Hands-on Recommender System Experiments with MyMediaLite

Zeno Gantner
zeno.gantner@nokia.com

Nokia gate5 GmbH, Berlin

13 September 2012

MyMediaLite Recommendation Algorithm Library

Major features:

- **scalable** implementations of many state-of-the-art recommendation methods – *tested on up to 700M events*
- evaluation framework for **reproducible** research
- **ready to be used**: command line tools, not programming necessary

MyMediaLite

- rating prediction
- item recommendation
- group recommendation

features

- **command-line tools**
- evaluation framework
- usable from C#, Python, Ruby, F#
- Java ports available

development

- written in C#, runs on Mono
- regular releases (ca. 1 every 2 months)

- simple
- free
- scalable
- well-documented
- well-tested
- choice

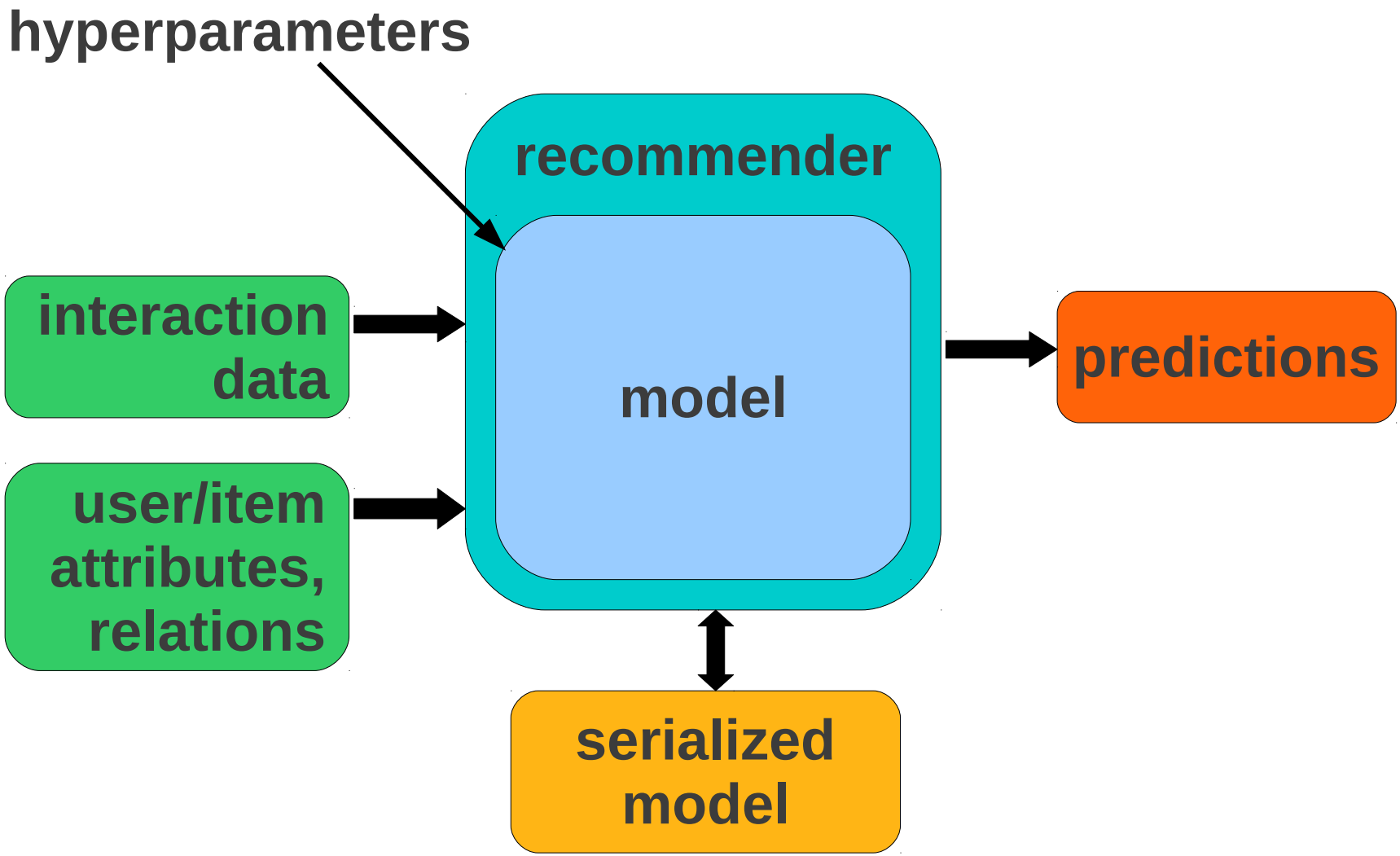
http://ismll.de/mymedialite

Methods in MyMediaLite

State-of-the-art recommendation methods in MyMediaLite:

- *kNN* variants
- *Online-Updating Regularized Kernel Matrix Factorization* [Rendle and Schmidt-Thieme, RecSys 2009]
- *SocialMF* [Jamali and Ester, RecSys 2010]
- *Asymmetric Factor Models* (*AFM*) [Paterek, KDD Cup 2007]
- *SVD++* [Koren, KDD 2008]
- *Weighted Regularized Matrix Factorization* (*WR-MF*) [Hu and Koren, ICDM 2008], [Pan et al., ICDM 2008]
- *BPR-MF* [Rendle et al., UAI 2009]

Simplified Architecture



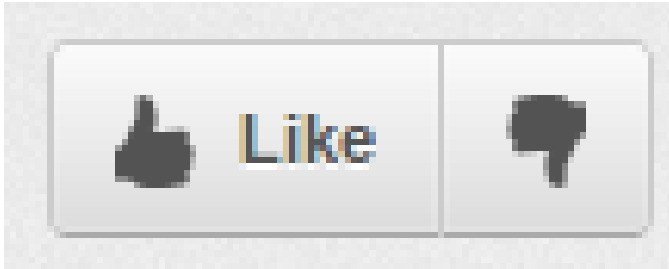
File Format: MovieLens

user ID	item ID	rating	timestamp
196	242	3	881250949
186	302	3	891717742
22	377	1	878887116
244	51	2	880606923

Remarks

- user and item IDs can be (almost) arbitrary strings
