Project Proposal

Movie Recommendation System

Problem Identification and Context:

Recommendation System is ubiquitous and is growing in popularity in the world of predictive modelling and Machine Learning.

The recommendation system is because of information overload, and can be called as an information filter system. It greatly influences with what we interact with the world: shopping (Amazon, Best Buy), music (Spotify), videos (YouTube, Netflix), and more.

To build a movie recommendation system is to provide recommendations to number of users with numerous movies. Recommendation is done based on similarity between users (Collaborative Filtering) or by considering particular user's activity (Content Based Filtering) which he wants to engage with.

Criteria for Success:

Solid updated data covering user behavior, product (movie) information, and the reviews that link them both is critical.

Selecting efficient features as well implementation of advanced machine learning tools and algorithms are also instrumental.

Constraints:

There are various challenges faced by Recommendation System. These challenges are Cold Start problem, Data Sparsity, Scalability.

Cold Start Problem: It needs enough users in the system to find a match. For instance, if we want to find similar user or similar item, we match them with the set of available users or items. At initial stage for a new user, their profile is empty as they have not rated any item and the system do not know about their preferences, so it becomes difficult for a system to provide them recommendation about any item. Same case can be with new item, as it is not rated by any user because it's new for the user. Both these problems can be resolved by implementing hybrid techniques.

Data Sparsity: The user or rating matrix is very sparse. It is very hard to find users that have rated the same items because most of the user do not rate the items. So, it becomes hard to find set of users who rate the items. To give recommendation is really tough when there is less information about any user.

Scalability: Collaborative Filtering use massive amount of data to make it reliable and better which require a greater number of resources. As information grows exponentially processing becomes expensive and inaccurate result from this Big data challenge.

Scope of solution space:

The scope of any recommender system is to build a model in such a way that their user gets proper recommendation and efficiency of the system is maintained.

Movie recommendation system scope is not limited to entertainment, but also in information sharing.

Incorporating time into a recommender system is important, because there are often preference seasonal effects. For example, it is likely that in December, more people are going to be watching family based and holiday-themed movies.

Stakeholders:

- Owners and shareholders/investors
- Employees
- Customers (Subscribers) and viewers
- Independent content producers
- Competitors

Data Source:

The data was obtained from MovieLens, an online recommendation service. The dataset has two sub-datasets, first one is the ratings dataset which consists of around 25 million user ratings and the second one is the movies dataset which provides information about the movie name, genre and the year of release.