Multiverse Recommendation: N-dimensional Tensor Factorization for Context-aware Collaborative Filtering

Alexandros Karatzoglou[1] Xavier Amatriain [1] Linas Baltrunas [2]
Nuria Oliver [2]

[1]Telefonica Research
Barcelona, Spain
[2]Free University of Bolzano
Bolzano, Italy

November 4, 2010

Context in Recommender Systems

Context in Recommender Systems

Context is an important factor to consider in personalized Recommendation

Context in Recommender Systems

Context is an important factor to consider in personalized Recommendation

Context in Recommender Systems

Context is an important factor to consider in personalized Recommendation

Context in Recommender Systems

Context is an important factor to consider in personalized Recommendation

Current State of the Art in Context- aware Recommendation

Current State of the Art in Context- aware Recommendation

- Pre-Filtering Techniques

Current State of the Art in Context- aware Recommendation

- Pre-Filtering Techniques
- Post-Filtering Techniques

Current State of the Art in Context- aware Recommendation

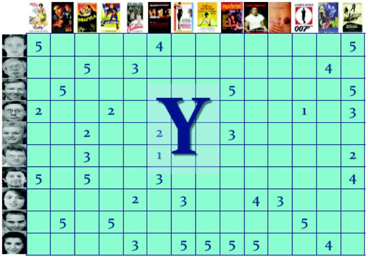
- Pre-Filtering Techniques
- Post-Filtering Techniques
- Contextual modeling

Current State of the Art in Context- aware Recommendation

- Pre-Filtering Techniques
- Post-Filtering Techniques
- Contextual modeling

The approach presented here fits in the Contextual Modeling category

Collaborative Filtering problem setting

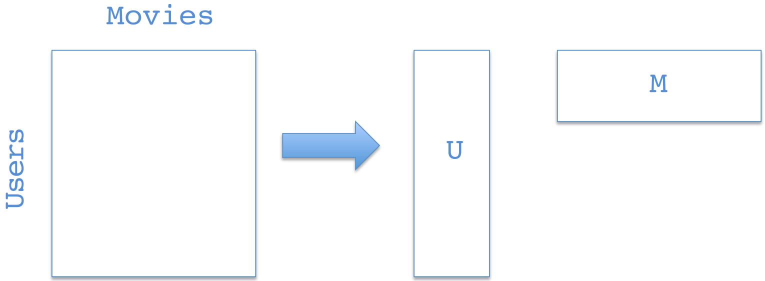


Typically data sizes e.g. Netlix data $n = 5 \times 10^5$, $m = 17 \times 10^3$

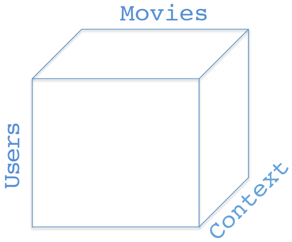
Standard Matrix Factorization

Find $U \in R^{n \times d}$ and $M \in R^{d \times m}$ so that $F = UM$

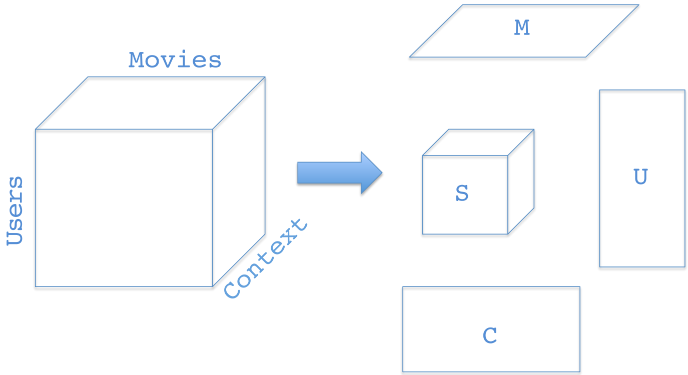
$$\text{minimize}_{U,M} L(F, Y) + \lambda \Omega(U, M)$$



