Partner Report for Spotify - Milestone 1

Spandan Madan, Mehul Smriti Raje
Timothy Lee, Benjamin Sanchez-Lengeling

Summary of points discussed in Skype meeting

- Spandan gave a demo of his word2vec model, and asked Aparna and Joe if we could get a sense of how spotify's existing systems work so that we can build on top of it, and not duplicate any previous failed experiments. As it is not possible divulge any information about spotify's existing systems, Aparna and Joe mentioned that we should approach this project independently of Spotify's existing recommendation systems.
- Aparna and Joe mentioned that it would be interesting to see how incorporating different types input information and data can be utilized, as opposed to focusing our energy on only different kinds of models. That is, focusing on data modalities like lyrics and youtube comments would be a promising direction. Another important source of auxiliary information mentioned was the order of songs in a playlist.
- Aparna mentioned that designing evaluation metrics will be important as the problem is open ended. Pavlos said the same in milestone discussion.
- Joe agreed that Spandan's idea of splitting the playlist continuation task into two parts - 1) Finding an approximate neighborhood i.e. a pool of interesting songs, and 2) Ranking songs in that pool only, was promising.
- Ben suggested Reinforcement Learning approaches. Joe said that is a interesting venue to look at but the metric is going to be hard.

Summary of status quo and next steps

As discussed in the meeting, we plan to focus on the problem of Automatic Playlist Continuation, and designing evaluation metrics for the same. We plan to do so using a 2 step process - Finding a pool of relevant songs, and then ranking the songs in this pool.

We have trained a model using the same algorithm as word2vec (CBOW) for the pooling step. Initial testing shows that the results of this model are sensible, and it is able to identify a bunch of artists/songs that are similar to the seed songs in a playlist. Two auto generated results for the same can be seen in these automatically generated playlists Playlist 1 and Playlist 2 (first song was seed).

Now, we are focusing our efforts on 2 directions - 1) Building a hierarchical model inspired by Microsoft's Groove Radio and 2) scraping relevant information in the form of lyrics and youtube comments. The 1^{st} step will give us a sensible base model that solves the task of Automatic Playlist Continuation and serves as a playground to experiment with new input features. Next steps include incorporating new features in the form of lyrics and comments into this ranking model. We plan to discuss these next steps in our next Skype meeting.