Movie Recommendation System

Dev Moxaj Desai

Data Science, Recommendation System, KNN Algorithm, Python, Clustering, Numpy, Movies

---

Team Member Name	Roll Number	Email-Id
Advait Padhye	200100014	advaitpadhye1@gmail.com
Sameep Chattopadhyay	20D070067	sameep.ch.2002@gmail.com

Brief Description

A Recommendation System is a filtration program whose prime goal is to predict the “rating” or “preference” of a user towards a domain-specific item. In our project, the domain-specific item is movie. Therefore the main focus of our recommendation system is to filter and predict only those movies which a user would like to watch. For this project, we are using Content based Collaborative filtering . So basically, the user feeds the name of a movie that he likes (Eg. Godfather) and the recommendation system would suggest 10 other movies that are similar to the given movie which the user may also like. This recommendation system makes use of the KNN algorithm and we have used an open source database named MovieLens-small for this project.

Progress

In the months of May & June, we went through all the reading materials provided to us by our mentor, this helped us learn about the general ML concepts such as Regression, Classification, KNN, PCA etc. We also went through various assignments and solved them. Once all the learning part was done, our mentor shared with us a list of recommended topics on which we could do our final project.
Our endsems got over at 25th July, and after that we went over a few of the topics, After going through the list multiple time,we finally zeroed in on this project about Movie recommendation system. We both are huge fans of movies and so decided to do this project as a part of our SoC. After selecting a topic,we had a meeting with our mentor and he gave us an idea about the project and how to proceed on that. It was his suggestion that we make use of the KNN algorithm for this. Then, over the next few days we both did research online about how to implement this project. We came across various articles and YouTube videos and learnt a lot from them. We also obtained an open source database that contained thousands of movies and their reviews.
Once we had what we needed, we were able to build the final project without any hitch. We spent some time after that to make the code and output look aesthetically soothing. Finally once we completed making it, we shared the project with our mentor , and he approved it with some minor changes. We incorporated the changes immediately. Lo and Behold, our project was done.

Results

Output :-

```
In [15]: get_movie_recommendation('godfather')
```

Out[15]:

	Title	Distance
1	Godfather: Part II, The (1974)	0.154159
2	Goodfellas (1990)	0.295961
3	Star Wars: Episode IV - A New Hope (1977)	0.349061
4	Pulp Fiction (1994)	0.352409
5	One Flew Over the Cuckoo's Nest (1975)	0.357212
6	Fight Club (1999)	0.358305
7	Reservoir Dogs (1992)	0.360162
8	American Beauty (1999)	0.360804
9	Silence of the Lambs, The (1991)	0.361809
10	Fargo (1996)	0.362356

Github Repository :- https://github.com/Adu3108/SoC.git

MovieLens-small Dataset :- https://drive.google.com/drive/folders/16Vv15_EsXlDCiYp-Q-yw7o20EFpNl2Hp?usp=sharing

Learning Value

With this project, we were able to apply all the concepts that we had learnt during the four weeks of SoC. We also got an opportunity to work with various datasets such as the MovieLens.

Software used

Jupyter Notebook , Google Colab

Suggestions for others

Before starting the project, one must have prior knowledge about KNN algorithms and clustering. There are various articles and YouTube videos online, if you are stuck at any point during the project.

References and Citations

Andrew NG Course -
https://youtube.com/playlist?list=PLLssT5z_DsK-h9vYZkQkYNWcItqhlRJLN/

Article Links-
https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1/

https://www.analyticsvidhya.com/blog/2021/04/simple-understanding-and-implementation-of-knn-algorithm/

https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761/

https://docs.scipy.org/doc/scipy/reference/generated/scipy.sparse.csr_matrix.html/

https://www.geeksforgeeks.org/python-implementation-of-movie-recommender-system/

Datasets-
https://grouplens.org/datasets/movielens/

Disclaimer

We have taken assistance of Resources such as the Stanford university Machine Learning course videos by Dr Andrew NG and certain articles from websites such as towardsdatascience.com , geeksforgeeks.com, and certain python library documentations.

Licenses

Pandas
https://pandas.pydata.org/pandas-docs/stable/getting_started/overview.html#license
Numpy
https://numpy.org/doc/stable/license.html
Scipy
https://www.scipy.org/scipylib/license.html
Matplotlib
https://matplotlib.org/stable/users/license.html
Seaborn
https://github.com/mwaskom/seaborn/blob/master/LICENSE
Python
https://docs.python.org/3/license.html