VARIous METHODS OF NEWS PERSONALISATION

Priya Chaudhari¹*, Richa Makhijani²

¹Department of Computer Engineering, JTM COE, Faizpur, Maharashtra, India
²Department of Computer and Science Engineering, SSGB COET, Bhusawal, Maharashtra, India

ARTICLE INFO

ABSTRACT

1. INTRODUCTION

The explosive growth of the Internet, the emergence of e-commerce, and the huge amount of multimedia materials generated in different fields, such as science and leisure, have led to the design of specific systems to help users to find relevant material. More specifically, the aim of personalization systems is to help users either to find what they are explicitly looking for, or what they would find useful but cannot easily access because they do not know about their existence, giving them a step ahead in the context of information retrieval (IR).

Personalisation aims to customize the results on a user's explicit and/or implicit interests and desires. The move to personalization is no longer an option, but a necessity. Several challenges however come into place for the personalization procedure to be successful, scalability, accuracy, evolving user interests, data collection and preprocessing, intergrading multiple sources of data, as well as privacy issues are only some of the aspects that our system faces up. From an architectural and algorithmic point of view, Personalisation is either based on content-filtering or collaborative filtering or both systems. Personalising news feeds is an interesting subtask of news filtering and personalisation. Its target is to effectively separate interesting news articles for a user from a large amount of documents.

The personalization of newspapers involves both social and technical issues. In social terms, the personalisation of newspapers raises some interesting issues. Readers may be tempted to restricted to reading only personally interesting articles, if their newspaper tried to show them only the articles that they are likely be interested in. However, from the readers' point of view, newspapers are not only means to read articles they are interested in but also are means to find information which they are not explicitly looking for. In regular newspapers, people tend to regard an article as important if it is located at an important location on the page or allocated more space than other articles. They may read the article not because it is interesting, but because it seems to be important. Eventually, they may be able to broaden their interests by reading such articles. A purely personalized newspaper would lack this important feature. In technical terms, the manner in which the user's interest is measured, and the strategy used to personalize the presentation are important.

The growing number of online newspapers in last years has presented a rich area which can immensely profit from personalised filtering approaches. Online news reading has become very popular as the Internet provides access to news from millions of sources all over the world. People typically read news to know and understand what happened, is happening and will happen in a town, region, country or the world. Thus, it is a key challenge for news websites to help users find, fast and accurately, those news which are of interest for the online readers.

News recommender systems offer recommendations, usually based on content similarity to previous visits by the user (content based approaches) or on news items visited by similar users (collaborative filtering). Nevertheless, it has two characteristics that distinguish it from other recommendation tasks.

2. LITERATURE SURVEY

2.1. TF-IDF

There are many term weighting methods available, such as for example probabilistic weighting, term frequency (TF) weighting, inverse document frequency (IDF) weighting, TF-IDF weighting, variations of TF-IDF weighting, etc.[5]. The main term weighting method that is focused on specifically in the work presented in this paper is the traditional TF-IDF weighting scheme. A classic approach in comparing documents is the use of TF-IDF together with the cosine similarity measure. TF-IDF is a statistical...
method used to determine the relative importance of a word within a document in a collection (or corpus) of documents.

2.2. CF-IDF

There has been some previous work on the use of TF-IDF with concepts (similar to CF-IDF). In [6] a conceptual indexing method based on WordNet, a large lexical database, is proposed. This approach represents document contents by the semantic network called document semantic core. The documents are mapped on the WordNet semantic network and converted from a set of terms to a set of concepts.

2.3. Content-Based Recommenders

In content-based approaches to news recommending, articles are recommended according to a comparison between their contents and the user profiles. The user profiles contain information about the users' content-based preferences. Both of these components have data-structures which are created using features extracted from text. A weighting scheme is often used to assign high weights to the most discriminating features/preferences, and low weights to the less informative ones.

2.3.1. Traditional Content-Based Approaches

Traditional content-based approaches are purely content based without any semantics. Concepts get weights assigned that are obtained without semantic knowledge of underlying relations between the concepts. User interests are often measured with machine learning algorithms, like Nearest Neighbor or Naive Bayes. In the traditional content-based approaches we review, articles are processed with TF-IDF by taking all terms (but the stop words) into account. The article is stored in a weighted vector of terms, and compared with a user profile by using a similarity measure.

