

## ABSTRACT

Ever increasing volume of data on the Web has resulted in information explosion. One of the major challenges faced by a user is to find relevant information catering to their interests. To alleviate this problem recommender system is used to search for and filter information based on the user's interest. In the news domain, recommendation technique particularly aims at collecting news articles according to the user's interests with the objective of creating a personal newspaper. The popularity of news recommender systems is attributed to availability and easy accessibility of news from websites like Google and Yahoo. However, the driving problem is to identify and recommend the most interesting news articles to each user balancing user interest with importance of the news.

Some of the challenges faced by the news domain include dynamic nature of the news domain, changing user interests, popularity and novelty of the news, lack of history information for new users and the volume of information. These challenges result in cold-start, sparse data situations along with scalability and overspecialization problems. The major goal of this research is to design and develop a personalized news recommender system capable of handling data sparsity, cold-start, dynamically changing user interests and scalability issues. Experiments are performed on NEWS and YOW datasets. The evaluation metrics used in this thesis are Precision, Recall and F-measure. The contributions of this research includes,

- Using rough set based Collaborative Filtering technique with Constrained Pearson Correlation Coefficient (CPCC) to impute missing category values in user profiles. This technique considers explicit user feedback to handle cold

start problem. Experimental analysis is performed on different missing value imputation techniques including mean imputation, user based imputation using PCC, user based imputation using cosine similarity and the proposed approach. The proposed approach shows a reduction in a Mean Absolute Error by 24.76%, 22.72%, 12.88% over mean imputation, user based imputation using PCC and user based imputation using cosine similarity on NEWS dataset. Mean Absolute Error is reduced by 20.4%, 15.02%, and 20.56% over mean imputation, user based imputation using PCC and user based imputation using cosine similarity on YOW dataset.

- Detecting novel documents using different novelty detection techniques like MeanCS based novelty, cosine similarity based novelty detection and the proposed novelty detection. On an average the proposed novelty detection technique provides 8% and 11% of improvements in F-Measure when compared to MeanCS based novelty detection (Allan et al. 2004) and cosine similarity based novelty detection (Margarita Karkali et al. 2013) approaches.
- Applying adaptive user profiling to accommodate change in dynamic user interests. Based on number of clicks made by each user on a category, the user profiles are updated. Experimental analysis is performed on different news recommendation frameworks including GOO, LOGO, Click-B and the proposed framework. On an average, the proposed framework provides 3%, 5%, 7% improvement in F-measure when compared to LOGO, ClickB and GOO systems on NEWS dataset. On an average, the proposed

approach provides 4%, 8%, 13% improvement in F-measure when compared to LOGO, ClickB and GOO systems on YOW dataset.

- Analyzing the existing similarity measures and proposing new Modified Heuristic Similarity Measure (MHSM) for Collaborative Filtering by considering local and global preferences of users. This effectively addresses data sparsity issue. The proposed similarity measure provides 12%, 10%, 8%, 4% of improvement in F-measure over LOGO, ClickB, GOO, and News recommendation using Roughset based CPCC on NEWS dataset. On an average, the proposed similarity measure provides 21%, 17%, 13%, 8% of improvement in F-measure over LOGO, ClickB, GOO and News recommendation using roughest based CPCC on YOW dataset.
- Reranking articles based on personalization attributes such as individual user profiles, group user Profiles, novelty, popularity and recency to avoid overspecialization problem. Hierarchical clustering is used to identify the articles read by other similar users but not read by the active user. Genetic Algorithm is used to weight personalization attributes for a particular user. Experimental analysis is performed on SCENE, user based k-nearest neighbor approach and the proposed approach. On an average, the proposed approach provides 10% and 4% of improvement in F-measure over SCENE and user based k-Nearest Neighbor approach respectively on NEWS dataset. The proposed approach also provides 10.75% and 2.63% improvement in F-measure over

SCENE and user based k-Nearest Neighbor approach respectively on YOW dataset.

- Using hypergraph based learning to represent higher order relationship between news articles and user profile. Experimental analysis is performed on news recommendation via hypergraph learning proposed by Lei li et al. model and proposed model. On an average, the proposed model provides 4.7% and 6.36% of improvement in F-measure over Lei et al. model on NEWS and YOW datasets respectively. The proposed model improves the execution time by 19.005%, when compared to conventional graph representation of user profiles.

Experimental results demonstrate that incorporating personalization attributes like individual user profile, group user profile, popularity, recency and novelty during recommendation can achieve a good balance between the user satisfaction and the recommendation process.