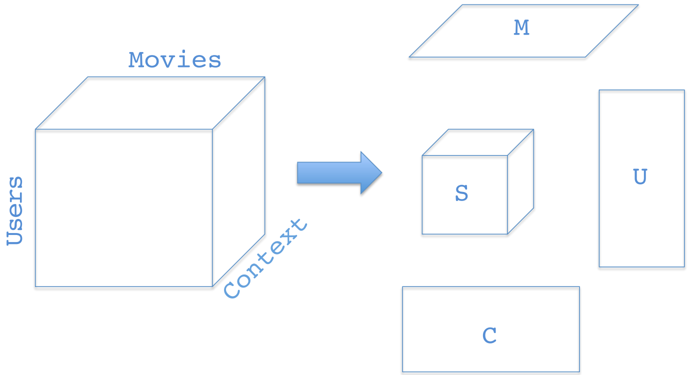
Multiverse Recommendation: Tensors for Context Aware Collaborative Filtering



Tensors for Context Aware Collaborative Filtering



Tensors for Context Aware Collaborative Filtering



$$F_{ijk} = S \times_U U_{i*} \times_M M_{j*} \times_C C_{k*}$$

$$R[U, M, C, S] := L(F, Y) + \Omega[U, M, C] + \Omega[S]$$

Regularization

$$\Omega[F] = \lambda_M \|M\|_F^2 + \lambda_U \|U\|_F^2 + \lambda_C \|C\|_F^2$$

$$\Omega[S] := \lambda_S \, \|S\|_{\mathrm{F}}^2 \tag{1}$$

Squared Error Loss Function

Many implementations of MF used a simple squared error regression loss function

$$l(f, y) = \frac{1}{2}(f - y)^2$$

thus the loss over all users and items is:

$$L(F, Y) = \sum_{i}^{n} \sum_{j}^{m} l(f_{ij}, y_{ij})$$

Note that this loss provides an estimate of the conditional mean

Absolute Error Loss Function

Alternatively one can use the absolute error loss function

$$l(f, y) = |f - y|$$

thus the loss over all users and items is:

$$L(F, Y) = \sum_{i}^{n} \sum_{j}^{m} l(f_{ij}, y_{ij})$$

which provides an estimate of the conditional median

Optimization - Stochastic Gradient Descent for TF

The partial gradients with respect to U, M, C and S can then be written as:

$$\partial_{U_{i*}} l(F_{ijk}, Y_{ijk}) = \partial_{F_{ijk}} l(F_{ijk}, Y_{ijk}) S \times_M M_{j*} \times_C C_{k*}$$
$$\partial_{M_{j*}} l(F_{ijk}, Y_{ijk}) = \partial_{F_{ijk}} l(F_{ijk}, Y_{ijk}) S \times_U U_{i*} \times_C C_{k*}$$
$$\partial_{C_{k*}} l(F_{ijk}, Y_{ijk}) = \partial_{F_{ijk}} l(F_{ijk}, Y_{ijk}) S \times_U U_{i*} \times_M M_{j*}$$
$$\partial_S l(F_{ijk}, Y_{ijk}) = \partial_{F_{ijk}} l(F_{ijk}, Y_{ijk}) U_{i*} \otimes M_{j*} \otimes C_{k*}$$

Optimization - Stochastic Gradient Descent for TF

The partial gradients with respect to U, M, C and S can then be written as:

