Hasso Plattner Institut

IT Systems Engineering | Universität Potsdam

Matrix Factorization Techniques For Recommender Systems

Collaborative Filtering

Markus Freitag, Jan-Felix Schwarz

28 April 2011

Agenda

Paper Backgrounds

MATRIX FACTORIZATION TECHNIQUES FOR RECOMMENDER SYSTEMS

Yehuda Koren, *Yahoo Research*
Robert Bell and Chris Volinsky, *AT&T Labs—Research*

As the Netflix Prize competition has demonstrated, matrix factorization models are superior to classic nearest-neighbor techniques for producing product recommendations, allowing the incorporation of additional information such as implicit feedback, temporal effects, and confidence levels.

Modern consumers are inundated with choices. Electronic retailers and content providers offer a huge selection of products, with unprecedented opportunities to meet a variety of special needs and tastes. Matching consumers with the most appropriate products is key to enhancing user satisfaction and loyalty. Therefore, more retailers have become interested in recommender systems, which analyze patterns of user interest in products to provide personalized recommendations that suit a user's taste. Because good personalized recommendations can add another dimension to the user experience, e-commerce leaders like Amazon.com and Netflix have made recommender systems a salient part of their websites.

Such systems are particularly useful for entertainment products such as movies, music, and TV shows. Many customers will view the same movie, and each customer is likely to view numerous different movies. Customers have proven willing to indicate their level of satisfaction with particular movies, so a huge volume of data is available about which movies appeal to which customers. Companies can analyze this data to recommend movies to particular customers.

RECOMMENDER SYSTEM STRATEGIES

Broadly speaking, recommender systems are based on one of two strategies. The *content filtering* approach creates a profile for each user or product to characterize its nature. For example, a movie profile could include attributes regarding its genre, the participating actors, its box office popularity, and so forth. User profiles might include demographic information or answers provided on a suitable questionnaire. The profiles allow programs to associate users with matching products. Of course, content-based strategies require gathering external information that might not be available or easy to collect.

A known successful realization of content filtering is the Music Genome Project, which is used for the Internet radio service Pandora.com. A trained music analyst scores

42 COMPUTER Published by the IEEE Computer Society 0018-9162/09/$26.00 © 2009 IEEE

Yehuda Koren, Yahoo Research

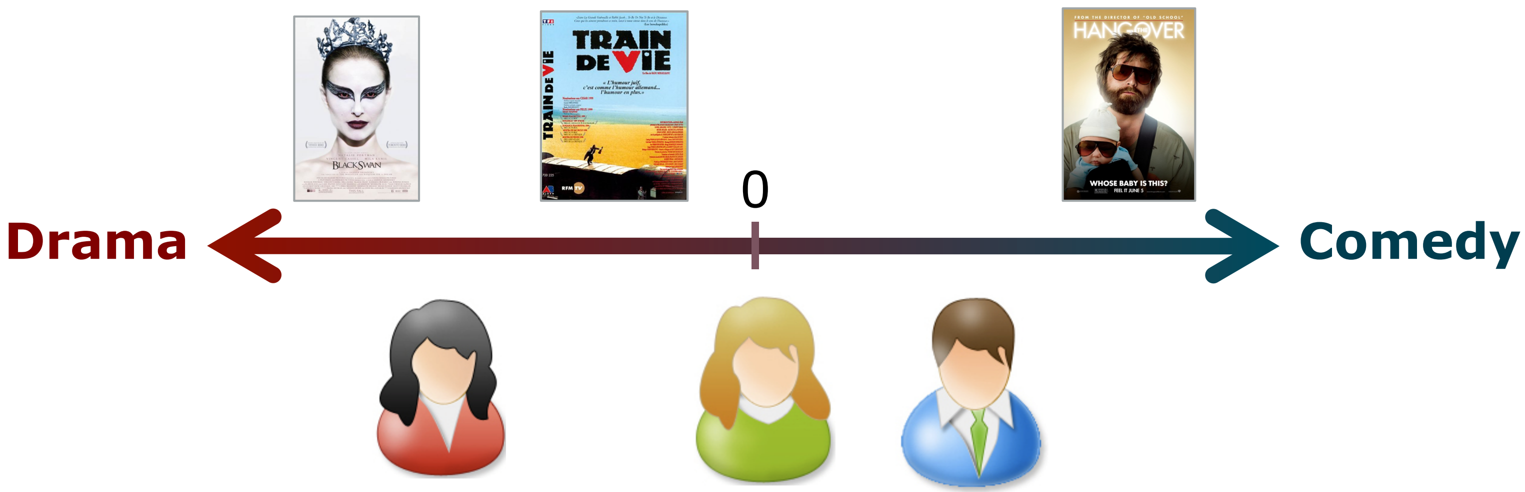
Robert Bell and Chris Volinsky, AT&T Labs-Research

Paper published in August 2009

Authors won the grand Netflix Prize in September 2009

Latent Factor Models

- Find features that describe the characteristics of rated objects
- Item characteristics and user preferences are described with numerical factor values
- Assumption: Ratings can be inferred from a model put together from a smaller number of parameters



Latent Factor Models

- Items and users are associated with a factor vector
- Dot product captures the user's estimated interest in the item:

$$\hat{r}_{ui} = q_i^T p_u$$

- Challenge: How to compute a mapping of items and users to factor vectors?

