

# Typicality Based Content-Boosted Collaborative Filtering Recommendation Framework

<sup>1</sup>N.SATISH KUMAR, <sup>2</sup>SUJAN BABU VADDE

<sup>1</sup>M.Tech (CSE), Department of Computer Science & Engineering, NRI Institute of Technology

<sup>2</sup>Assistant Professor, Department of Computer Science & Engineering, NRI Institute of Technology

**Abstract:** - Collaborative filtering (CF) is significant and admired technology for recommender systems. Recommender frameworks have been turned out to be significant means for web online clients to adapt to the information overload and have ended up a standout amongst the most effective and prevalent tools in electronic commerce. Recommending and personalization are critical ways to deal with combating information overload. Machine Learning is an imperative piece of frameworks for these assignments. Collaborative filtering has issues, substance based routines address these issues integrating both is best.

**Keywords:** Collaborative Filtering, Content-based Recommender System, Neighbor Selection.

## 1 INTRODUCTION

Recommendation systems found their application in the field of e-commerce and internet where items suggest to a group of user on the basis of their requirement based on their area of interest. A recommendation system is an information filtering system that built a model from the characteristic of an item according to the rating or prediction, given by a user to an item. Recommendation system has an important component in social media sites (such as Amazon, IMDB, Movie Lens), social sites giants such as Amazon have been greatly gained from the capability of their recommenders in accurately delivering the correct item to the correct user [17]. Collaborative filtering (CF) is an important and popular technology for recommender system. CF methods are classified into user based CF and item-based CF. The basic idea of user-based CF approach is to find out a set of users who have similar favor patterns or interest to a given user and the basic idea of item-based CF approach is to find out a set of items having highest correlation with the given item. In reality, people may like to group items into categories, and for each category there is a corresponding group of people who like items in the

category [18]. Cognitive psychologists find that objects (items) have different typicality degrees in categories in real life [19], [20], [21]. But these collaborative filtering methods have facing some problems.

## II. SYSTEM ANALYSIS:

### 2.1. Recommender Systems

Frameworks for recommending items (e.g. books, films, CD's, site pages, newsgroup messages) to clients taking into account cases of their inclinations. Numerous on-line stores give suggestions (e.g. Amazon, CDNow).Recommenders have been appeared to considerably increase deals at on-line stores. There are two essential ways to deal with recommending: Collaborative Filtering (a.k.a. social filtering), Substance based

### Book Recommender

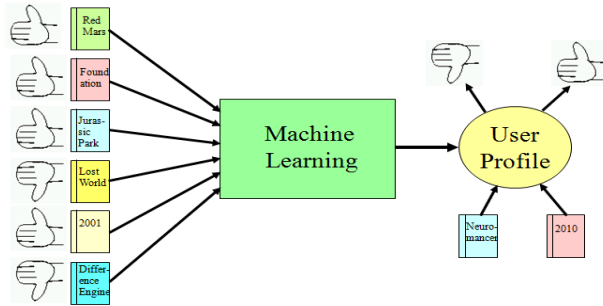


Fig 1. Book recommender architecture

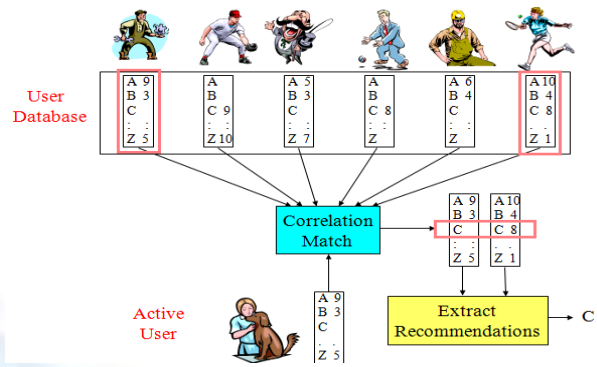


Fig 2. Collaborative Filtering

**Personalization**

- Recommenders are occurrence of personalization software.
- Personalization distresses get used to to the being needs, interests, and favorite of each user.
- Includes:
  - Recommending
  - Filtering
  - Predicting (e.g. form or calendar appt. completion)

From a business perspective, it is viewed as part of Customer Relationship Management (CRM).