- News Dude [8] is a personal news recommending agent that uses TF-IDF in combination with the Nearest Neighbor algorithm and uses the full text of an article. News Dude first considers the short-term interests to look for similar items and if this does not return satisfiable results, long-term interests are considered.

- The next related work is Daily Learner [4]. This is an adaptive news service which allows users to personalize the news to their own taste. First a user gives his preferences of what type of news he is interested in. Based on this user profile, the system then delivers those stories that best match this user's interests. A new article is processed with TF-IDF, and represented as a vector. Then this article is compared with the user profile (also a vector with TF-IDF weights), using cosine similarity. Finally, the user explicitly provides feedback using four ratings (interesting, not interesting, more information, already known). Short-term interests are determined by analyzing the N most recently rated stories, based on the Nearest Neighbor Algorithm. Long term interests are modeled with the Naive Bayes Classifier.

- Your News [1] is another example of a content-based news recommendation system. It is a personalized news system, which intends to increase the transparency of adapted news delivery. It allows the user to view and edit his interest profile. To support this, Your News highlights the key terms in news items. The news items are represented as weighted vectors of terms. The weight of each term is calculated using TF-IDF.

CF-IDF does not consider the full text, but only the concepts that exist in the knowledge base. With the semantic knowledge about the concepts it is possible to consider more than just the text at hand. The strength of the CF-IDF algorithm depends on the quality of the knowledge base.

2.3.2. Semantic Content-Based Approaches

Semantic content-based approaches aim to recommend news items by combining content-based techniques with domain semantics. Weights for concepts take into account the semantic knowledge about these concepts. Each of the reviewed recommenders has a different approach of applying the semantic knowledge provided by the ontology. The CF-IDF recommender only records the concepts to calculate weights.

- The approach proposed in [2], which was created in the same environment as our CF-IDF approach, calculates a similarity based on not only the concepts themselves but also based on the directly and indirectly related concepts, which are described in an ontology.

- OntoSeek [3] is a content-based approach which aims to retrieve information from online yellow pages and product catalogs. It matches content with the help of the large Sensus ontology, which comprises a simple taxonomic structure of approximately 70,000 nodes. OntoSeek does not employ a user profile. Instead, OntoSeek uses lexical conceptual graphs to represent queries and resource descriptions, i.e., a tree structure where nodes are nouns from the descriptions and arcs are concepts inferred by the corresponding nouns. The ontology is used for classifying items, and to match an item with a query. The user is required to disambiguate the meaning of his queries. This process is performed by the user interface that tries to identify the concept provided and asks the user to choose between potential solutions.

- The authors of [7] propose News@hand, a news-based recommendation system which uses Semantic Web technologies to describe and relate news items and user preferences in order to recommend items to a user. To represent news contents and user preferences the authors make use of concepts which appear in a set of domain ontologies. News@hand looks very similar to the Hermes News Personalization framework. Both approaches classify news items to gain key concepts, and work with a domain ontology. For recommending, News@hand makes use of 3 different semantic methods for recommendations: content-based, collaborative filtering, and a hybrid approach. The latter two are not discussed since this paper focuses primarily on content based approaches. The semantic content-based recommendation approach employs a certain similarity measure that utilizes the semantic preferences (weighted concepts gained by observing and profiling user behavior) of the user and the semantic annotations (the key concepts weighted by the classification) of an item. CF-IDF is mainly created to prove that a term-based recommender can be significantly improved with the help of the semantic annotations, whereas the content-based approach in News@hand is mainly used for comparison with the hybrid recommendation approach.

CONCLUSIONS

Various method of News Personalisation personalise the news to the user as per that user’s choice. When any user wants to read news of particular category, he has to search that news. So various methods of news personalization provides this facility. Some methods provides news to the
user based on content based approach, some based on collaborative based approach and some based on both approaches. When subscribed user gets login, the personalization system provides news to that user of his choice from the news database as per user’s choice is known to the system.

REFERENCES
[6] Baziz, M., Boughanem, M., Traboulsi, S.: A Concept-Based Approach for Indexing Documents in IR. IRIT, Campus universitaire ToulouseIII 118 rte de Narbonne, F-31062 Toulouse Cedex 4, France {baziz, boughane, traboul@irit.fr