

Online News Recommender System

Mahta Moattari and Weichang Du
University of New Brunswick

Introduction

Recommender Systems (RS), have become a significant research area in the mid-1990s, after the first papers on collaborative filtering. They can be categorized as content-based, collaborative or hybrid.

- In Content Based Filtering (CBF) method the algorithm will suggest new items to user based on user interests in the past.

- Collaborative Filtering (CF) act based on the user-user similarity.

- Hybrid recommender system is combination of content-based and collaborative filtering. These two methods can be combined in different ways.

Purpose

we aim to design a recommender system for news sites like Canadian Broadcasting Corporation (CBC). Reading news online means you can access various news providers' articles around the world, that's why nowadays, online news reading is very popular. But having so many varieties, engaged us in the problem of how to find the desired ones as easy and fast as we need.

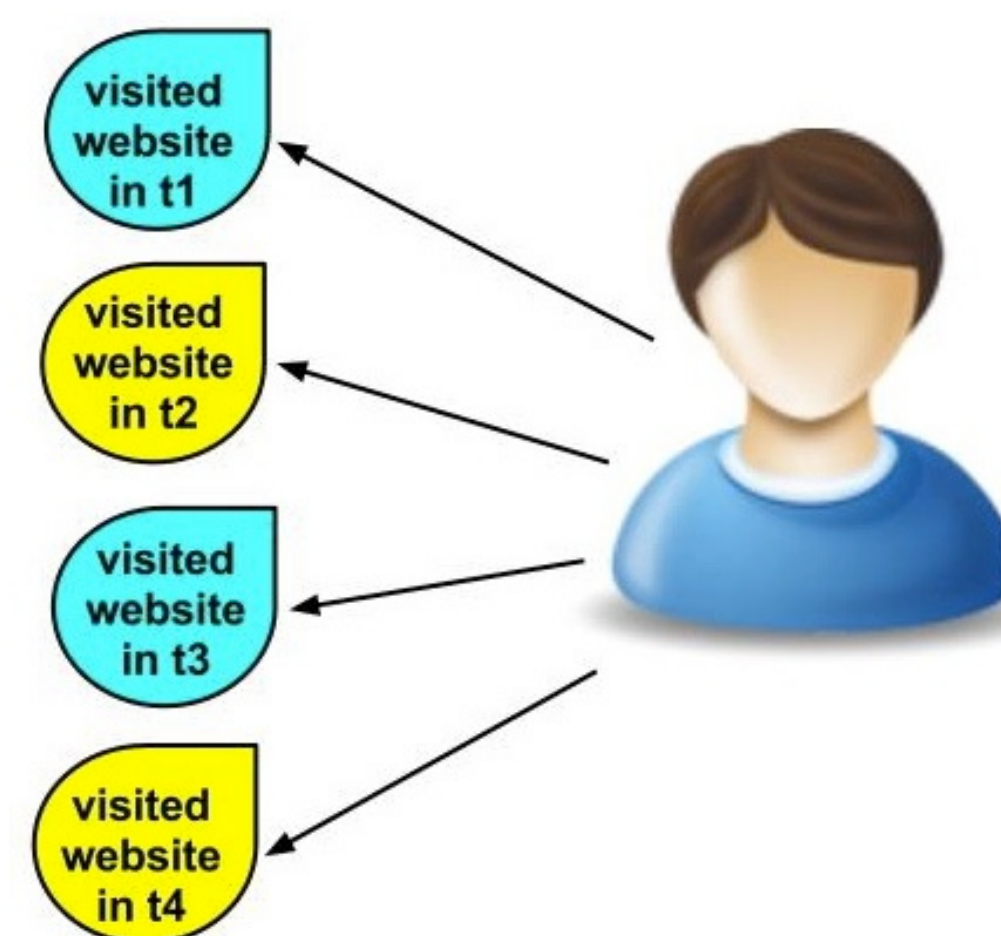
As it is clear the answer is using recommender systems and the most popular one in this field is content-based recommender system. We would like to create a RS based on CBF, which can perform well even with small amount of data.

Methods

- All the items that are going to be recommended should be classified. For this matter in most RS takes advantage of the classification algorithms like KNN, Naïve Bayes, ID3, etc. But the articles in a news site are already categorized based on their subject. So we can use the same categories.



