****Abstract****

Introduction

Recommender systems facilitate finding of the apposite object for user based on various parameters. Parameters are the attributes of the object in question. Recommendations can be provided on the basis of object relevance as well as shared review of other users about the object. Playing an important role of descriptor, the attributes define the route to finding appropriate recommendations such as products and literature. The purpose of this paper is to propose recommendation technique and suggest appropriate papers from IEEE.

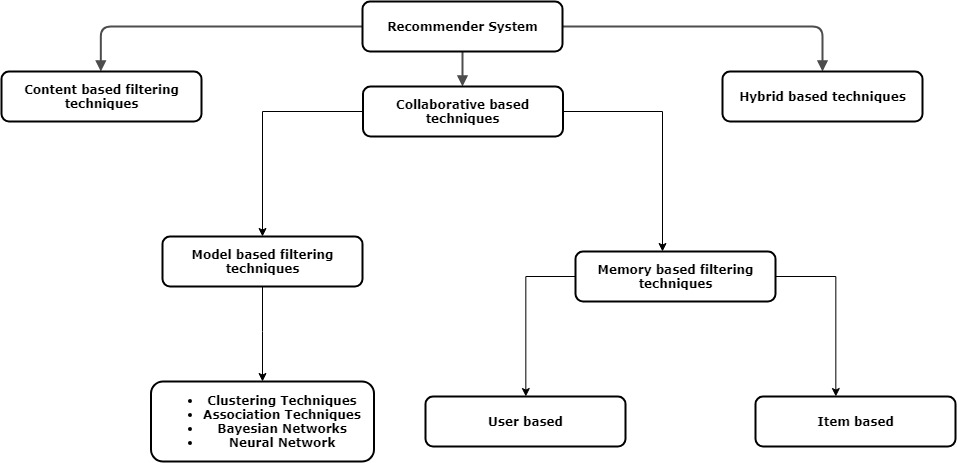
Usage of Recommendation systems is quite extensive in today’s era as it's able to track and understand requirements of user through elegant methods and algorithms which fall under several filtering techniques of machine learning.

Recommendation filtering techniques

Filtering techniques are used in Recommendation systems to order or rank and suggest required article to user through different approaches, based on the types of parameters used to predict the content.

These parameters are those features which are provided to the filtering algorithms in use; the necessary input on basis of which system decides the output.

Figure shows the anatomy of different recommendation filtering techniques.



Content based**:**

Content based filtering concentrates on document relevancy and suggestions, which is suited for model based recommender system utilizing data mining techniques, creating an apposite model. The difference is that data taken is only from one user or item and recommendation is made on the basis of studying one particular subject.

The methods that can be used to achieve the recommendation task are:

1. Neural Network
2. Data Mining Techniques

Hybrid system tends to apply both up to a certain extent.

Detailed description is in order as this particular technique is used to obtain the recommendations in this paper.

The method that can be used to achieve the recommendation task are:

1. Neural Network

Neural Network achieve great success in different field of AI such as NLP(Natural Language Processing), Image recognition, speech recognition and so many fields are there in which neural network can be used and it gives best result.

Neural network can be applied in recommendation system. We obtain features of users and items and feed into neural network and output of which gives probability of the score that user might give.

Neural network can be applied in recommendation system for content and collaborative recommendation.

1. Data Mining Techniques

Certain Data Mining techniques are present in market which assist in prediction of required item.

As stated in Introduction portion, there exists a three phase method in doing so.

Phases: -

1. Pre-processing. This reduces and filters out unnecessary items which do not pertain to the user input query. Methods such as Cosine Similarity, Pearson Correlation and Jaccard Similarity exists to pre-process the data and convert the input into suitable matrix.
2. Analysis Phase. Methods such as Bayesian Networks, SVM, Decision Trees, K Nearest Neighbours, Artificial Neural Networks fall under this domain to analyse the given input matrix/data and analyse the pattern of the data and classify / cluster / associate rule mine. These methods analyse the input matrix and find pattern in the data and either group or predict and passes the data to interpretation phase.

Collaborative based techniques

Collaborative approach takes multiple user's experience and rating into account as opposed to content based filtering which takes relevance of object to given query considering its attributes. It assumes that user with similar opinion on a particular topic will have the same preference for another. This requires finding patterns of preferences and reviews between multiple users and classify them. The next step to be taken is to offer the predictions on the basis of class or the preferences of the nearly similar user pattern. Amazon is known to use item to item collaborative filtering[1].

For example, Rating/Review/Opinion of any paper or movie or books collected from users for content based system.

