Multi-purpose Library of Recommender System Algorithms for the Item Prediction Task

Presentation of my Bachelor Thesis

Julius Kolbe

Fakultät für Elektrotechnik und Informatik
Institut für Verteilte Systeme

June 25, 2013

Contents

1. Background
   - Item Prediction Task and Implicit Feedback
   - Evaluation
   - Evaluation Metrics
2. Recommendation Algorithms
3. Recsyslab

Contents

1. Background
   - Item Prediction Task and Implicit Feedback
   - Evaluation
   - Evaluation Metrics
2. Recommendation Algorithms
3. Recsyslab

Implicit Feedback

	Anna	Berta	Claudia	Dagmar
The Shawshank Redemption	1		1	
The Godfather		1	1	
The Godfather: Part II		1		1
Pulp Fiction	1	1		1
The Good, the Bad and the Ugly	1		1	

Item Prediction Task

	Anna	Berta	Claudia	Dagmar
The Shawshank Redemption	1		1	?
The Godfather		1	1	?
The Godfather: Part II		1		1
Pulp Fiction	1	1		1
The Good, the Bad and the Ugly	1		1	?

Notation

	Anna	Berta	Claudia	Dagmar
The Shawshank Redemption	1		1	
The Godfather		1	1	
The Godfather: Part II		1		1
Pulp Fiction	1	1		1
The Good, the Bad and the Ugly	1		1	

Items
Users
Interactions
Basket of u

Leave-one-out Protocol

1. Randomly choose one interaction per user and hide them
2. Train the recommender system with the remaining interactions
3. Get recommendations for every user
4. Compute the chosen evaluation metric with the hidden items and the recommendations

Hitrate/Recall@N [Karypis(2001), Sarwar et al.(2000)Sarwar, Karypis, Konstan, and Riedl]

$$\text{Recall@N} = \frac{\sum_{u \in U} H_u \cap \text{topN}_u}{|H|} \quad (1)$$

H hidden interactions

H_u the hidden interaction of u

U set of users

$topN_u$ N recommendations for u

Precision [Sarwar et al.(2000)Sarwar, Karypis, Konstan, and

$$\text{Precision} = \frac{\sum_{u \in U} H_u \cap \text{topN}_u}{N \times |U|} \tag{2}$$

H hidden interactions

H_u the hidden interaction of u

U set of users

$topN_u$ N recommendations for u

F1 [Sarwar et al.(2000)Sarwar, Karypis, Konstan, and Riedl]

$$\text{F1} = \frac{2 \times \text{Recall@N} \times \text{Precision}}{\text{Recall@N} + \text{Precision}}. \quad (3)$$

H hidden interactions

H_u the hidden interaction of u

U set of users

$topN_u$ N recommendations for u

Mean Reciprocal Hitrate [Ning and Karypis(2011)]

$$\mathrm{MRHR} = \frac{1}{|U|} \sum_{u \in U} \frac{1}{\mathrm{pos}(\mathrm{topN}_u, H_u)}, \tag{4}$$

H hidden interactions

H_u the hidden interaction of u

U set of users

$topN_u$ N recommendations for u

$\mathrm{pos}(\mathrm{topN}_u, H_u)$ position of the hidden item in the list of recommendations

Area under the ROC (AUC) [Rendle et al.(2009)Rendle, Freudenthaler, Gantner,

$$\mathrm{AUC} = \frac{1}{|U|} \sum_{u \in U} \frac{1}{|E(u)|} \sum_{(i,j) \in E(u)} \delta(x_{ui} > x_{uj}), \tag{5}$$

$$\delta(x) = \begin{cases} 1, & \text{if x is true,} \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

$$E(u) = \{(i,j) | (u,i) \in H \wedge (u,j) \notin (H \cup T)\}. \tag{7}$$

H hidden interactions

H_u the hidden interaction of u

U set of users

x_{ui} predicted score of the interaction between u and i

Contents

1 Background
- Item Prediction Task and Implicit Feedback
- Evaluation
- Evaluation Metrics

2 Recommendation Algorithms

3 Recsyslab

Non-Personalized

Constant Recommend the most popular items

Random Recommend randomly chosen items

k-Nearest-Neighbor [Karypis(2001)]

1. Compute similarity of each item, item pair
2. For each item, save the k items with the highest similarity ($=$ neighbors)
3. Compute the union of the neighbors of the basket of u
4. For each item in this set compute the sum of similarities to the basket of u
5. Sort by this score and return the first N items

$$\text{sim}(i, j) = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{||\vec{i}||_2 ||\vec{j}||_2} \tag{8}$$

k-Nearest-Neighbor [Karypis(2001)]

