Analyzing interest transition in movies genre over different age-group

Syed Asad Shah 22001127

Hamza Islam 22101350

Scientific studies show how human mind learn, mature, and adapt from infancy to adulthood to elderly phases of life. While every person is a little bit different, human development tends to follow a remarkably predictable pattern. This development can be realized in his choices and behavior. In this project we are going to analyze human interest development in movies genre over the passage of time. Common assumption is people with the same age-group tends to have the same behavior, feelings, and interests. We will analyze this assumption based on movies genre watch by the group of people of same age. We will try to find co-relation between the age group and the genre of movie, which might help in explaining behavior and thoughts felt by that age group. We will try to partition the data on other demographics like gender (male and female) and country (US and non-US), perform analysis on these partitions of data, compare the results and try to relate these results with age-group behavior based on region.

We have obtained two IMDB datasets [1] from Kaggle i.e., movies.csv and rating.csv. Collectively, we will be analyzing around 170k+ rows of data. Firstly, the data would be needed to be in the right format for which we'll use python, then use that in some tool. The generated network would be 2-mode as there would two kinds of nodes. We will identify nodes as users and movies and then create link between them using the ratings given by users to movies where movie ratings value would be used as edge-weight. We will create clusters of movies which would be labeled as the genre of the movie and the other nodes would be clustered as the age groups of the users that viewed those movies. And then we would be looking at the relations between those clusters. The dataset obtained has the age groups defined as 0-18, 18-30, 30-45 and 45+ whereas movie genres are drama, comedy, action, adventure, fiction, romantic etc. and their combination as group.

We will perform one of the clustering algorithms such as Girvan Newman or Markov cluster, to cluster the nodes by their attributes i.e., movies genre, age group etc. We will also analyze both the network and node-based centralities such as degree distribution analysis, centrality analysis, page rank analysis, density analysis and path analysis etc. We can perform these analysis and others more that we might learn as the course progresses using timestamps also. This can help us find the change in relations or likeness or dislikeness over a course of time. We can visualize the trend over the period using Gephi or ORA. This analysis can also help us predict the popularity of an upcoming movie of a certain genre in different groups of people, which eventually will make filmmakers take better decisions when targeting the movie for a particular group or it can help marketing team to better focus on the right targeted audience for the promotion of a movie.

References

[1] Stefano Leone. IMDb movies extensive dataset. https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset. Accessed: 2021-10-25

[2] http://dept.clcillinois.edu/psy/LifespanDevelopment.pdf

[3] Kubrak T. (2020). Impact of Films: Changes in Young People's Attitudes after Watching a Movie. *Behavioral sciences (Basel, Switzerland)*, *10*(5), 86. https://doi.org/10.3390/bs10050086

[4] Fearing, F. (1947). Influence of the Movies on Attitudes and Behavior. *The Annals of the American Academy of Political and Social Science*, *254*, 70–79. http://www.jstor.org/stable/1026142