**Two techniques which are commonly used for collaborative filtering:**

1. Model based

Involves model building using several algorithms falling in the domain of data mining, machine learning etc., to find relation between unrated data and recommend appropriately.

1. Memory based

Takes help of user ratings to find similarity between users or items. This is usually done by Pearson correlation or cosine similarity.

**Characteristics of Recommendation System**

1. Response Time.

Ideally, the response should be instantaneous irrelevant of the size of data set.

1. Scalability.

Models can be trimmed limiting the number of entries and keeping the model scalable. This should however not affect the relevance of results.

3. Flexibility.

4. Relevant Results.

5. Dynamic Nature.

If the data set is changed or the query is altered, the effect should be visible in the result. The priority of attributes should be set in a way that the user finds recommendations accurate. In case of Document recommendation, the query words should be provided appropriate weights to find accurate recommendations.

Literature review

1. Data mining methods for recommender systems by Xavier Amatriain, Alejandro Jaimes, Nuria Oliver, and Josep M. Pujol.

Certain Data Mining techniques are present in market which assist in prediction of required item.

As stated in Introduction portion, there exists a three phase method in doing so.

First is Pre-processing. This reduces and filters out unnecessary items which do not pertain to the user input query. Methods such as Cosine Similarity, Pearson Correlation and Jaccard Similarity exists to pre-process the data and convert the input into suitable matrix.

Second is Analysis Phase.

Methods such as Bayesian Networks, SVM, Decision Trees, K Nearest Neighbours, Artificial Neural Networks fall under this domain to analyse the given input matrix/data and analyse the pattern of the data and Classify / Cluster / Associate Rule mine. These methods analyse the input matrix and find pattern in the data and either group or predict and passes the data to interpretation phase.

1. A Document Recommendation System Using a Document-Similarity Ontology by Rodrigo Vences Nava, Victor Hugo Menendez Dominguez, Jorge Gomez Montalvo.

Use of Tf-Idf to match the document and find similarity is done and the most similar documents are recommended to the user.

Transitivity similarity is used and whenever the user choses a particular document, by the transitive rule the documents similar to chosen document are predicted and are recommended. The Tf-Idf measures the similarity by checking the frequency of occurring keywords in the original document.

1. Product Recommendation System by Jianfeng Hu Bo Zhang.

Vector Space Model is used to find item similarity considering rating (prominent part of Collaborative System), following which Bipartite graph is used. In the graph the nodes define ratings given by customer and edges exist between nodes pertaining to different sets exclusively. Finally the paper discusses the use of spanning tree finding novel means to express similarity between customers and products.

Methodology

In this paper we propose solution to recommend papers to users based on their search query with help of content and collaborative recommendation. The Recommendations will be based on relevancy between searched keywords and concept words of a papers. This similarity will be found by applying TF-IDF and Vector space model. To obtain concept words Latent Semantic Analysis (LSA) has been applied on top 10 research papers obtained from search query of users, searched as keywords on IEEE API. Restriction on number of keywords and weight to appropriate keywords haven't been applied, making each search parameter have equal value in recommended material.

1. Obtain search keyword from user: -

Users will feed in the keywords according to their requirement. These keywords will be used as main search query which will be than fired through IEEE API to obtain top 10 research papers.

TF-IDF and vector space model is applied on abstracts of downloaded papers. Further, LSA is implemented to find the concept words of these documents. With the obtained list of concept words, user can select the concept which is more suitable to him and a new query will be fired with his searched keyword and his selected concept words to get better and more relevant results.

1. Term Frequency – Inverse Document Frequency(TF-IDF): -

The TF-IDF algorithm is used to weigh a keyword in any content and assign the importance to that keyword based on the number of times it appears in the document. The TF-IDF algorithm is statistical process of reflect the importance of a word in respect to a document in a collection. TF-IDF can be evaluated as follows: -

1. Term frequency:

It measures how frequently a term occurs in a document. Since every document is different in length, it is possible that a term would appear much more times in long documents than shorter ones.

**Formula:**

1. Inverse document frequency:

Measures how important a term is. While computing TF, all terms are considered equally important. However, it is known that certain terms, such as "is", "of", and "that", may appear a lot of times but have little importance.

**Formula:**

1. Vector space model: -

A vector can be expressed as a sum of elements such as,

Where ak­ are scalar and v­ik as component of elements.

A vector space is defined by a set of linearly independent basis vectors. Here, a document is D1 is presented in an n-dimensional vector where each dimension correspond to a distinct term and n is the total number of terms used in collection of documents.

