



Improving Recommendation Lists Through Topic Diversification

April 06, 2005



C.-N. Ziegler¹, S. M. McNee², J. A. Konstan², G. Lausen¹


¹Institute of Computer Science (DBIS),
University of Freiburg
Germany

²GroupLens Research,
Univ. of Minnesota
MN, USA


Outline



- Introduction and Motivation
- Topic Diversification
- Empirical Evaluation
- Outlook and Conclusion

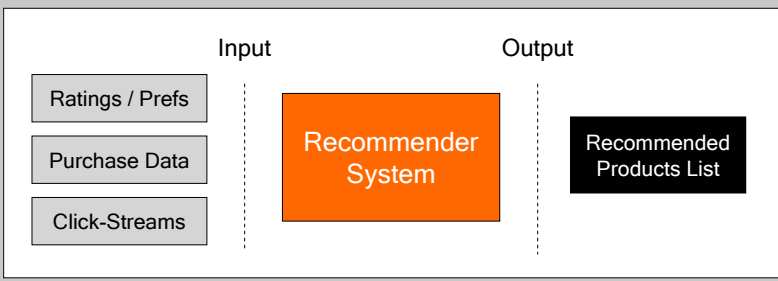


Introduction and Motivation



Motivation

- **Recommender Systems**
 - Suggest products that users are likely to appreciate
 - Personalized recommendations with respect to interest



```
graph LR; subgraph Input; direction TB; I1[Ratings / Prefs]; I2[Purchase Data]; I3[Click-Streams]; end; subgraph Output; direction TB; O1[Recommended Products List]; end; RS[Recommender System]; Input -.-> RS; RS -.-> Output;
```

- **Problem**
 - In many cases, recommended lists lack diversity!

Diversity Issue Example (Item-based CF)

- Example: Amazon.com (own recommendations)

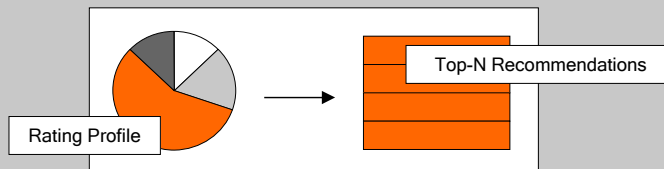
The screenshot shows a list of book recommendations from Amazon.com, all related to the Tolkien universe. The books listed are:

- Sauron Defeated
- The War of the Ring
- Treason of Isengard
- The Ring of Middle-Earth

Below the list, a text box states: "All books taken from the Tolkien universe".

Diversity Issue Analysis

- «Winners Take All» behavior (over-fitting)
 - Inherent to certain recommender system types
 - Item-based CF (portfolio effect [Ali and van Stam, 2004])



- Effect largely unnoticed by accuracy metrics
 - Typical **off-line** metrics
 - Precision/Recall
 - Mean Absolute Error (MAE)
 - Low diversity does **not** exclude high accuracy

Contributions

1. « Topic Diversification » method

- Make lists reflect the whole range of interests
- Minimize accuracy trade-off for diversification
 - Intra-list similarity metric to measure diversity

2. Empirical evidence on user satisfaction

- Hypothesis:
 - Diversity **increases** user satisfaction

$$\text{Satisfaction} = \text{Accuracy} + X,$$

where

$$X \subseteq \{ \text{Diversity, Novelty, Serendipity, ...} \}$$



Topic Diversification

Topic Diversification

- **Idea**

- Maximize overall recommendation list utility

- Represent all major topics of the user's interest
- Proportional weighting and representation