- separator: whitespace, tab, comma, ::
- alternative date/time format: yyyy-mm-dd
- rating and date/time fields are optional
- import script; Unix tools, Perl, Python ...

Explicit Feedback



★★★★★ = Must See
★★★★☆ = Will Enjoy
★★★☆☆ = It's OK
★★☆☆☆ = Fairly Bad
★☆☆☆☆ = Awful

Getting Help: Usage Information

```
rating_prediction --help
```

Data

rating_prediction --training-file=u1.base --test-file=u1.test

Recommender Options

rating_prediction --training-file=u.data --test-ratio=0.2

Fixing the Random Seed

rating_prediction . . . --random-seed=1

Choosing a Recommender

rating_prediction . . . --recommender=UserAverage

Choosing a Recommender

rating_prediction . . . --recommender=UserItemBaseline

Iterative Recommenders

```
rating_prediction
. . . --recommender=BiasedMatrixFactorization
--find-iter=1 --max-iter=30
```

Recommender Options (Hyperparameters)

```
rating_prediction
. . . --recommender-options="num_factors=5"
```

Recommender Options (Hyperparameters)

rating_prediction

. . . --recommender-options="num_factors=5 reg=0.05"

SVD++

rating_prediction . . . --recommender=SVDPlusPlus
--recommender-options="num_factors=5 reg=0.1
learn_rate=0.01"

Personalized Item Recommendation

amazon.de

Hallo, Zeno Gantner. Wir haben Empfehlungen für Sie. (Ausloggen)

Zenos Amazon | Sonderangebote | Wunschzettel | Gutschein

Alle Kategorien ansehen | Suche | Alle Kategorien

Mein Amazon | Ihre persönliche Seite | Ihre Empfehlungen | Diese Artikel bewerten | Verbesser

Zeno, willkommen bei Amazon.de (Wenn Sie nicht Zeno Gantner sind, klicken Sie bitte hier.)