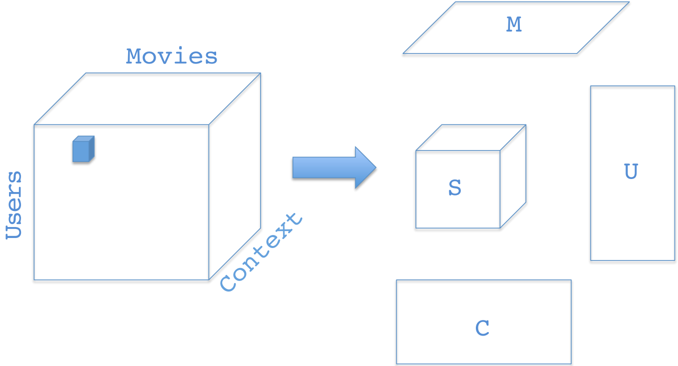
$$\partial_{U_{i*}} l(F_{ijk}, Y_{ijk}) = \partial_{F_{ijk}} l(F_{ijk}, Y_{ijk}) S \times_M M_{j*} \times_C C_{k*}$$
$$\partial_{M_{j*}} l(F_{ijk}, Y_{ijk}) = \partial_{F_{ijk}} l(F_{ijk}, Y_{ijk}) S \times_U U_{i*} \times_C C_{k*}$$
$$\partial_{C_{k*}} l(F_{ijk}, Y_{ijk}) = \partial_{F_{ijk}} l(F_{ijk}, Y_{ijk}) S \times_U U_{i*} \times_M M_{j*}$$
$$\partial_S l(F_{ijk}, Y_{ijk}) = \partial_{F_{ijk}} l(F_{ijk}, Y_{ijk}) U_{i*} \otimes M_{j*} \otimes C_{k*}$$

We then iteratively update the parameter matrices and tensors using the following update rules:

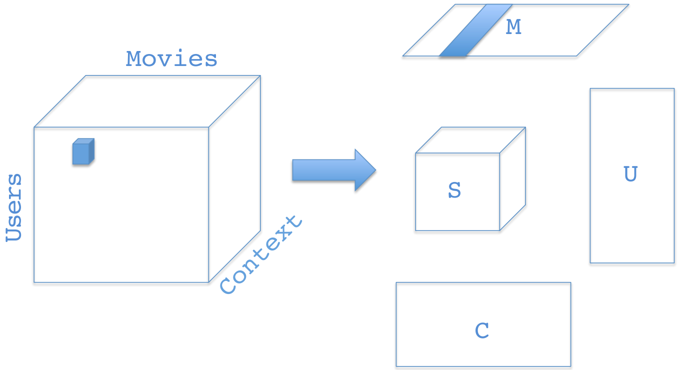
$$U_{i*}^{t+1} = U_{i*}^{t} - \eta \partial_U L - \eta \lambda_U U_{i*}$$
$$M_{j*}^{t+1} = M_{j*}^{t} - \eta \partial_M L - \eta \lambda_M M_{j*}$$
$$C_{k*}^{t+1} = C_{k*}^{t} - \eta \partial_C L - \eta \lambda_C C_{k*}$$
$$S^{t+1} = S^{t} - \eta \partial_S l(F_{ijk}, Y_{ijk}) - \eta \lambda_S S$$

where η is the learning rate.

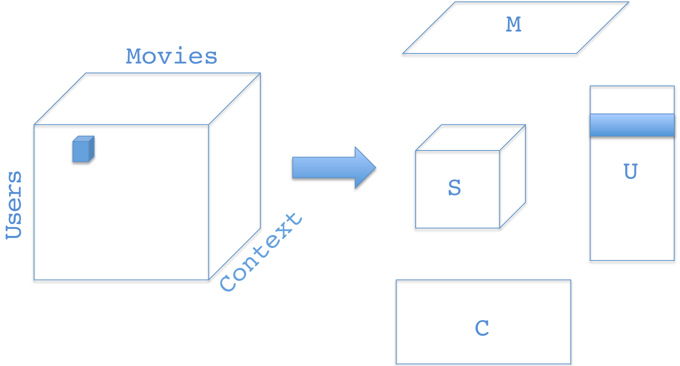
Optimization - Stochastic Gradient Descent for TF



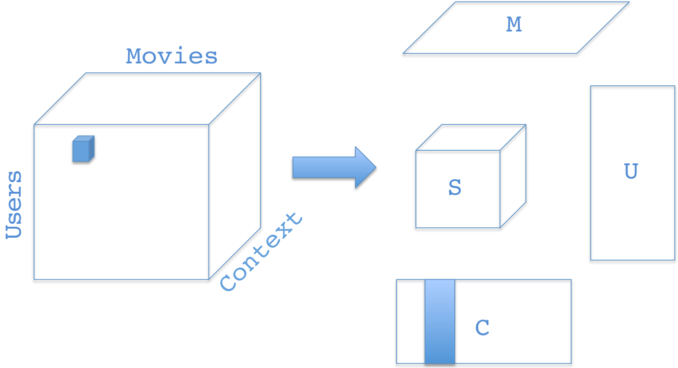
Optimization - Stochastic Gradient Descent for TF



