Movie Rating Predictor Using Nonnegative Matrix Factorization on Tag Data

Claire Chang, Thaxter Shaw, and TJ Tsai[1]

Abstract— **The goal of recommendation systems is to suggest suitable content such as movies, groceries, and more. This article aims to predict what someone will rate a movie, which could in turn be applied to a movie recommendation system. We propose a method of addressing this problem**

I. INTRODUCTION

The goal of this paper is to predict what someone will rate a movie on a five-star scale. This prediction can be used in recommendation systems (citation?) for not only movies, but also grocery shopping, music, and more.

There are several obstacles that make this problem challenging. One issue, as investigated in [citation Netflix Bellsmth] is the effect of time. Over time, users may change their baseline rating (for example, have higher expectations of graphics) and begin to prefer different genres and actors. Movies themselves also become more or less popular over time.

One approach that has been taken is to use a "neighborhood approach" that compares different movies, and recommends a highly rated movie that is similar to other movies a user has liked [citation]. Another approach is to use collaborative filtering (CF) methods (citation) such as nonnegative matrix factorization (NMF) and singular value decomposition (SVD). We could also use a hybrid approach of the neighborhood and CF approahces [citation Koren 2008].

In this paper, we use data from MovieLens [citation]. The dataset we use contains ratings of over 1000 movies made by various users, filtered such that each user has rated at least 20 movies. The dataset also relates each movie to a series of short tags, such as "sci-fi," "cliché," or "overrated." This data was collected through users applying labels to movies, and machine learning [citation Tag Genome].

Our system uses NMF on this tag data to create clusters of tags, which we will equate to "genres". We then use these genres predict ratings. Finally, we use mean movie ratings to adjust this predicted rating.

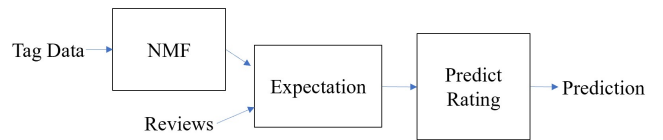## II. SYSTEM DESCRIPTION

The overall architecture for our system is shown in Figure 1.

Fig. 1. Block diagram of proposed system.

III. RESULTS

IV. CONCLUSION

REFERENCES

[1] test

[1]T. Tsai is with the Department of Engineering at Harvey Mudd College, 301 Platt Blvd., Claremont, CA 91711. E-mail: ttsai@hmc.edu