- Approaches:
  - Singular Value Decomposition (SVD)
  - Matrix Factorization

Singular Value Decomposition

Rating Matrix (N x M)

	5	3	5
	4	2	1
	0	3	3

User Feature Matrix (F x N)

f_1	1	-4	1
f_2	-2	0	-3
f_3	0	-5	1

Movie Feature Matrix (F x M)

f_1	-1	0	-2
f_2	4	-4	1
f_3	0	2	2

Singular Value Decomposition

User Feature Matrix (F x N) · Movie Feature Matrix (F x M) = Rating Matrix (N x M)

$$\left(\begin{array}{c|ccc} & & & \\ \hline f_1 & 1 & -4 & 1 \\ f_2 & -2 & 0 & -3 \\ f_3 & 0 & -5 & 1 \end{array}\right)^{T} \bullet \begin{array}{c|ccc} & & & \\ \hline f_1 & -1 & 0 & -2 \\ f_2 & 4 & -4 & 1 \\ f_3 & 0 & 2 & 2 \end{array} = \begin{array}{ccc} 5 & 3 & 5 \\ 4 & 2 & 1 \\ 0 & 3 & 3 \end{array}$$

SVD - Problems

- Conventional SVD is undefined for incomplete matrices!

- Imputation to fill in missing values
  - Increases the amount of data
  - "SVD of ginormous matrices is... well, no fun" (Simon Funk)

- We need an approach that can simply ignore missing ratings

Matrix Factorization

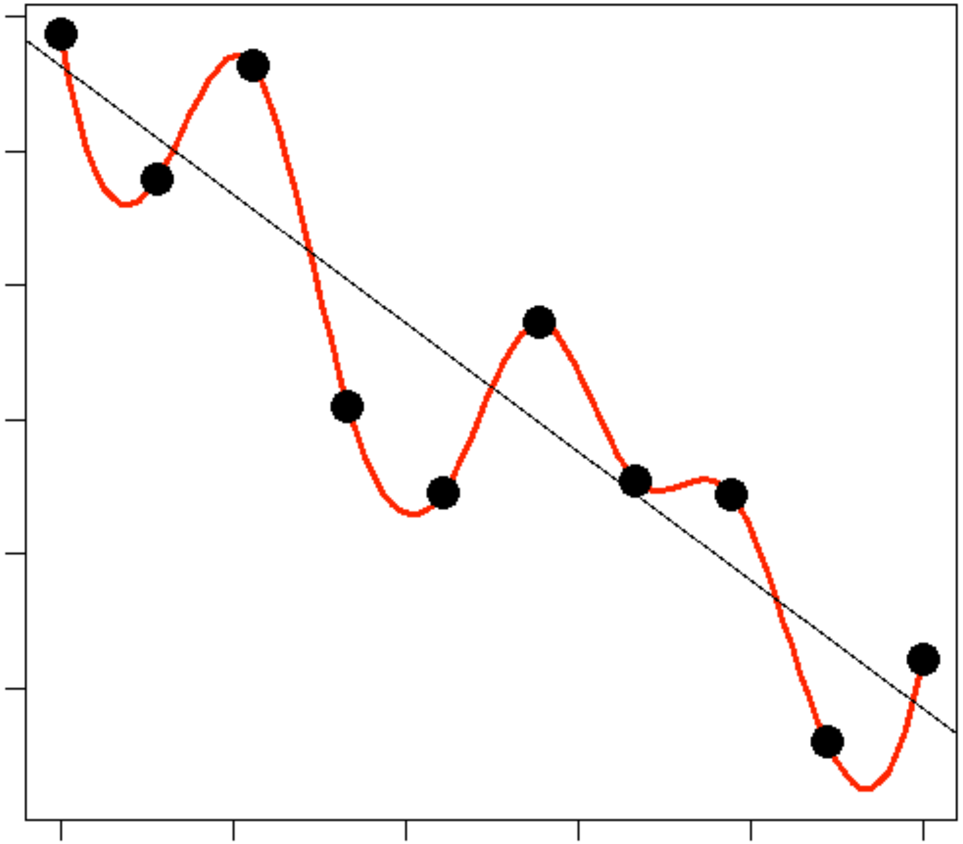
$$\min_{q^*,p^*} \sum_{(u,i)\in\kappa} (r_{ui} - q_i^T p_u)^2$$

r_{ui} : known rating of user u for item i

remember:
predicted rating $\hat{r}_{ui} = q_i^T p_u$

Matrix Factorization - Overfitting

A model is built to *represent* the training data – not to *reproduce* the training data.



