Hybrid algorithms for recommending new items

ROBERTO TURRIN – Moviri, R&D
Paolo Cremonesi – Politecnico di Milano
Fabio Airoldi – Moviri, R&D

http://dx.doi.org/10.1145/2039320.2039325
in a nutshell

- Hybrid algorithms
- Real domain requirements
  - scalability
  - modularity
  - many unrated items
- New-item stressing experiments
- Datasets
  - Private TV dataset
  - MovieLens

Credits: http://dpaki.com/?p=2591
Traditional recommender systems

**Collaborative (CF)**
- **Pros**
  - High quality
- **Cons**
  - New items problem (since they do not have ratings)
  - Popularity bias

**Content-based (CBF)**
- **Pros**
  - Work on new items
- **Cons**
  - Low quality (since user ratings are ignored)
  - Profile overfitting

---

R. TURRIN, P. Cremonesi, F. Airoldi - Hybrid algorithms for recommending new items
..so CF or CBF? ..many variables

quality

new system

?  

mature system

CF

CBF

time

R. TURRIN, P. Cremonesi, F. Airoldi - Hybrid algorithms for recommending new items
• The EPG is characterized by many unrated, new TV programs
• The percentage of new-item cannot be neglected
Several hybrid algorithms mix CF and CBF (but also demographics, social)

e.g.:


**Pros**
- Some approaches show better quality than CF/CBF

**Cons**
- Low scalability / no real-time recommendations
- Only partial focus on new-item problem
- Not working with implicit, binary ratings
GOALS
- New-item
- Quality comparable to collaborative

REQUIREMENTS:
- Batch/real-time scalability/complexity
  - Updated recommendations
- Modularity: ability to re-use existing CF and CBF algorithms.
- Implicit/explicit ratings
Main contributions

- **GOALS**
  - New-item
  - Quality comparable to collaborative

- **REQUIREMENTS:**
  - Batch/real-time **scalability**/complexity
    - Updated recommendations
  - **Modularity**: ability to re-use existing CF and CBF algorithms.
  - Implicit/explicit ratings

- Two hybrid algorithms:
  - extension of SimComb algorithm
  - introduction of a new hybrid algorithm

- New-item stressing evaluation

R. TURRIN, P. Cremonesi, F. Airoldi - Hybrid algorithms for recommending new items
STATE-OF-THE-ART RECOMMENDER ALGORITHMS
Implemented strategies:

- **Item-item neighborhood-based (NNCos)**
  - Recommendations are based on item-item similarities computed as the *cosine metric*

- **Latent factor models (PureSVD)**
  - Recommendations are based on hidden factors implicitly discovered by means of a matrix factorization (SVD)
Content-based algorithm

Weight of feature $f$ in item $i$.
- Computed as TF-IDF
- Example of features: genre, actors, directors,...

Item-content matrix (ICM)

LSA (Latent Semantic Analysis)

The ICM is factorized by means of SVD in order to discover latent semantic
**Hybrid algorithms**

**Interleaved (INTL)**
- Trivial hybrid implementation where the final recommendation list is formed by alternating items recommended by the CF algorithm with items recommended by the CBF algorithm.

**SimComb [Mobasher et al. 2004]**
- Two item-item similarity matrices are computed and linearly combined:

\[
(1-\alpha) \text{ CF item-item similarities} + \alpha \text{ CBF item-item similarities} = \text{ HYBRID item-item similarities}
\]
PROPOSED HYBRID ALGORITHMS

• FFA (Filtered Feature Augmentation)
• SIMinjKnn (Similarity Injection Knn)
Collaborative filtering as main brick

We trust CF recommendations when the model has been trained with “enough” information (i.e., ratings).

We add CBF-based data (i.e., rating) for better training the CF when no enough information is available.

R. TURRIN, P. Cremonesi, F. Airoldi - Hybrid algorithms for recommending new items
Collaborative filtering as main brick

We trust CF recommendations when the model has been trained with “enough” information (i.e., ratings).

We add CBF-based data (i.e., features) for better training the CF when no enough information is available.

