Abstract
In this project we aim to detect user communities in a news stream. We define a user community as a set of news readers with similar interest in the news documents. The similarity in the interest of users can be due to similar news categories, entities, subjects, etc. We will use MIND[1] news recommendation dataset in this work. The MIND dataset is a collection of news articles, users as news readers, and users’ click log on news articles. More specifically, our aim is to detect user communities by using the click logs of the users.

We define the problem of online user community detection as follows. First, we use Word2Vec embedding to embed the news articles in the dataset. Then, the users are represented by the mean embedding vector of the news articles they clicked on. We use a dimension reduction algorithm like PCA to reduce the dimensionality of the user representation vectors. Followed by that, Voronoi diagram in 2D will be used to visualize and detect communities.

References