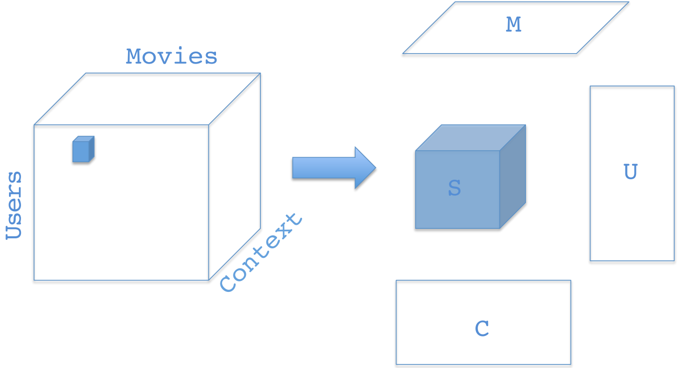
Optimization - Stochastic Gradient Descent for TF



Optimization - Stochastic Gradient Descent for TF



Optimization - Stochastic Gradient Descent for TF



Experimental evaluation

We evaluate our model on contextual rating data and computing the *Mean Absolute Error (MAE)*,using 5-fold cross validation defined as follows:

$$MAE = \frac{1}{K}\sum_{ijk}^{n,m,c} D_{ijk}|Y_{ijk} - F_{ijk}|$$

Data

Data set	Users	Movies	Context Dim.	Ratings	Scale
Yahoo!	7642	11915	2	221K	1-5
Adom.	84	192	5	1464	1-13
Food	212	20	2	6360	1-5

Table: Data set statistics

Context Aware Methods

- Pre-filtering based approach, (*G. Adomavicius et.al*), computes recommendations using *only* the ratings made in the same context as the target one
- Item splitting method (*L. Baltrunas, F. Ricci*) which identifies items which have significant differences in their rating under different context situations.

Results: Context vs. No Context

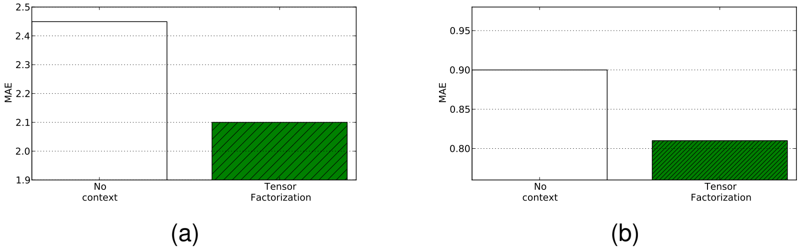


Figure: Comparison of matrix (no context) and tensor (context) factorization on the Adom and Food data.

