Wanyea Barbel
2 December 2022

Final Project Report

In choosing my final project for this I considered what I thought to be most relevant to my experiences. I have just concluded Senior Design II and I chose to build an itinerary planning application that takes in user sentiments and recommends places accordingly. However, this application does not take into account places has previously been and I believe including this functionality will make the application feel more complete and give users more accurate recommendations based on who they are and the places they like to visit. With this in mind, I thought it would be interesting to investigate how Amazon chooses to recommend items to users based on some of the items they have already browsed. I believe by understanding this system, it may give me some insight into my own project and I will be able to provide a better user experience.

I chose to use the Stanford Large Network Dataset Collection for my dataset. Specifically, I used dataset *amazon0601* which includes the Amazon product co-purchasing network from June 1st 2003. I chose this dataset because it was relatively large with 403,394 vertices and 3,387,388 edges which would provide me enough sample points to predict what a user might pick. This sample set is a directed unweighted graph that includes two columns: a *from vertex* and a *to vertex*.

Amazon uses a powerful *collaborative filtering algorithm* that allows the company to make personalized recommendations to its customers. By analyzing the purchasing habits and preferences of its users, the algorithm is able to identify patterns and make predictions about what products a particular customer is likely to be interested in. The

algorithm works by taking into account the products that a customer has previously purchased, as well as the products that other customers who have similar purchasing habits have purchased. It then uses this information to create a profile for the customer, which is used to make personalized recommendations. One of the key advantages of Amazon's collaborative filtering algorithm is its ability to make accurate predictions. By analyzing the vast amount of data that is available to the company, the algorithm is able to identify patterns and trends that would be impossible for a human to detect. This allows Amazon to make recommendations that are highly relevant to each individual customer, increasing the chances that they will make a purchase. Another advantage of the algorithm is its ability to constantly evolve and improve. As more data is collected and analyzed, the algorithm is able to identify new patterns and trends, allowing it to make even more accurate predictions. This is important because it allows Amazon to make relevant recommendations for its customers that are always improving, making it easier for them to find products that they are more willing to purchase. Overall, Amazon's collaborative filtering algorithm is a powerful tool that allows the company to make personalized recommendations to its customers. Although this is not an artificial intelligence or machine learning course, graph problems are often found in these disciplines and Amazon’s use of collaborative filtering algorithms shows this. By using data analysis and machine learning, the algorithm is able to identify patterns and make accurate predictions, improving the customer experience and increasing the chances of customers purchasing an item they see.

Wanyea Barbel
2 December 2022

To implement Amazon's collaborative algorithm (assuming we're coding in Java), we would need to create a class that can collect and analyze data about a customer's purchasing habits and preferences. This class would need to have methods for storing data about a customer's purchases, as well as methods for analyzing this data to make predictions about what products the customer is likely to be interested in. We would also need to create a data structure to store the data about a customer's purchases, as well as the data about the purchases made by other customers with similar purchasing habits. Once the data has been collected and stored, we can implement the algorithms that are used to make predictions about a customer's interests. This would involve using machine learning techniques to identify patterns and trends in the data, and using this information to make personalized recommendations to users. Like I mentioned earlier, I am able to obtain some of this customer information from the Stanford Large Network Dataset Collection. However, the nature of this dataset is primitive and only provides a graph that shows a relationship between items and does not include metadata about each specific item which would be used in a complete implementation of the collaborative algorithm.

Ultimately, I decided that the nature of this course did not call for a machine learning model and instead I decided on a variation of the original algorithm. Instead of passing the dataset to train a model, I evaluate a *prediction value* based on the average of the neighboring vertices in the graph. These neighbors are found by providing the vertex in question and iterating through each edge to find directly connected vertices. The average is found by taking the summation of the data for every vertex in the graph. The

prediction value is found by taking the vertex chosen by the user (the item they would have viewed on Amazon), adding it by the average value and dividing by two. Each vertex has a prediction value associated with it accordinging to the initial matrix chosen. Once the algorithm is finished running I grab the top five nodes with the highest prediction value and return these items to the user.

Overall, I believe this project was a great chance to investigate an interesting algorithm and its relation to network optimization. In the future, I will definitely be implementing a variation of this algorithm within my own itinerary planning application so users can visit places that are more relevant to their interestings. Alternatively, this graph algorithm is also useful for recommending friends and places users wouldn't normally visit to get them out of their comfort zone. Although I did not implement the full capabilities of this algorithm (no machine learning model), I still believe the algorithm I implemented is an interesting way to provide users with useful recommendations based on their personal interests.