**STATS 402 - Interdisciplinary Data Analysis**

**<Recommender System for Social Media>**

Milestone Report: Stage 1

**<zehcen yuan, lian zeng, yechen wang>**

**<zy96@duke.edu,lz166@duke.edu,yechen.wang@dukekunshan.edu.cn>**

**Abstract**

A recommendation system is a system designed to recommend items of interest to end users. Since the 1990s, recommendation systems have become an important research area and have been applied to recommend movies, pictures, goods, TV shows, music, and many other areas. With the massive use of social networking services (SNS), recommendation systems have become increasingly important considering the user experience and profitability of SNS platforms. Common approaches to building recommendation systems include demographic filtering, content-based filtering, and collaborative filtering. In the past few years, deep learning methods such as natural language processing (NLP) have been applied to Recommender Systems to improve the quality of recommendations. However, training such deep learning models usually requires a large amount of user data on SNS platforms, which is difficult to obtain. Therefore, we used a dataset from a movie recommendation website, which can also be considered as a small SNS platform, as the substitution. Combining deep learning methods with three common filtering approaches, we want to develop an effective recommendation system for SNS platforms.

**1. The project rationale**

Despite the mass usage of SNS platforms, they are becoming increasingly popular in the last decade. According to statistics, famous SNS platforms such as Facebook and Youtube have about 2 billion users by 2018, which is as four times as the number of users back in 2010 [1]. The combination of SNS platforms with Recommender Systems brought new opportunities for businesses, considering their product marketing and advertising based on the Recommender Systems [2]. Also, Recommender Systems are beneficial to SNS platforms that want to improve their user experience by personalizing the content it pushes to its users [2].

Deep learning is becoming increasingly popular in the past few decades. The great success of the deep learning methods employed in many application domains such as computer vision and speech recognition has demonstrated its ability in solving complex tasks such as computer vision and speech recognition. Recently, deep learning methods have been proved able to overcome the obstacles encountered in traditional models and build Recommender Systems with higher quality [3]. Previous research has shown that hybrid Recommender Systems combining two or more recommendation strategies are more robust because of their complementary advantages [4], we will apply three of the most popular techniques used in Recommender Systems are Content-based filtering (CB) Demographic Filtering (DF) and Collaborating Filtering (CF) [5] to build a hybrid model for our Recommender System.

**2. The research content/objectives of this project and critical scientific problems to be solved**

The rapid growth of data collection has led to a new era of information. Data is being used to create more efficient systems and this is where Recommender Systems come into play. Recommendation enginesdiscover data patterns in the data set by learning consumers' preferences and produces the outcomes that co-relates to their needs and interests. Our project intends to find an efficient recommender algorithm works for our own SNS platform and encompasses three parts:

1. Popularity Based.

We will provide recommendations to every user, based on popularity and/or genre. The System recommends the same movies to users with similar demographic features. To be specific, this system will recommend more popular and critically acclaimed movies since they are more likely to be liked by average users. Same as the posts in social media platforms, the system will suggest popular posts by like count. However, each user is different and has various preferences , this approach is considered to be too simple. Thus, we decide to add two more methods to provide more accurate recommendations.

1. Classification Based

Diagram

Description automatically generatedThe second way to build a recommendation system is a classification model. It uses the features of both users as well as movies in order to predict whether this movie is liked or not by the user. It recommends similar items based on a particular feature. Different movies have different features. In the metadata we use, it includes movies’ genre, director, description, actors, etc. The basic mechanism for this system is that if a person likes a particular item or feature, he or she will also like an item that is similar to it. This idea also works for the posts in social network sites.  For instance, the description for the movies can be assumed as the contents of their posts or comments, the genre and keywords of the movies can represent the hashtag mentioned in the posts.