Matrix Factorization - Regularization

Idea: penalize complexity

$$\min_{q^*,p^*} \sum_{(u,i)\in\kappa} (r_{ui} - q_i^T p_u)^2 + \lambda(\| q_i \|^2 + \| p_u \|^2)$$

λ : constant to control the extend of regularization

→ determined by cross-validation

Learning Algorithms

- Stochastic gradient descent
  - Modification of parameters (q_i, p_u) relative to prediction error
  - Recommended algorithm
- Alternating least squares
  - Allows massive parallelization
  - Better for densely filled matrices

Learning Algorithms

- Calculation of the prediction error
  - Error = actual rating – predicted rating
  - $e_{ui} = r_{ui} - q_i^T p_u$
- Modification
  - By magnitude proportional to γ
  - In the opposite direction of the gradient
  - $q_i \leftarrow q_i + \gamma \cdot (e_{ui} \cdot p_u - \lambda \cdot q_i)$

Biases

- Item or user specific rating variations are called biases
- Example:
  - Alice rates no movie with more than 2 (out of 5)
  - Movie X is hyped and rated with 5 only
- Matrix factorization allows modeling of biases

- Including bias parameters in the prediction:

$$\hat{r}_{ui} = \mu + b_i + b_u + q_i^T p_u$$

Temporal Dynamics

- Ratings may be affected by temporal effects
  - Popularity of an item may change
  - User’s identity and preferences may change
- Modeling temporal affects can improve accuracy significantly

- Rating predictions as a function of time:

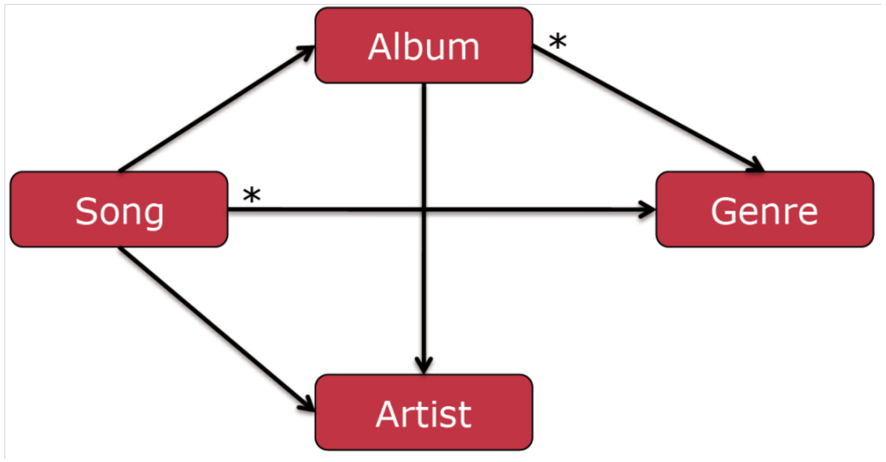
$$\hat{r}_{ui}(t) = \mu + b_i(t) + b_u(t) + q_i^T p_u(t)$$

Paper Evaluation

- High-level overview of matrix factorization techniques
- Mathematical foundations are pointed out but not elaborated
- Many useful references to related work
- Authors do not reveal their secret implementation tidbits

Application For KDD Cup

- Different item types
  - We must assume different prediction models!
  - We have explicit dependencies between items
- How to apply Matrix Factorization to the KDD data set?
  - Segment training data in separate sets for each type
  - Consider ratings for dependent items to make a prediction



Application For KDD Cup - Hypotheses

- Users change their taste in music
- Users tend to rate new songs better
  - Users get tired of songs
- Some artists and albums are hyped for a while
- Evergreens
  - Loved by many, but maybe also hated by some
- ...

References

[1] C. Volinsky et al.: "*Matrix Factorization Techniques for Recommender Systems*" *In: IEEE Computer, Vol. 42 (2009) , pp. 30-37 .*

[2] Simon Funk on the Stochastic Gradient Descent algorithm: *http://sifter.org/~simon/journal/20061211.html (Dec 2006)*

[3] Y.Koren et al.: *"Collaborative filtering with temporal dynamics" In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (2009), pp. 447-456.*

[4] Chih-Chao Ma*: "A Guide to Singular Value Decomposition for Collaborative Filtering" In: csientuedutw (2008)*

[5] Y.F. Hu, Y. Koren, C. Volinsky: *"Collaborative Filtering for Implicit Feedback Datasets" In: Proc. IEEE Int'l Conf. Data Mining (2008), pp. 263-372*

Discussion

Summary

- Matrix factorization is a promising approach for collaborative filtering
- Factor vectors are learned by minimizing the RSME
- Regularization to prevent overfitting
- Addition of bias parameters and temporal dynamics further improve accuracy

Outlook

- Develop strategies for applying matrix factorization on our data set with different item types
- Make use of the available dependencies between items
- Explorer biases and rating behaviors specific for our music domain