Yahoo Artificial Data

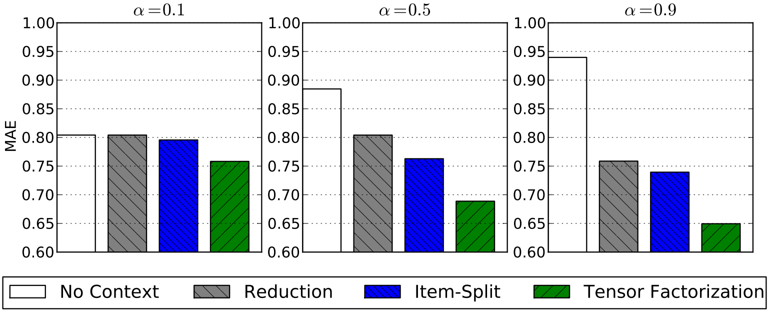
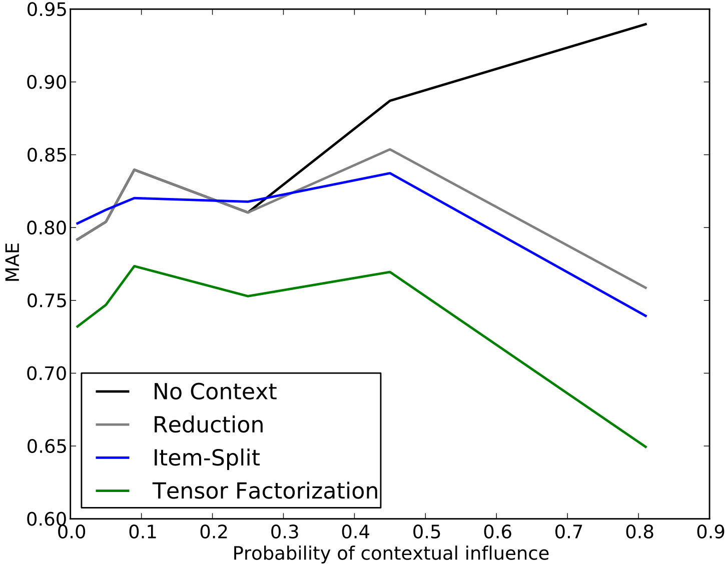


Figure: Comparison of context-aware methods on the Yahoo! artificial data

Yahoo Artificial Data



Tensor Factorization

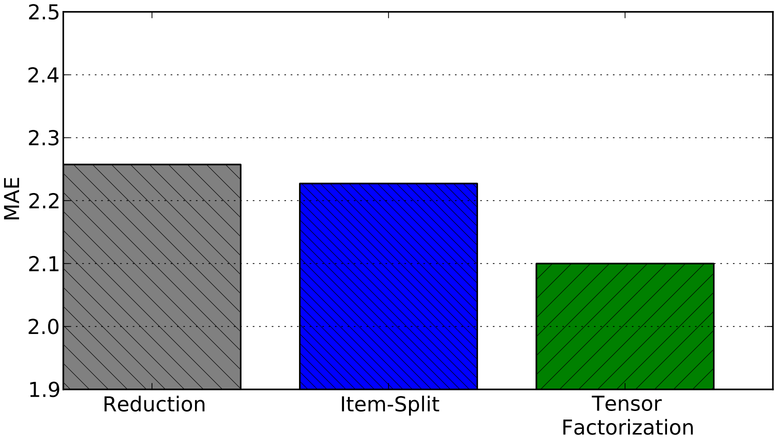


Figure: Comparison of context-aware methods on the Adom data.

Tensor Factorization

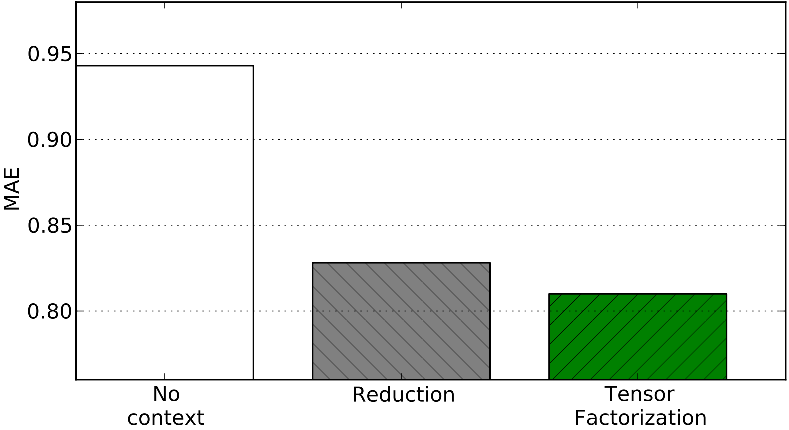


Figure: Comparison of context-aware methods on the Food data.

Conclusions

- Tensor Factorization methods seem to be promising for CARS
- Many different TF methods exist
- Future work: extend to implicit taste data
- Tensor representation of context data seems promising

Thank You !