|  |
| --- |
| **Creating a Movie Recommendation Engine** |

**Dylan McCardle Corey Myers Evan Sheehan**

COMP 5600 COMP 5600 COMP 5600

*res0038@auburn.edu*

**1 Problem formulation**

Our final project for COMP 5600 will be creating a movie recommendation engine. The user will offer one or more movies they enjoy that they would like new recommendations based on. Therefore, the input will be some number of movies the user enjoys, and some number of movies the user dislikes, probably no more than five of each. The engine will take these movie titles and find the respective pages on IMDB. The engine will parse every review for each movie (possibly up to a given boundary) and create a word bank/bag based on these reviews. This word bank will keep track of the occurrences of unique words. The difficulty here is determining which words should be counted as a unique word and which should be ignored. Certainly, articles such as ‘the’ and ‘a’ will be ignored. **WE NEED TO IDENTIFY BASELINE AND ORACLE**

**2 Identifying data source**

The current plan is to use IMDB as a data source for reviews and use movie-map.com to find similar movies that can be parsed further for review data. The purpose of using movie-map.com for similar movies instead of IMDB is simply because movie-map.com is easier to automatically traverse and parse for results than IMDB.

<https://link.springer.com/article/10.1007/s11042-006-0082-7>

A hybrid approach for movie recommendation

“A hybrid approach for movie recommendation” proposes creating a movie recommendation engine using a hybrid of content-based and collaborative filtering techniques. Collaborative filtering predicts similarities between the active user and other users. The closest group of similar users is then used to make predictions for the active user. In contrast, content-based filtering is a broad term used to describe the extraction of some features from a source and comparing these features to features of other sources in order to make recommendations. The more similar the features, the more likely a recommendation will be made. This technique is closer to what we will be implementing. However, the article lists the pros and cons of both techniques, leading to the decision to combine them. If we find that our content-based filtering is not providing satisfactory results, this paper may be a useful reference to improve performance. The paper also mentioned a movie recommendation system called MoRe that we may consider using instead of movie-map.com to provide a dataset for recommendations.

**3 Literature review with reference list**

**References**

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D. S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609-616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System*. New York: TELOS/Springer-Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hiippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.