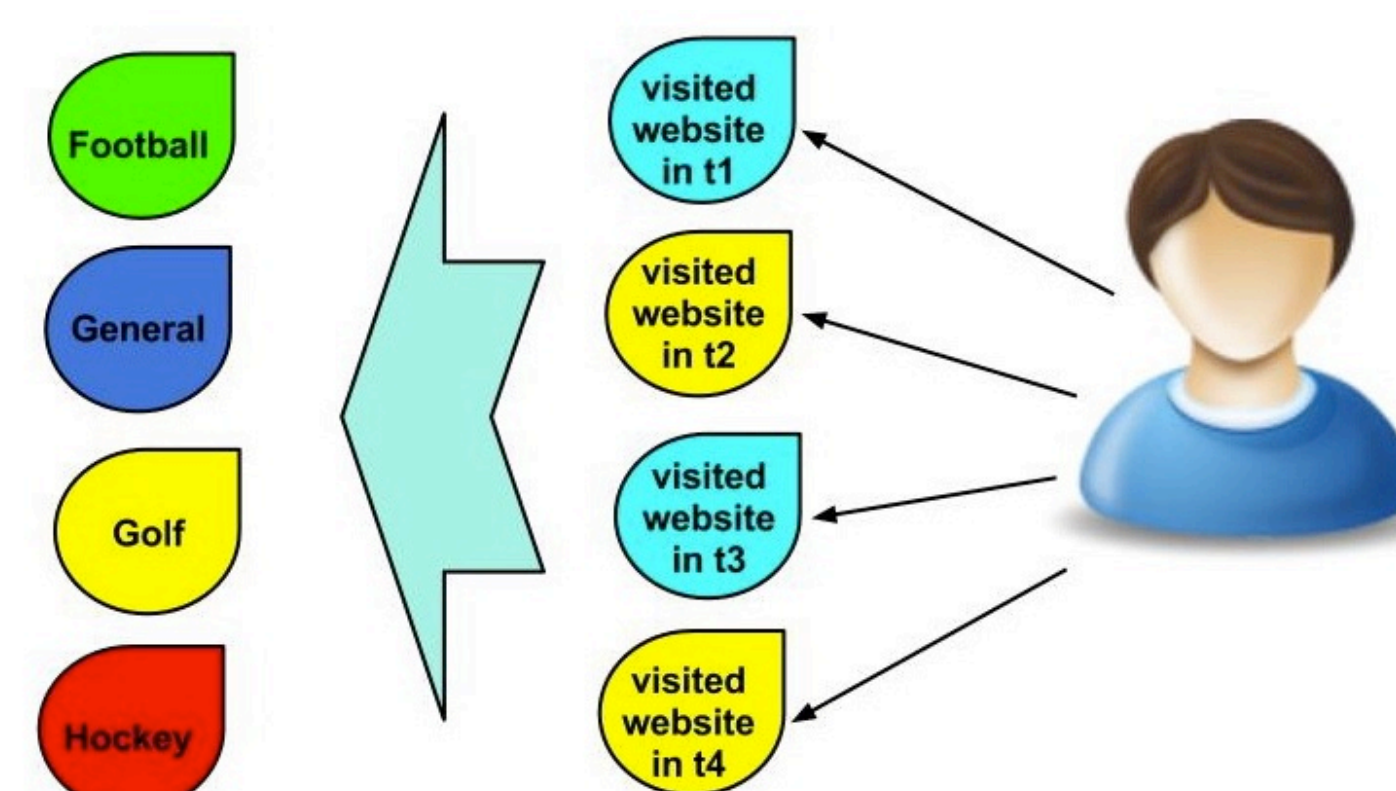
- Our next step is to create a user profile. user profile demonstrate user interests. A user profile can be created manually (explicitly). On the other hand monitoring the user activity over the web can provide implicit feedback.



- Using Machine Learning algorithms we can estimate future user interests based on user interests in the past (user profile). We used Bayesian Rule to make this estimation. The user top n will be recommended to the user.

$$interest(category = c_i) = \frac{\sum_t \left(N^t \frac{P^t(category = c_i | CM) P^t(CM)}{P^t(category = c_i)} \right)}{\sum_t N^t}$$

- Where N is the number of articles that user visited in time t, and it is used for normalization. Interest is the probability that a specific user post a comment (CM) for an article in specific category given the category.

