Neha M

Reg. No: 14CO127

8095299474

[nehamohan1996@gmail.com](mailto:nehamohan1996@gmail.com)

**Improving Scalability of Personalized Recommendation Systems for Enterprise Knowledge Workers**

Implementation Details

* Metadata feature extraction: Folder features, token features (path feature) and extension features.
* Metadata modelling: Creating models for each user using the metadata features.
* Collaborative filtering aware modelling: Use the metadata models of all the users to create a CF model for each of the users.
* Apply Active Features-based Model Selection (AFMS) to reduce the time complexity during testing phase.
* Evaluation : Varying testing period and training period
* Calculate Recall and precision and check the performance of the system.

Work done so far

1. Collecting the dataset :

* Collected around 500 files from 20 students, such that there is some collaboration.
* Various types and formats: txt, pdf, jpg, png, py, c, pptx, docx, etc.

1. User Access: A binary vector having cardinality=total number of files, has been created for each user. The value is 1 if the user has accessed or created it.
2. Extracting necessary metadata features:

* Folder feature: X (file, Folder) (i) is 1 if Folder (i) lies in the path of file f .
* Extension feature: X (file, Extensions) (i) is 1 if file has the extension Extensions (i).
* Token feature: Tokenize the file path and construct a vocabulary based on the popular tokens.
* The python module ‘os.path’ has been used to carry out all the operations regarding file path.
* For tokenization of the file paths, bag of words representation is done using scikit-learn.

Work to be done

1. Metadata Modelling: The features which are extracted will be used as a parameter to build metadata models for each of the user.
2. Collaborative filtering aware modelling: Use the metadata models of all the users to create a CF model for each of the users.
3. Apply Active Features-based Model Selection (AFMS) to reduce the time complexity during testing phase.
4. Evaluation : Varying testing period and training period
5. Calculate Recall and precision and check the performance of the system.