**Machine Learning and Personalization**

- Machine Learning can allow learning a user model or profile of a particular user based on:
    - Sample interaction
    - Rated examples
  - This model or profile can then be used to:
    - Recommend items
    - Filter information
- Predict behavior

**2.2. Collaborative Filtering**

Maintain a database of numerous clients' ratings of an assortment of items. For a given client, find other comparative clients whose ratings unequivocally associate with the current user. Recommend items appraised vary by these comparative clients, however not evaluated by the current user. Almost all existing commercial recommenders utilize this methodology (e.g. Amazon).

- Weight all clients as for comparability with the dynamic user. Select a subset of the clients (neighbors) to use as predictors. Normalize ratings and process an expectation from a weighted combination of the selected neighbors' ratings. Present items with most noteworthy anticipated ratings as proposals.

Typically use Pearson correlation coefficient between ratings for active user, *a*, and another user, *u*.

$$C_{a,u} = \frac{\text{covar}(r_a, r_u)}{\sigma_{r_a} \sigma_{r_u}}$$

*r<sub>a</sub>* and *r<sub>u</sub>* are the ratings vectors for the *m* items rated by

**both** *a* and *u*

*r<sub>ij</sub>* is user *i*'s rating for item *j*

**Covariance and Standard Deviation**

**Covariance:**

$$\text{covar}(r_a, r_u) = \frac{\sum_{i=1}^m (r_{a,i} - \bar{r}_a)(r_{u,i} - \bar{r}_u)}{m}$$

$$\bar{r}_x = \frac{\sum_{i=1}^m r_{x,i}}{m}$$

**Standard Deviation:**

$$\sigma_{r_x} = \sqrt{\frac{\sum_{i=1}^m (r_{x,i} - \bar{r}_x)^2}{m}}$$

**2.3. Neighbor Selection**

For a given dynamic client,  $a$ , select connected clients to serve as the wellspring of predictions. The standard methodology is to utilize the most comparative  $n$  clients,  $u$ , taking into account closeness weights,  $w_{a,u}$ . Exchange methodology is to include all clients whose similitude weight is over a given limit.

- For a given dynamic client,  $a$ , select related clients to serve as the wellspring of forecasts.
- Standard approach is to use the most similar  $n$  users,  $u$ , based on similarity weights,  $w_{a,u}$
- Alternate approach is to include all users whose similarity weight is above a given threshold.

**Rating Prediction**

- Predict a rating,  $p_{a,i}$ , for each item  $i$ , for active user,  $a$ , by using the  $n$  selected neighbor users,  $u \in \{1,2,\dots,n\}$ .
- To represent clients diverse ratings levels, base expectations on contrasts from a client's normal rating.
- Weight clients' ratings commitment by their closeness to the active user.

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n w_{a,u} (r_{u,i} - \bar{r}_u)}{\sum_{u=1}^n w_{a,u}}$$

