Ethics Pledge

Consistent with the above statements, all homework exercises, tests and exams that are designated as individual assignments MUST contain the following signed statement before they can be accepted for grading.

I pledge on my honor that I have not given or received any unauthorized assistance on this assignment/examination. I further pledge that I have not copied any material from a book, article, the Internet or any other source except where I have expressly cited the source.

Signature: Haodong Zhao Date: Apr 7th. 2019

Please note that assignments in this class may be submitted to

www.turnitin.com, a web-based anti-plagiarism system, for an evaluation of their originality.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Reading review**

**A Survey of Collaborative Filtering Techniques**

The recommendation system can help people screen information that they are interested in and valuable. The database used by CF technology is other topics or products that users may like to predict for credit. The challenge of collaborative filtering is that the CF algorithm needs to have the ability to process highly sparse data, expand with the number of users and projects, and make satisfactory recommendations in a short period of time.

Early collaborative filtering systems used user rating data to calculate similarities or weights between users or projects. However, memory-based CF technology has limitations, such as similarity based on common terms, so it is unreliable when data is sparse and public items are few. In order to obtain better prediction performance and overcome the shortcomings of memory-based CF algorithm, the model-based CF method is studied. Model-based CF techniques use pure rating data to estimate and learn models (data mining or machine learning) for prediction.

Characteristics of collaborative filtering: The degree to which high quality predictions or recommendations are generated depends on the challenges they are dealing with.

Collaborative filtering challenges:

1. Data sparsity. Collaboratively filtered user term matrices are often very sparse, and the predicted or recommended performance of the CF system is challenged. It can be optimized by dimensionality reduction.
2. Scalability. As the number of users and projects grows dramatically, traditional CF algorithms will encounter scalability challenges that exceed actual or acceptable levels of computing resources. The item-based Pearson correlation CF algorithm and SVD can make the recommendation system overestimate scalability.
3. Synonymous. Most recommendation systems are unable to find potential links to projects with very similarities, and thus handle these projects differently. SVD technology can handle this problem.
4. Gray sheep. Gray sheep refers to the users whose opinions do not consistently agree or disagree with any group of people and thus do not benefit from collaborative filtering
5. Shilling Attacks and other challenges.

Memory-Based Collaborative Filtering Techniques

1. Similarity Computation: Correlation-Based Similarity; Vector Cosine-based Similarity; Other Similarities
2. Prediction and Recommendation Computation: Weighted Sum of Others’ Ratings; Simple Weighted Average.
3. Top-N Recommendations: User-Based Top-N Recommendation Algorithms; Item-Based Top-N Recommendation Algorithms.
4. Extensions to Memory-Based Algorithms: Default Voting; Inverse User Frequency; Case Amplification; Imputation-Boosted CF Algorithms; Weighted Majority Prediction.

Model-Based Collaborative Filtering Techniques

1. Bayesian Belief Net CF Algorithms: Simple Bayesian CF Algorithm; NB-ELR and TAN-ELR CF Algorithms; Other Bayesian CF Algorithm.
2. Clustering CF Algorithms.
3. Regression-Based CF Algorithms.
4. MDP-Based CF Algorithms.
5. Latent Semantic CF Models.
6. Other Model-Based CF Techniques.

Hybrid Collaborative Filtering Techniques

1. Hybrid Recommenders Incorporating CF and Content-Based Features.
2. Hybrid Recommenders Combining CF and Other Recommender Systems.
3. Hybrid Recommenders Combining CF Algorithms.

Evaluation Metrics

The quality of a recommender system can be decided on the result of evaluation. Metrics evaluating recommendation systems can be broadly classified in to following broad categories:

1. Mean Absolute Error (MAE) and Normalized Mean Absolute Error (NMAE).
2. Root Mean Squared Error (RMSE).
3. ROC Sensitivity.

After solving the current challenges, future CF technologies should be able to make accurate predictions with shilling attacks and noisy data and can also be effectively applied in fast-growing mobile applications. And because of the nature of the CF task, manual data is usually not reliable, so the real data set of the field test is more suitable for CF research.