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 24, 2013

Contents

Contents

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 [1, 2]

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

H hidden interactions

H_u the hidden interaction of u

U set of users

$topN_u$ N recommendations for u

Precision [2]

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

H hidden interactions

H_u the hidden interaction of u

U set of users

$topN_u$ N recommendations for u

F1 [2]

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

- H hidden interactions
- H_u the hidden interaction of u
- U set of users
- $topN_u$ N recommendations for u

Mean Reciprocal Hitrate [3]

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

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) [4]

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

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

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

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

Non-Personalized

Constant Recommend the most popular items

Random Recommend randomly chosen items

k-Nearest-Neighbor [1]

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

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

k-Nearest-Neighbor [1]

	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

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

k-Nearest-Neighbor [1]

	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

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

Matrix Factorization

Slope One

Contents

Motivation

General Structure

G. Karypis, “Evaluation of item-based top-n recommendation algorithms,” in *Proceedings of the tenth international conference on Information and knowledge management*, CIKM '01, (New York, NY, USA), pp. 247–254, ACM, 2001.

B. M. Sarwar, G. Karypis, J. A. Konstan, and J. T. Riedl, “Application of dimensionality reduction in recommender system – a case study,” in *IN ACM WEBKDD WORKSHOP*, 2000.

X. Ning and G. Karypis, “Slim: Sparse linear methods for top-n recommender systems,” in *ICDM* (D. J. Cook, J. Pei, W. Wang, O. R. Zaïane, and X. Wu, eds.), pp. 497–506, IEEE, 2011.

S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, “Bpr: Bayesian personalized ranking from implicit feedback,” in *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, (Arlington, Virginia, United States), pp. 452–461, AUAI Press, 2009.