Heutige Empfehlungen für Sie

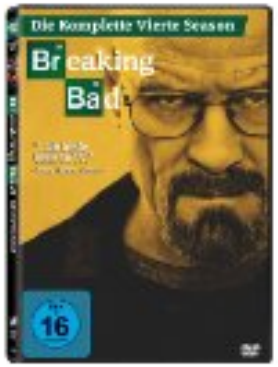
Hier sind einige der Ihnen empfohlenen Artikel.
Klicken Sie hier, um **alle Empfehlungen anzuzeigen**.

Seite 3 von 88 (Zum Anfang)

Boardwalk Empire Season 1 (L... DVD ~ Steve Buscemi

(26) EUR 44,99

Diese Empfehlung korrigieren



Breaking Bad - Die komplett... DVD ~ Bryan Cranston

(44) EUR 22,95

Diese Empfehlung korrigieren

Implicit Feedback

Behavior that is not an immediate expression of preference

- views
- clicks
- purchases

Advantages over explicit feedback:

- easy to collect
- available in abundance

positive-only feedback

Item Recommendation Tool: Very Similar Usage

```
item_recommendation --training-file=u.data
--test-ratio=0.2
```

Item Recommendation Tool

item_recommendation . . . --recommender=UserKNN

Choosing a Different Correlation/Similarity

item_recommendation
. . . --recommender-options="correlation=Jaccard"

Option Shortcuts

item_recommendation
. . . --recommender-options="cor=Cosine w=true q=1.5"

Iterative Recommenders / Save Predictions to Disk

```
item_recommendation ... --recommender=WRMF
--find-iter=1 --max-iter=10 --prediction-file=pred.txt
```

Left out from this presentation

- parallelization
- --cross-validation=K --chronological-split=2012-01-01
- limiting the test users/candidate items
- attribute- and relation-aware recommenders
- user-to-item recommendation
- top-n evaluation of rating predictors: rating_based_ranking
- --online-evaluation
- --repeated-items
- --save-model=FILE --load-model=FILE
- --cutoff=1.05 --measure=RMSE --epsilon=0.001
- tricks to save memory, e.g. --no-id-mapping --rating-type=byte
- ...

Instead of a Conclusion: 2-Hour Projects

Parallel processing

- similarity computation
- BPR matrix factorization

Correlations

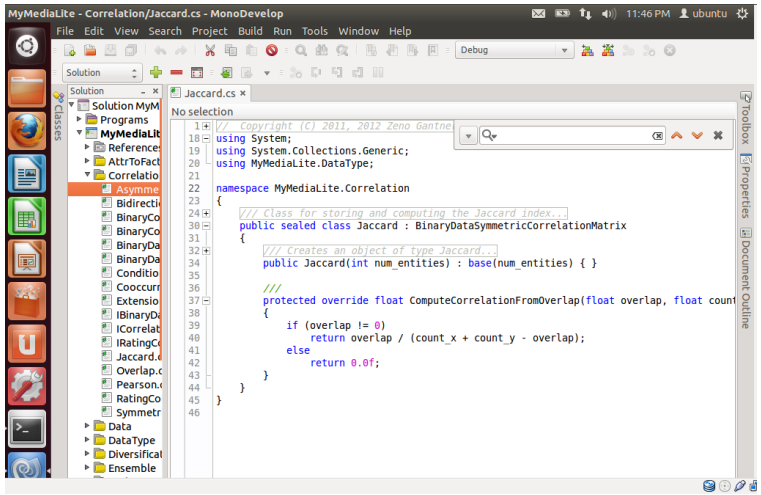
- Dice, Tyversky
- Jaccard index for $\{1, -1, ?\}$

Algorithms

- SGD learning for WRMF
- ALS learning for MF
- *your favorite algorithm*

Evaluation

- expected reciprocal rank (ERR)
- Kendall's Tau; Spearman



http://recsyswiki.com/wiki/MyMediaLite/Workshop_projects

Want to work in Berlin?

Nokia is hiring!

zeno.gantner@nokia.com