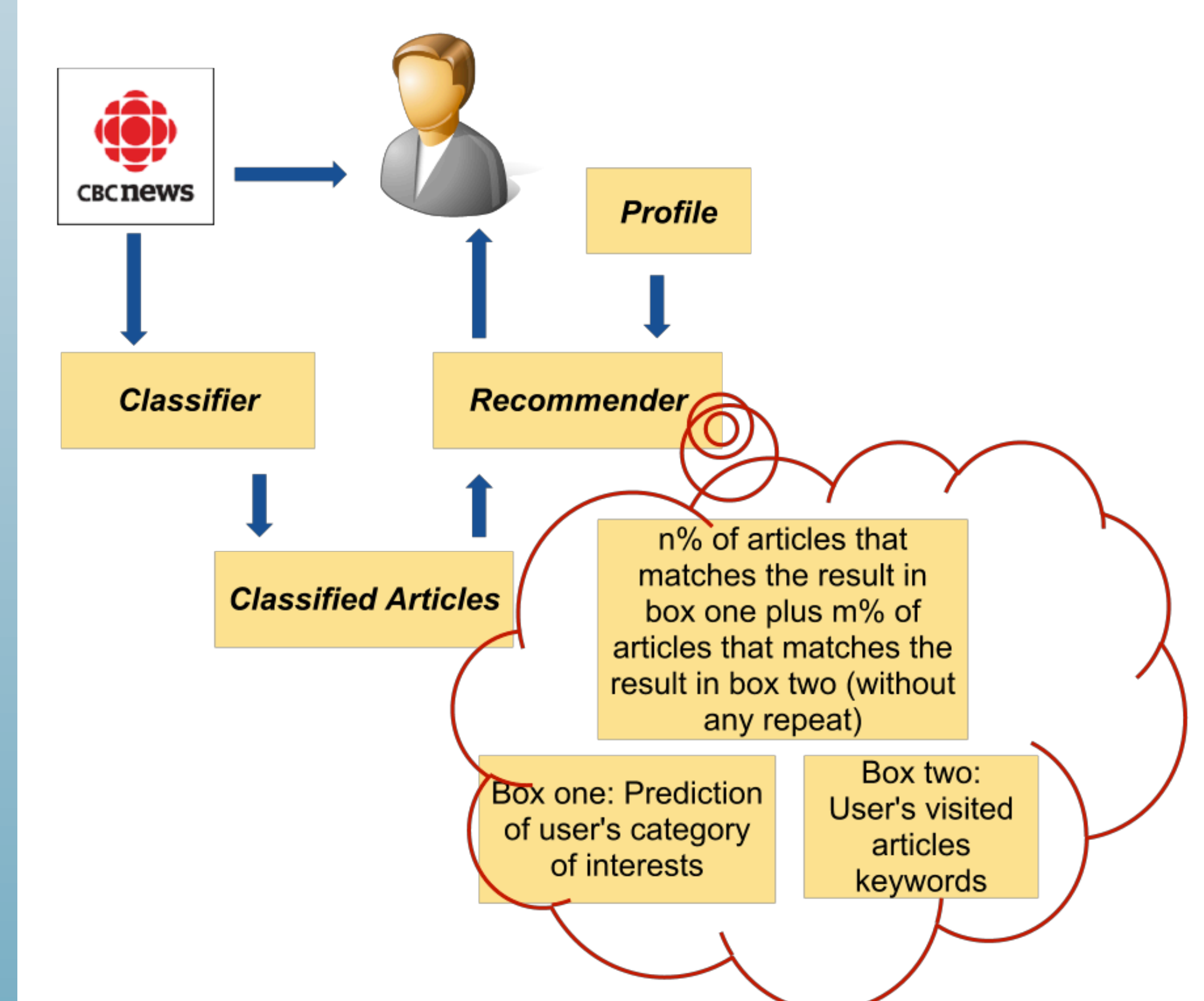


- we used another concept called keyword (a set of most repeated words in an article) to ranking websites that have been seen by the user and used this ranking to estimate user interest and recommend articles to her/him as well.
- We combine the results of these different methods to recommend new articles to users. Our motivation for combing these methods was our user interests crossover

System Architecture Diagram

In the following figure the system architecture is demonstrated. News stories are classified and then the classified articles are stored. Later recommender system used the classified articles and user profile to recommend new articles to the user.

Note that keywords are stored with classified articles as well and our recommender system uses both keywords and articles categories (using to different algorithm) to estimate user interests.



Conclusion

Recommender system is wide area of research and still needs lots of improvement. Regardless of the algorithm that we use to estimate user interests, inputs of the algorithm play an important role in the results. Item classification affect the accuracy of results in recommender systems. Different techniques can be used to improve item classification. Using keywords to define items is one of them. Also classifying articles from different perspective can give us more options.