Mark Shinozaki
Cpts 315
Homework #2
Analytical Part

Q1. Consider the following ratings matrix with three users and six items. Ratings are on a 1-5 star scale. Compute the following from data of this matrix: (20 points)

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	4	5		5	1	
User 2		3	4	3	1	2
User 3	2		1	3		4

Table 1: Data of ratings from three users for six items.

a) Treat missing values as 0. Compute the jaccard similarity between each pair of users.

b) Treat missing values as 0. Compute the cosine similarity between each pair of users.

c) Normalize the matrix by subtracting from each non-zero rating, the average value for its user. Show the normalized matrix.

d) Compute the (centered) cosine similarity between each pair of users using the above normalized matrix.

a)

HW #2

a) Jaccard similarity Index: consider two sets x and y.

User 1 = {4, 5, 0, 5, 1, 0} = {0, 1, 4, 5}

User 2 = {0, 3, 4, 3, 1, 2} = {0, 1, 2, 3, 4}

User 3 = {2, 0, 1, 3, 0, 4} = {0, 4, 3, 1, 2}

• Jaccard Similarity between user 1 and user 2

$$J(\text{user } 1, \text{user } 2) = \frac{|\text{user1} \cap \text{user2}|}{|\text{user1} \cup \text{user2}|} \times 100$$

$\frac{3}{6} \times 100 = 50\%$ J user 1, user 2 = 50% or .5

$J(\text{user } 1, \text{user } 3) = \frac{2}{6}$

$J(\text{user } 2, \text{user } 3) = \frac{3}{6}$

b) Cosine Similarity between user x and y $= \dfrac{r_x \cdot r_y}{|r_x|\,|r_y|}$

User 1 = {4, 5, 0, 5, 1, 0}

User 2 = {0, 3, 4, 3, 1, 2}

User 1, User 2 = 60.6% or .61

User 1, User 3 = .51

User 2, User 3 = .61

$\|\text{user }1\| = \sqrt{4^2+5^2+0^2+5^2+1^2+0^2}$

$= \sqrt{16+25+0+25+1+0}$

$= \sqrt{67}$

$\|\text{user }2\| = \sqrt{0^2+3^2+4^2+3^2+1^2+2^2}$

$= \sqrt{0+9+16+9+1+4} = \sqrt{39}$

Mark Shinozaki
Cpts 315
Homework #2

c) Normalized Matrix

	item 1	item 2	item 3	item 4	item 5	item 6
User 1	.25	1.25	0	1.25	-2.75	0
User 2	0	0.4	1.4	.4	-1.6	-0.6
User 3	-0.5	0	-1.5	.5	0	1.5

d)

Sim(User 1, User 2) = .72

Sim(User 1, User 3) = 0.07

Sim(User 2, User 3) = -0.55

Q2. Please read the following two papers and write a brief summary of the main pointers in at most TWO pages.

Two Decades of Recommender systems at Amazon.com

- The first part of the article, it discusses how Amazon.com has been personalizing the shopping experience for each customer using item-based collaborative filtering since 1998. The algorithm recommends items based on a customer's past behavior and current context, and has been widely adopted by other websites. The success of the algorithm lies in its simplicity, scalability, and ability to provide useful recommendations while being easily understandable. The article also highlights the challenges and improvements in collaborative filtering over the years, and discusses the future of recommender systems and personalization. This next part of the article explains how the item-based collaborative filtering algorithm works by finding related items for each item in the catalog based on purchase patterns. This approach has advantages over user-based collaborative filtering, as most of the computation is done offline, and recommendations can be generated in real-time through lookups. The algorithm is scalable and updates immediately on new information. The article also provides examples of other companies, such as YouTube and Netflix, that have successfully used recommender systems to improve the user experience and increase engagement. In the section, defining "related items", this part discusses the challenges of creating effective recommendation systems, which rely on identifying useful patterns in customer behavior. The article explains a method for estimating the expected number of customers who have brought both X and Y, and for evaluating whether the observed number of customers who brought both items is higher or lower than expected. The article also discusses the importance of balancing popularity with the power law distribution of unpopular items when creating relatedness scores. The article concludes by discussing how compatibility and the meaning of related items can emerge from people's behavior, and how machine learning and controlled online experimentation can be used to optimize recommendation systems based on customer preferences. The next part of this article is, The importance of Time, This section explains the importance of time in making these recommendations for Amazon.com customers. It also explains, how the timing and sequence of purchases can affect the strength of correlations between items, and how cold-start problems can arise for both new customers and items. The article also highlighted the importance of diversity in recommendations, as well as the need to balance immediate intent with long-term optimization. Finally, it discusses the challenges of modeling customer interests as they change over time, and the need to develop techniques for learning which purchases lead to useful recommendations. In the next section, or The Future: Recommendations Everywhere, in this section, the focus is on the future of recommendations and the personalization in technology. The author envisions a future where recommendations will be seamlessly integrated into our lives, based on a deep understanding of our preferences and needs. They propose a new way of thinking about recommendations, where every interaction should reflect who we are and what we like. The author believes that the future of recommendations will be built on intelligent computer algorithms leveraging collective human intelligence, resulting in more personalized, relevant, and engaging experiences for everyone. Then in the final conclusion paragraph, this discusses the history of recommendation algorithms, specifically item-based collaborative filtering, which remains popular due to its simplicity, scalability, and adaptability. However, the author notes that there is still a lot of opportunities to add intelligence and personalization to every part of every system, creating experiences that feel like a friend who understands and anticipates your needs. The author emphasizes that recommendations are about discovery and should offer surprise and delight by uncovering new options for the user, and envisions a future where every interaction is a recommendation.

