Ravi Rajan

Behavioral clustering of Users from Twitter Data

The goal of this demonstration is to illustrate that users / items could be clustered together purely from a data driven approach and could resemble reality. One possible application of this would be in targeted advertising.
Matrix factorization techniques were used to demonstrate the association of items with each other, and this is based on their interaction with users in the sample population.
In the following graph, the blue dots are users while the red dots are the items and are labeled appropriately.

It is interesting to see how
- Retailers (walmart, target, costco, sears etc) cluster to the top right part of the graph
- Electronics (Tablets and Phones such as nexus, nook, iPad, kindle, surface, iPhone etc) to the far left
- Entertainment (Popular Games and consoles such as xbox, playstation, halo, call o duty etc) to the bottom right

Looking deeper into one of the clusters, for example the Retailers, it can be claimed that the population (users) in the vicinity of this cluster is "similar" and can be targeted as a whole. This claim holds true because walmart, costco, sears etc are indeed Retailers and have clustered together. This establishes the validity for using unsupervised machine learning on large scale Twitter data.

A couple of noteworthy observations
- Gamestop (to the far right), which falls somewhere in between Retailers and Entertainment.
- Applestore is more closely related to Tablets and Phones than other Retailers; which could arguably be the true perception

Cols