On papers downloaded from IEEE, Vector Space Model will be applied to determine relevancy ranking of documents in a keyword search can be calculated, using the assumptions of document similarities theory, by comparing the deviation of angles between each document vector and the original query vector where the query is represented as the same kind of vector as the documents.

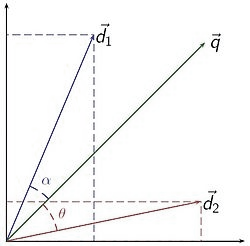


Figure Vector space model

1. Latent semantic analysis: -

Latent semantic analysis (LSA) is one method of analysing relationships between documents. Assumption made by LSA is that the words in semantic proximity should occur in documents which are closely related. Word count in a paragraph matrix is constructed and Singular Value Decomposition is applied to lower the rank of the matrix by reducing number of rows.

LSA can use term document matrix and reduce the sparse nature of matrix and create dense matrix. The term document matrix might happen to have large amount of text and LSA by using the property of Eigen Value and Eigen Vector can reduce the dimensionality of a matrix.

Eigen value λ is a scalar value corresponding to the Eigen vector X when a linear transformation is applied on Vector X.

A\*X = λX.

X = Term Document matrix.

X’ = Transpose of X.

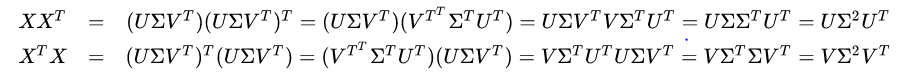
X = UΣV’.

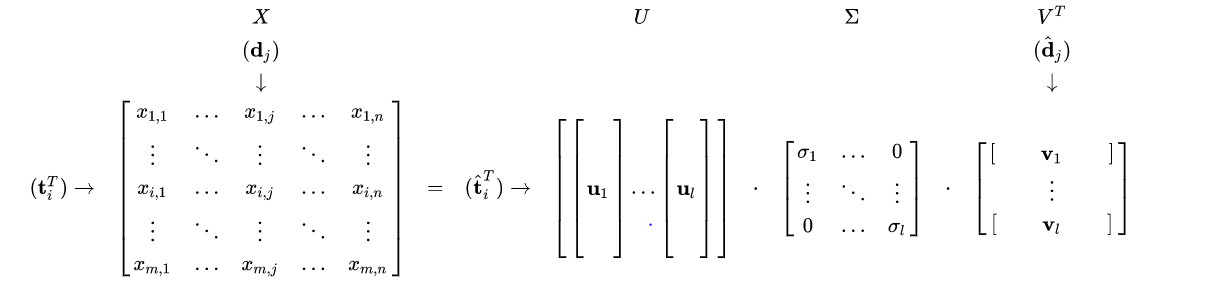
Singular Value Decomposition factorizes a given matrix and provides with three matrices explained as follows.

U → Unitary Matrix.

Σ → Diagonal Matrix.

V’→ Unitary Matrix.

The columns of U are u (i) which are left singular vectors of X. The columns of V are v (i) which are right singular vectors of X. The properties of X in terms of U and V are as follows. U is document-concept matrix where rows are documents and columns are concepts. V is term-concept matrix where rows are terms and columns are concepts. Σ is square matrix where its elements are amount of variation in each concept.



U and V (transpose) are left and right singular vectors and are singular values.

Matrix representation: 

The matrix U(k) gives concept – document matrix and V(k) transposed gives concept – term matrix.

For Recommendation system, U(k) can be used to fetch the concept words and further recommend a few words which can then be added to complete and personalize a query.

Vector Space suffers from POLYSEMY and SYNONYMY which can be solved using LSA model as it is able to create and group the concept words.

This mitigates the problem of identifying synonymy, as the rank lowering is expected to merge the dimensions associated with terms that have similar meanings. It also mitigates the problem with polysemy, since components of polysemous words that point in the "right" direction are added to the components of words that share a similar meaning.

**Future scope:**

**Conclusion:**

The system provides recommendation based on the relevancy on a document which is obtained by applying tf-idf and extracting concept words with help of LSA and vector space model, based on the final query. The system however can use multitude of intermediate algorithms such as K-means, decision trees and Neural Networks to improve output accuracy, with latter being the most preferable.

However, the results featuring range of precise articles were duly appreciated by users.

Due to iterative suggestions and corrections in user query as well as attribute weights the system can achieve excellent results.

**Reference**

[1] Amazon.com Recommendations Item-to-Item Collaborative Filtering by Greg Linden,Brent Smith,and Jeremy York