**3. CONTENT-BOOSTED COLLABORATIVE FILTERING**

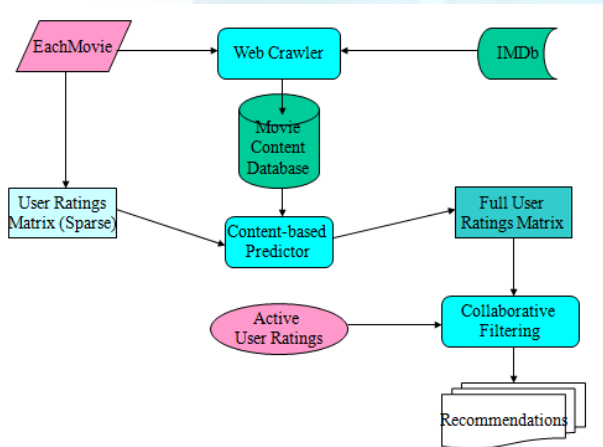


Fig 3. Content-boosted collaborative filtering architecture

**Content-Boosted CF – I**

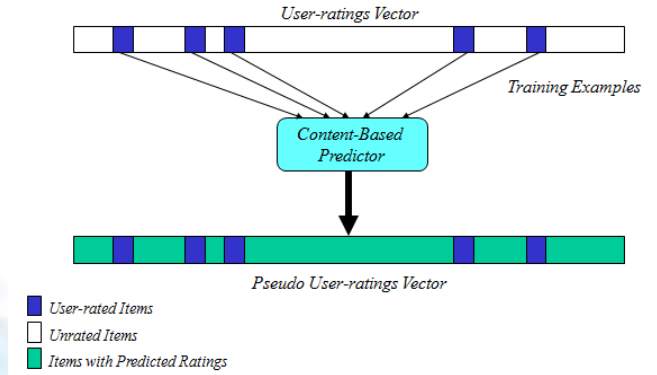


Fig 4. Content-Boosted CF – I

**Problems with Collaborative Filtering**

**Cold Start:** There should be sufficient different clients as of now in the framework to find a match.

**Sparsity:** If there are numerous items to be suggested, regardless of the fact that there are numerous clients, the client/ratings lattice is inadequate, and it is elusive clients that have appraised the same items.

**First Rater:** Can't suggest a thing that has not been already appraised.

**New items**

- Esoteric items

**Popularity Bias:** Can't prescribe items to somebody with special tastes.

Tends to recommend popular items.

**3.1. Content-Based Recommending**

Proposals depend on information on the substance of items as opposed to on other clients' opinions. Utilizes a machine learning calculation to induce a profile of the client's inclinations from samples taking into account a featural depiction of the substance.

Some previous applications:

Newsweeder (Lang, 1995), Syskill and Webert (Pazzani et al., 1996)

**Advantages of Content-Based Approach**

- No need for data on other users.
- No cold-start or sparsity problems.
- Able to recommend to users with unique tastes.

- Able to recommend new and unpopular items
- No first-rater problem.
- Can provide explanations of recommended items by listing content-features that caused an item to be recommended.

#### Disadvantages of Content-Based Method

Requires content that can be encoded as meaningful features. Users' tastes must spoke to as a learnable capacity of this substance features. Unable to adventure quality judgments of different users. Unless these are some way or another included in the substance highlights.

### 4. LIBRA (LEARNING INTELLIGENT BOOK RECOMMENDING AGENT)

Content-based recommender for books using information about titles extracted from Amazon.

Uses information extraction from the web to organize text into fields:

- Author
- Title
- Editorial Reviews
- Customer Comments
- Subject terms
- Related authors
- Related titles

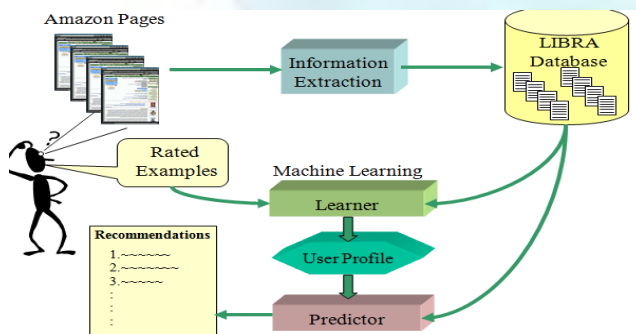


Fig5. LIBRA System Architecture

#### Sample Extracted Information

Title: <The Age of Spiritual Machines: When Computers Exceed Human Intelligence>

Author: <Ray Kurzweil>

Price: <11.96>

Publication Date: <January 2000>

ISBN: <0140282025>

Related Titles: <Title: <Robot: Mere Machine or Transcendent Mind>

Author: <Hans Moravec> >

...