	Anna	Berta	Claudia	Dagmar
The Shawshank Redemption	1	0	1	0
The Godfather	0	1	1	0
The Godfather: Part II	0	1	0	1
Pulp Fiction	1	1	0	1
The Good, the Bad and the Ugly	1	0	1	0

$$\mathrm{sim}(i,j) = \cos(\vec{i},\vec{j}) = \frac{\vec{i}\cdot\vec{j}}{||\vec{i}||_2||\vec{j}||_2} = \frac{0}{2} = 0$$

k-Nearest-Neighbor [Karypis(2001)]

	Anna	Berta	Claudia	Dagmar
The Shawshank Redemption	1	0	1	0
The Godfather	0	1	1	0
The Godfather: Part II	0	1	0	1
Pulp Fiction	1	1	0	1
The Good, the Bad and the Ugly	1	0	1	0

$$\text{sim}(i,j) = \cos(\vec{i},\vec{j}) = \frac{\vec{i}\cdot\vec{j}}{||\vec{i}||_2||\vec{j}||_2} = \frac{2}{\sqrt{2}\sqrt{3}}$$

Matrix Factorization [mat(2013)]

	Anna	Berta	Claudia	Dagmar
The Shawshank Redemption	1	0	1	0
The Godfather	0	1	1	0
The Godfather: Part II	0	1	0	1
Pulp Fiction	1	1	0	1
The Good, the Bad and the Ugly	1	0	1	0

Find W and H so: $\hat{M} = W\ H^{\top}$.

$$\text{Score}(u, i) = W_u I_i^{\top}. \tag{9}$$

Matrix Factorization, Training

```
U = randomly chosen user
I = randomly chosen item U interacted with
J = randomly chosen item U did not interact with

X=H[i] - H[j]
wx = dot product of W[u] and X
dloss = (derivative of the
        loss function of wx and 1) *
        learningRate

W[u] += dloss * (H[i] - H[j]) #These three lines
H[i] += dloss * W[u]          #have to be
H[j] += dloss * -W[u]         #executed at once
```

Contents

1 Background
- Item Prediction Task and Implicit Feedback
- Evaluation
- Evaluation Metrics

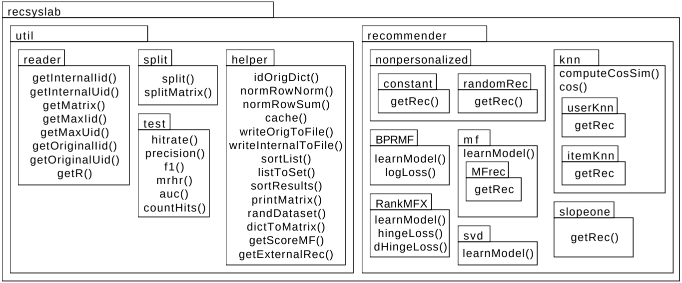
2 Recommendation Algorithms

3 Recsyslab

Introduction of recsyslab

- Python for easy readable source code
- simple usage
- for education
- for research
- open source license: GPLv3

General Structure



Get recsyslab

github.com/Foolius/recsyslab

github.com/Foolius/recsyslab/archive/master.zip

```
$ git clone
    https://github.com/Foolius/recsyslab.git
```

Matrix factorization, June 2013.
URL
http://en.wikipedia.org/wiki/Matrix_factorization.

George Karypis.
Evaluation of item-based top-n recommendation algorithms.
In *Proceedings of the tenth international conference on Information and knowledge management*, CIKM '01, pages 247–254, New York, NY, USA, 2001. ACM.
ISBN 1-58113-436-3.
doi: 10.1145/502585.502627.
URL http://doi.acm.org/10.1145/502585.502627.

Xia Ning and George Karypis.
Slim: Sparse linear methods for top-n recommender systems.
In Diane J. Cook, Jian Pei, Wei Wang, Osmar R. Zaïane, and Xindong Wu, editors, *ICDM*, pages 497–506. IEEE, 2011.
ISBN 978-0-7695-4408-3.

Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme.
Bpr: Bayesian personalized ranking from implicit feedback.
In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, pages 452–461, Arlington, Virginia, United States, 2009. AUAI Press.
ISBN 978-0-9749039-5-8.
URL
http://dl.acm.org/citation.cfm?id=1795114.1795167.

Badrul M. Sarwar, George Karypis, Joseph A. Konstan, and John T. Riedl.
Application of dimensionality reduction in recommender system – a case study.
In *IN ACM WEBKDD WORKSHOP*, 2000.