offer customers product suggestions based on the items in their shopping cart. The feature is similar to the impulse items in a supermarket checkout line, but impulse items are targeted to each customer. Amazon.com extensively uses recommendation algorithms to personalize its Web site to each customer's interests. Because existing recommendation algorithms cannot scale to Amazon.com's tens of millions of customers and products, they developed their own. Their algorithm, item-to-item collaborative filtering, scales to massive data sets and produces high-quality recommendations in real time.
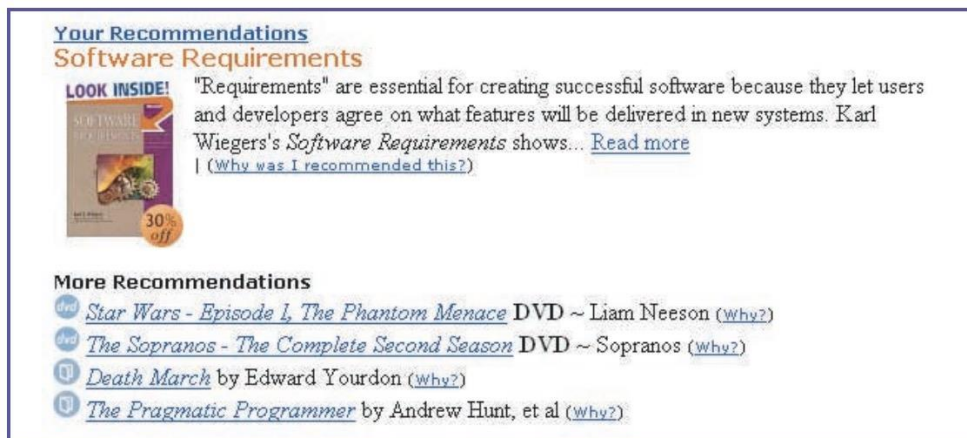
*Figure 1. The "Your Recommendations" feature on the Amazon.com homepage. Using this feature, customers can sort recommendations and add their own product ratings.*



*Figure 2. Amazon.com shopping cart recommendations. The recommendations are based on the items in the customer's cart:* The Pragmatic Programmer *and* Physics for Game Developers.

## Scalability

A Comparison Amazon.com has more than 29 million customers and several million catalog items. Other major retailers have comparably large data sources. While all this data offers opportunity, it's also a curse, breaking the backs of algorithms designed for data sets three orders of magnitude smaller. Almost all existing algorithms were evaluated over small data sets. For example, the MovieLens data set4 contains 35,000 customers and 3,000 items, and the EachMovie data set3 contains 4,000 customers and 1,600 items. For very large data sets, a scalable recommendation algorithm must perform the most expensive calculations offline. As a brief comparison shows, existing methods fall short:

- Traditional collaborative filtering does little or no offline computation, and its online computation scales with the number of customers and catalog items. The algorithm is impractical on large data sets, unless it

uses dimensionality reduction, sampling, or partitioning — all of which reduce recommendation quality.

- <u>Cluster models</u> can perform much of the computation offline, but recommendation quality is relatively poor. To improve it, it's possible to increase the number of segments, but this makes the online user–segment classification expensive.

- <u>Search-based models</u> build keyword, category, and author indexes offline, but fail to provide recommendations with interesting, targeted titles. They also scale poorly for customers with numerous purchases and ratings.

The key to item-to-item collaborative filtering's scalability and performance is that it creates the expensive similar-items table offline. The algorithm's online component — looking up similar items for the user's purchases and ratings — scales independently of the catalog size or the total number of customers; it is dependent only on how many titles the user has purchased or rated. Thus, the algorithm is fast even for extremely large data sets. Because the algorithm recommends highly correlated similar items, recommendation quality is excellent. Unlike traditional collaborative filtering, the algorithm also performs well with limited user data, producing high-quality recommendations based on as few as two or three items.

**Conclusion**
Recommendation algorithms provide an effective form of targeted marketing by creating a personalized shopping experience for each customer. For large retailers like Amazon.com, a good recommendation algorithm is scalable over very large customer bases and product catalogs, requires only subsecond processing time to generate online recommendations, is able to react immediately to changes in a user's data, and makes compelling recommendations for all users regardless of the number of purchases and ratings. Unlike other algorithms, item-to-item collaborative filtering is able to meet this challenge. In the future, we expect the retail industry to more broadly apply recommendation algorithms for targeted marketing, both online and offline.

**References**

1. J.B. Schafer, J.A. Konstan, and J. Reidl, "E-Commerce Recommendation Applications".

2. K. Goldberg et al., "Eigentaste: A Constant Time Collaborative Filtering Algorithm".

3. B.M. Sarwar et al., "Item-Based Collaborative Filtering Recommendation Algorithms"

4. B.M. Sarwar et al., "Analysis of Recommendation Algorithms for E-Commerce".