→ Similarity metric $c_b(b_i, b_j)$ for products b_i, b_j required

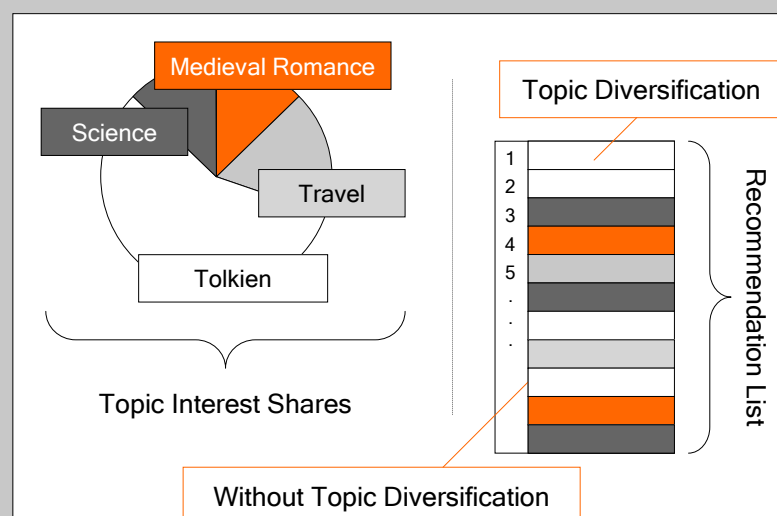
- **Implementation**

- Applicable as « Add-On » for recommenders

- **Post-processing** of recommendation lists
- Original recommendation list (size $Y \times N$) as input
- Diversified list (size N) as output

- Very suitable for **cross-domain** recommending

Solving the Overfitting Issue



Algorithm Outline

- Requirements

- Input list has **more than top-N** products ($Y \times N$)

- Experimental analysis with $5 \times N$ item slots
- «Hidden» products not shown to the user

- Similarity metric $c_b : B \times B \rightarrow [0, 1]$

- Similarity of product sets
- For our experiments:
 - Taxonomy-driven metric [Ziegler et al., 2004]

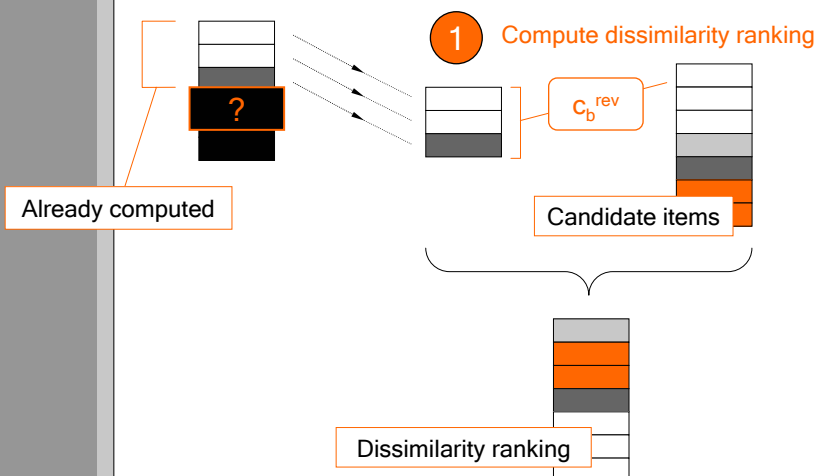
- Dissimilarity measure $c_b^{rev} : B \times B \rightarrow [0, 1]$

- Similarity opposed to diversity

→ $c_b^{rev} := 1 - c_b$

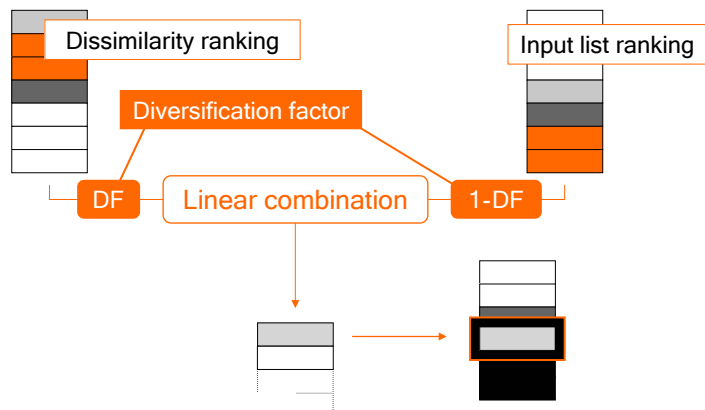
Top N

Algorithm Outline (II)



Algorithm Outline (III)