Amazon.com Recommendations, Item to Item Collaborative Filtering

In the first part of this article until the Cluster Models part, the article discusses, recommendation algorithms and their use in e-commerce websites. These algorithms use inputs about a customers interest, including purchased and rated items, items viewed, demographic data, subject interests, and favorite artists, to generate a list of recommended items. Recommendation algorithms personalize online stores for each customer and have higher click-through and conversation rates than untargeted content such as banner advertising and top-seller lists. However, e-commerce recommendation algorithms operate In a challenging environment due to large amounts of data, real-time requirements, and limited customer information. The three common approaches to solving the recommendation problem are traditional collaborative filtering, cluster models, and search-based methods. The article compares these methods with an algorithm called item-to-item collaborative filtering. Traditional collaborative filtering represents a customer as an N-dimensional vector of items, where N is the number of distinct catalog items, and generates recommendations based on a few customers who are most similar to the user. However, using collaborative filtering to generate recommendations is computationally expensive, making it difficult to scale large data sets. The article concludes by suggesting partial solutions to reduce data size, including random sampling, discarding customers with few purchases, and dimensionality reduction techniques such as clustering and principal component analysis. Then in the next part about cluster models, this part of the article can be summarized to be, the use of cluster models in recommendation systems to find similar customers and generate recommendations. Cluster models divide the customer base into segments and assign users to the segment containing the most similar customers. The segments are created using clustering or unsupervised learning algorithms, and similarity metics are used to match users to segments. Cluster models have better online scalability than collaborative filtering, but the recommendation quality is lower because the similar customers found are not always the most similar. The recommendations produced by cluster models are often too general or too narrow and fail to help customers discover new, relevant, and interesting items. Then in the next section, Search-Based methods, its about, Search-based recommendation methods that treat the recommendation problem as a search for related items by constructing a query based on the users purchased and rated items. They recommend items with similar keywords, subjects, or attributes to the user's purchased and rated items. These algorithms perform well for users with few purchases or ratings, but for users with a large amount of data, it is impractical to base a query on all items, and recommendation quality is relatively poor. The recommendations tend to be either too general or too narrow, which does not help the user discover new, relevant, and interesting items. The next few parts of the article go on to discuss, Item-to-item collaborative filtering and how that works, this part of the article discusses, how Amazon user recommendations to personalize their website to each customer's interests using their own algorithm called item-to-item collabortive filtering, which matches each of the user's purchased and rated items to similar items and combines them into a recommendation list. The algorithm builds a similar-items table by finding items that customers tend to purchase together, and then finds items similar to each of the user's purchases and ratings, aggregates those items, and recommends the most popular or correlated items. This algorithm scales to massive data sets and produces high-quality recommendations in real-time. The next part is the section entitled scalability, this part of the article discusses the challenge of scalability in recommendation algorithms for large datasets like Amazon's, with traditional methods failing to preform offline computation or providing poor quality recommendations. Item-to-item collaborative filtering is a scalable solution that creates the expensive similar-items table offline and provides excellent recommendation quality based on highly correlated similar items. The algorithm's online component scales independently of catalog size or total number of customers and preforms well even with limited user data. In the conclusion part of this article, it says, the recommendation algorithms are effective in providing personalized shopping experiences for customers, and a good recommendation algorithm for large retailers should be scalable, require sub second processing time, react immediately to changes in user data, and provide compelling recommendations for all users regardless of the number of purchases and ratings. The item-to-item collaborative filtering algorithm is said to be able to meet this challenge. This article also expects the retail industry to more broadly apply recommendation algorithms for targeted marketing both online and offline in the future.