Reviews: <Author: <Amazon.com Reviews> Text: <How much do we humans...>>

...

Comments: <Stars: <4> Author: <Stephen A. Haines> Text:<Kurzweil has ...>>

...

Related Authors: <Hans P. Moravec> <K. Eric Drexler>...

Subjects: <Science/Mathematics> <Computers> <Artificial Intelligence> ...

#### Libra Content Information

- Libra uses this extracted information to form “bags of words” for the following slots:
  - Author
  - Title
  - Description (reviews and comments)
  - Subjects
  - Related Titles
  - Related Authors
- User rates selected titles on a 1 to 10 scale.
- Libra uses a naïve Bayesian text-categorization algorithm to learn a profile from these rated examples.
  - Rating 6–10: Positive
  - Rating 1–5: Negative
- The learned profile is used to rank all other books as recommendations based on the computed posterior probability that they are positive.
- User can also provide explicit positive/negative keywords, which are used as priors to bias the role of these features in categorization.

### 5. SYSTEM IMPLEMENTATION

Stop words removed from all bags. A book's title and author are added to its own related title and related author slots. All probabilities are smoothed using Laplace estimation to account for small sample size. Lisp implementation is quite efficient: Training: 20 exs in 0.4 secs, 840 ex. in 11.5 secs, Test: 200 books per second

### Explanations of Profiles and Recommendations

Feature strength of word  $w_k$  appearing in a slot  $s_j$  :

$$\text{strength}(w_k, s_j) = \log \frac{P(w_k | \text{positive}, s_j)}{P(w_k | \text{negative}, s_j)}$$

## 6. EXPERIMENTAL DATA

Amazon searches were used to find books in various genres.

Titles that have at least one review or comment were kept.

Data sets:

Literature fiction:	3,061 titles
Mystery:	7,285 titles
Science:	3,813 titles
Science Fiction:	3,813 titles

### 6.1. Rated Data

4 users rated random examples within a genre by reviewing the Amazon pages about the title:

LIT1	936 titles
LIT2	935 titles
MYST	500 titles
SCI	500 titles
SF	500 titles

### 6.2. Method

10-fold cross-validation to generate learning curves.

Measured several metrics on independent test data:

**Precision at top 3:** % of the top 3 that is positive

**Rating of top 3:** Average rating assigned to top 3

**Rank Correlation:** Spearman's  $r_s$ , between system's and users complete rankings.

Test ablation of related author and related title slots (LIBRA-NR).

Test influence of information generated by Amazon's Collaborative approach.

## 7. CONCLUSION

In this paper we investigate on Collaborative filtering (CF). Recommending and personalization are

imperative ways to deal with combating information overload. Machine Learning is a vital piece of frameworks for these assignments. Community oriented filtering has issues. Substance based techniques address these issues (yet have issues of their own). Integrating both is best.

## REFERENCES

- [1] Z. Huang, H. Chen, and D. Zeng, "Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering," ACM Trans. Information Systems, vol. 22, no. 1, pp. 116- 142, 2004.
- [2] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 6, pp. 734-749, June 2005.
- [3] K.M. Galotti, Cognitive Psychology In and Out of the Laboratory, third ed. Wadsworth, 2004.
- [4] G.L. Murphy, The Big Book of Concepts. MIT Press, 2002.
- [5] L.W. Barsalou, Cognitive Psychology: An Overview for Cognitive Scientists. Lawrence Erlbaum Assoc., 1992.
- [6] S. Schiffer and S. Steele, Cognition and Representation. Westview Press, 1988.
- [7] D.L. Medin and E.E. Smith, "Concepts and Concept Formation," Ann. Rev. of Psychology, vol. 35, pp. 113-138, 1984.
- [8] W. Vanpaemel, G. Storms, and B. Ons, "A Varying Abstraction Model for Categorization," Proc. Cognitive Science Conf. (CogSci'05), pp. 2277-2282, 2005.