2 Compute next output list entry



Empirical Evaluation

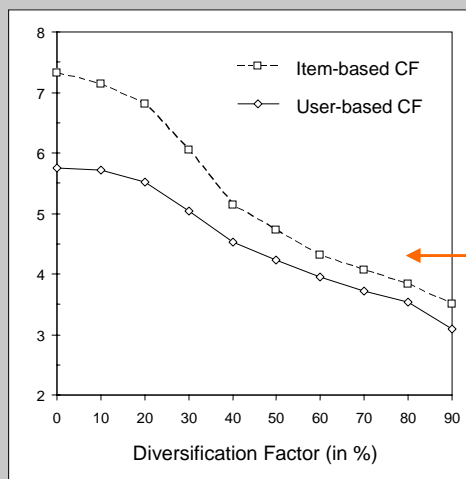


Experiment Setup

- **Dataset makeup**
 - Mined from « **BookCrossing.com** »
 - 278,858 members, 271,379 books, 1,157,379 ratings
 - Background knowledge from « **Amazon.com** »
 - Content-based classifiers for each book
- **Recommender algorithms**
 - **Item-based** } Collaborative Filtering
 - **User-based** }
- **Experiments**
 - Off-line accuracy metrics (**precision / recall**)
 - On-line user surveys

Hypothesis: Diversity increases user satisfaction

Off-Line Analysis



Precision/Recall Analysis

- Increasing diversification factor
 - $DF \in [0.1, 0.9]$
- Item-based / User-based CF

Recall

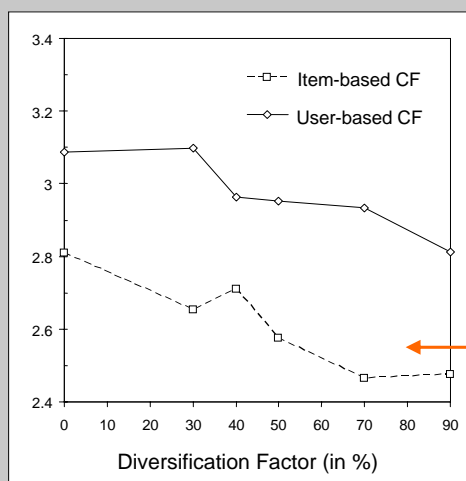
Result

Topic Div. **decreases** accuracy

On-Line Survey Study

- Survey
 - Individual recommendations for each member computed
 - 2,125 Book Crossing members completed the survey
- Design
 - Inter-subject study
 - 10 recommendations per survey
 - User-based / Item-based, varying DFs
 - Rating for each recommendation demanded
 - Questionnaire
 - Overall list value
 - Perceived diversity
 - ...

On-Line Analysis

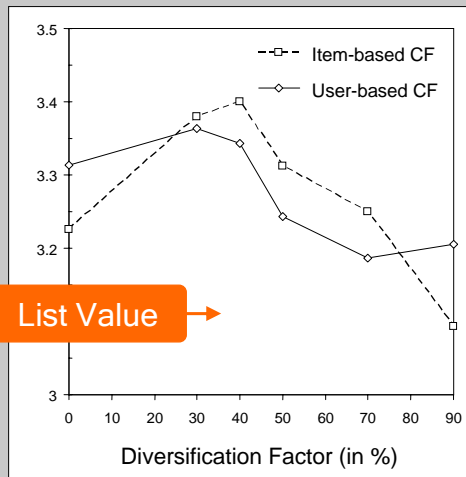


Single-Vote Averages

- For each subject
 - \emptyset of max. 10 votes
- **Disregards** list aspect
- In line with **precision / recall**
 - similar downward trend

Single-Vote Averages

On-Line Analysis (II)



List Value →

Overall List Value

- User-based CF
 - No significance for inc. DF
 - DF's effect is **negligible**
- Item-based CF
 - Peak at DF \in [0.3, 0.4]
 - DF \geq 0.5 appears too much

Result

TDiv **may increase** satisfaction

Conclusion and Outlook

- Off-line versus on-line analysis
 - Off-line accuracy **deteriorates** for increasing DF
 - «Real» satisfaction with lists behaves differently
 - User-based CF largely insensitive to increasing DF
 - Item-based CF substantially benefits for DF \in [0.3, 0.4]

Satisfaction is more than pure accuracy (empirically shown)

- Future directions
 - Apply Topic Diversification to other domains
 - Sequential consumption scenario (Launch, Internet Radio ...)