Figure 1 The Flow Chart of Classification Based Modeldata

1. Collaborative Filtering

Although everyone is unique, we can still find some similarities among their tastes. collaborative filtering models assumes that people like things similar to other things they like, and things that are liked by other people with similar taste. It will match persons with similar interests and provide recommendations based on this matching and recommend movies or posts to a user that similar users have liked.

**3. The proposed research plan and feasibility analysis (including research methods, technical routes, experimental methods, key technologies, etc.)**

1. **Demographic Filtering**

Recommender Systems based on Demographic filtering (DF) classify users according to their demographic information and make recommendations accordingly. User profiles are created by categorizing stereotypical descriptions of users, which represent the characteristics of user categories. [5, 6].

1. **Content Based Filtering**

This filtering system will analyze a particular item based on the item’s metadata. Including genre, director, description, actors, etc. In this system we might apply some methods including Natural Language Processing (NLP). The general idea behind these Recommender Systems is that all the items will be clustered into several groups, and if one user likes a particular item in a particular group, then he is more likely to like the item that falls into the same group [7].

1. **Collaborative Filtering**

This system matches persons with similar interests and provides recommendations based on this matching. Collaborative Filter-based recommendation systems use past activities or preferences, such as user ratings of products, rather than user or product content information [8]. In this filtering system we will invest in the performance of the attention model etc.

**4. Features and innovations of this project and the expected results**

Although our final goal is to build a recommendation system for our social network platform, the datasets that contain both user behavior and contents they post in a SNS are actually the trade secrets of those platforms. They are considered to be highly commercially sensitive. Thus, we decided to build our recommendation system on a dataset of movies first and try to apply this model we trained on our own SNS platform. During the training process, we assume the scores the users give to the movies as the user behaviors and the text description of the movies as the contents they post in the SNS platform.

There are basically three types of Recommender Systems including Demographic Filtering, Content Based Filtering and Collaborative Filtering. In this project, we innovatively create a Hybrid Recommendation system which leverages three methods together in order to provide users with more accurate recommendations. Once we have built up our own SNS, we will apply the model to the platform and test its functionality, reliability, and performance. We will be working on the perfection of the recommendation system to provide a better customer user experience.

**References**

[1] Ortiz-Ospina. 2019. The rise of social media. <https://ourworldindata.org/rise-of-social-media#licence>

[2] Magdalini Eirinaki, Jerry Gao, Iraklis Varlamis, Konstantinos Tserpes. 2018. Recommender Systems for Large-Scale Social Networks: A review of challenges and solutions. <https://doi.org/10.1016/j.future.2017.09.015>

[3] Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. 2018. Deep Learning based Recommender System: A Survey and New Perspectives. ACM Comput. Surv. 1, 1, Article 1 (July 2018), 35 pages.

[4] [Erion Çano](https://arxiv.org/search/cs?searchtype=author&query=%C3%87ano%2C+E), [Maurizio Morisio](https://arxiv.org/search/cs?searchtype=author&query=Morisio%2C+M). 2019. Hybrid Recommender Systems: A Systematic Literature Review. <https://arxiv.org/abs/1901.03888>

[5] Iateilang Ryngksai, L. Chameikho. 2014. Recommender Systems: Types of Filtering Technique. International Journal of Engineering Research & Technology (IJERT) IJERT ISSN: 2278-0181

[6] Adrian Benton, Raman Arora, Mark Dredze. Learning Multiview Embeddings of Twitter Users. <https://doi.org/10.18653/v1/P16-2003>

[7] Nada Ben-Lhachemi, El Habib Nfaoui. Using Tweets Embeddings For Hashtag Recommendation in Twitter. Procedia Computer Science, Volume 127, 2018, Pages 7-15, ISSN 1877-0509. <https://doi.org/10.1016/j.procs.2018.01.092>

[8]  Hao Wang, Naiyan Wang, Dit-Yan Yeung. 2015. Collaborative Deep Learning for Recommender Systems. <https://arxiv.org/pdf/1409.2944.pdf>