

ITCS 6162 Programming Assignment – Pixie-Inspired Algorithm Explanation

Emory Soper

A Pixie-inspired recommendation system uses a graph-based approach to find recommendations for users. To develop this graph, the data for users and items is merged together. This data will show which users have rated or interacted with which items. From this data, a graph can be formed where nodes represent users and items, and the edges represent the connections between the users and items. If a user has rated or interacted with an item, an edge will connect their representative nodes on the graph. Since both users and items will only have connections to the opposite type, this will form a bipartite graph with users and items being the two groups.

With this bipartite structure, two users connected to the same item or two items connected to the same user will be more closely related. And two users related in this way will henceforth have their other connected items be more closely related to the other user. This means that random walks starting at a specific will be more likely to hit items that are closely related to the starting user or item. This property means that the more times a node is visited on a random walk, the more likely it is to be closely related to the starting node. A pixie-inspired recommendation system will start a user or item node to make recommendations from, then take large random walks while counting the number of visits made to each item node. Due to the connected structure of the graph, the item nodes with the most visits should be ones that are most connected to the starting node. These item nodes will make good recommendations because they are based on the connections and communities formed by the graph.

The Pixie recommendation system was initially developed for Pinterest. They were faced with a massive pool of items and large list of users. They wanted their recommendation system to handle a massive amount of data and many empty connections while being fast and scalable. By employing this graph-based approach, they were able to make a system that can generate recommendations quickly while constantly updating the connections in the graph to account for new interactions.