R. TURRIN, P. Cremonesi, F. Airoldi - Hybrid algorithms for recommending new items
A number of recommendation (CF and CBF) algorithms allow to compute item-item similarity.
Item-item model: real-time recommendations

R. TURRIN, P. Cremonesi, F. Airoldi - Hybrid algorithms for recommending new items
Item-item model: real-time recommendations

Real-time requirements:
- **Memory**: $K \times \#\text{items}$
- **Time**: $f(\#\text{ratings}, K) \times \#\text{items}$
- Use of existing algorithms
- Updated recommendations
- Implicit/explicit ratings

User ratings
Filtered Feature Augmentation (FFA)

**Idea:** add *pseudo-ratings* to the item profiles

**Motivation**
- *Pseudo-ratings* model new items
- *Less sparse* item-profiles

---

R. TURRIN, P. Cremonesi, F. Airoldi - Hybrid algorithms for recommending new items
Filtered Feature Augmentation (FFA)

Idea: add pseudo-ratings to the item profiles

Motivation
- Pseudo-ratings model new items
- Less sparse item-profiles

Entropy-based filtering (e.g., Gini impurity measure)

predicted ratings

CONTENT → CBF → Filter → ratings → CF → Model

R. TURRIN, P. Cremonesi, F. Airoldi - Hybrid algorithms for recommending new items
**Motivation**
- Discovering relationships between new and old items

**Idea:** mixing CF and CBF similarities
EVALUATION
Datasets

- **1M Movielens**
  - ~6K users, ~3.9K items, 1M ratings

- An implicit, binary dataset collected from 15’000 IPTV users over a period of six months
  - ~15K users, ~800 rated items/~4K, ~26K ratings
  - Multilanguage (mainly German, French) content data

available at http://home.dei.polimi.it/cremones/memo/downloads/TV2.zip
Testing methodology (1)

- $H_1$: set of **existing** items
- $H_2$: set of **new** items
- **Training set** (extracted from $H_1$)
  - $(100-\beta)$% **existing** items: extracted from $H_1$
  - $\beta$% **new** items: extracted from $H_2$
- Discarded ratings

R. TURRIN, P. Cremonesi, F. Airoldi - Hybrid algorithms for recommending new items
Testing methodology (1)

- $H_1$: set of **existing** items
- $H_2$: set of **new** items

- Training set (extracted from $H_1$)
- Test set
  - $(100-\beta)\%$ **existing** items: extracted from $H_1$
  - $\beta\%$ **new** items: extracted from $H_2$

- Discarded ratings

R. TURRIN, P. Cremonesi, F. Airoldi - Hybrid algorithms for recommending new items
Testing methodology (2)

- For each <user, item> <u,i> in H_{1+2}:
  - Generate rating prediction for i
  - Generate rating prediction for every other items
  - Sort the items according to predicted rating
- There is a “hit” if rank(i) < N
  - i.e., item i appears in the top-N.
  In our tests, N=20

R. TURRIN, P. Cremonesi, F. Airoldi - Hybrid algorithms for recommending new items
Non-hybrid algorithms

![Graphs comparing recall when recommending 20 items against the percentage of new items.](#)

R. TURRIN, P. Cremonesi, F. Airoldi - Hybrid algorithms for recommending new items
Hybrid algorithms: ML

Recall when recommending 20 items

- LSA
- NNCOS
- INTL: LSA+NNCOS
- SIMCOMB: LSA+NNCOS

Percentage of new items $\beta$
Hybrid algorithms: ML

- LSA
- NNCOS
- INTL: LSA+NNCOS
- SIMCOMB: LSA+NNCOS
- FFaG: LSA+NNCOS
- SIMINJknn: LSA+NNCOS

Recall when recommending 20 items

Percentage of new items $\beta$
Hybrid algorithms: TV

Recall when recommending 20 items vs. Percentage of new items $\beta$

- LSA
- NNCOS
- INTL: LSA+NNCOS
- SIMCOMB: LSA+NNCOS
- FFA: LSA+NNCOS
- SIMINJknn: LSA+NNCOS

Recall when recommending 20 items vs. Percentage of new items $\beta$

- LSA
- NNCOS
- INTL: LSA+NNCOS
- SIMCOMB: LSA+NNCOS
- FFA: LSA+NNCOS
- SIMINJknn: LSA+NNCOS
Toy sample

R. TURRIN, P. Cremonesi, F. Airoldi - Hybrid algorithms for recommending new items
Conclusions / Future work

- Proposed 2 hybrid algorithms:
  - Higher recall than CF and CBF in the presence of new items
  - Scalable / non-affecting real-time performance
  - Handling implicit/explicit ratings

- Future work:
  - Subjective evaluation
  - Improving the filter with other information
  - Other domains
Thank you

Roberto TURRIN, PhD
Moviri, R&D – Italy
roberto.turrin@moviri.com

Q&A