Linking Content Information with Bayesian Personalized Ranking via Multiple Content Alignments

Ladislav Peška

Department of Software Engineering,
Charles University in Prague,
Czech Republic

Challenge

- Provide ranked list of top-k best objects matching to the user's preference
  - Extend collaborative-filtering with content-based alignments
- Combine content-based and collaborative approaches
  - Reduce the effect of the **cold-start problem**
    - Extend collaborative filtering (matrix factorization) by content-based component
  - Reduce the problem of **non-informative content**
    - Find and integrate additional CB data sources
    - Incorporate relevance estimation for the data sources
  - Reduce the overspecialization and obvious recommendations problems
    - Incorporate content-based method into the collaborative framework

Algorithm

- Matrix factorization/completion

$$\mathbf{R} \approx \mathbf{U}\mathbf{O}^T = \underbrace{\begin{bmatrix} \mu_1^T \\ \mu_2^T \\ \vdots \end{bmatrix}}_{n \times f} \times \underbrace{\begin{bmatrix} v_1 & v_2 & \dots \end{bmatrix}}_{f \times m} \qquad \hat{r}_{i,j} := \mu_i^T \times v_j$$

Standartní BPR ze 2. přednášky

- Ranking-oriented optimization
  - Based on BPR MF[1]
  - Binary implicit feedback
  - Train set triples *(**u**ser, **g**ood, **b**ad object)*
    - *Maximize distance in rating of good and bad obj.*
    - *Stochastic Gradient Descend*

$$maximize \overbrace{\sum_{(u,g,b)\in T_S}}^{\text{Train set}} \ln \sigma\underbrace{(\hat{r}_{u,g} - \hat{r}_{u,b})}_{\text{Ranking correctness}} - \lambda\overbrace{(\|\mathbf{U}\|^2 + \|\mathbf{O}\|^2)}^{\text{Regularization}}$$

[1]Rendle et al.: BPR: Bayesian Personalized Ranking from Implicit Feedback, UAI 2009

BPR-MCA, Content Alignments

- **Content alignments:** *Highly similar (CB) objects/users should also have similar latent factors*
  - There are many possible knowledge bases/sources for content-based similarity
    => Multiple similarity matrices => **Multiple Content Alignments**
    - Usually, similarity matrices are not equally relevant => learn relevance

To pro vás není v tuto chvíli zásadní – důležitá je věta nahoře

BPR-MCA, Content Alignments

- **Content alignments:** *Highly similar (CB) objects/users should also have similar latent factors*
  - There are many possible knowledge bases/sources for content-based similarity
    => Multiple similarity matrices => **Multiple Content Alignments**
    - Usually, similarity matrices are not equally relevant
- For each train set triple
  - Select *top-k (10 in evaluation)* most similar users/objects w.r.t. each sim. matrix
  - Add penalty for the difference in latent factors of source and similar users/objects
  - Add parameter ω estimating relevance for each similarity matrix
    - Learned via Stochastic Gradient Descend

$$maximize \underbrace{\sum_{\forall(u,g,b)}}_{\text{Train set}} \ln \overbrace{\sigma\left(\hat{r}_{u,g} - \hat{r}_{u,b}\right)}^{\text{Ranking correctness}} - \lambda\left(\underbrace{\|\mathbf{U}\|^2 + \|\mathbf{O}\|^2}_{\text{Regularization}} + \underbrace{\sum_{\forall \bar{u}\in s(u)_p}}_{\text{For all similar users/objects, For all sim. matrices}} \overset{\text{Sim. mat. relevance}}{\omega_p} \overset{\text{Similarity of users}}{s^U_{p,u,\bar{u}}} \|\underbrace{\boldsymbol{u}_u - \boldsymbol{u}_{\bar{u}}}_{\text{Difference of latent factors}}\|^2\right)$$

Experimental Dataset

Dál jsou konkrétní experimenty a výsledky – můžete přeskočit, nechávám jen pro zajímavost

- MovieLens 1M dataset
  - 6000 users, 4000 movies, 1M ratings
  - Binarized rating*(3*, 4*, 5* => positive preference)*
  - Users: Gender, Age group, Occupation, ZIP code
  - Movies: Title, Genres
- Extensions *(cosine sim., TF-IDF weighting)*
  - ZIP code statistics (*UnitedStatesZipCodes.org*)
    - Population density, singles ration, vacant houses, racial groups,…
  - IMDB API
    - Directors, actors, average rating, year, language,…
  - DBTropes
    - „Intrinsic plot features“

Experimental Dataset

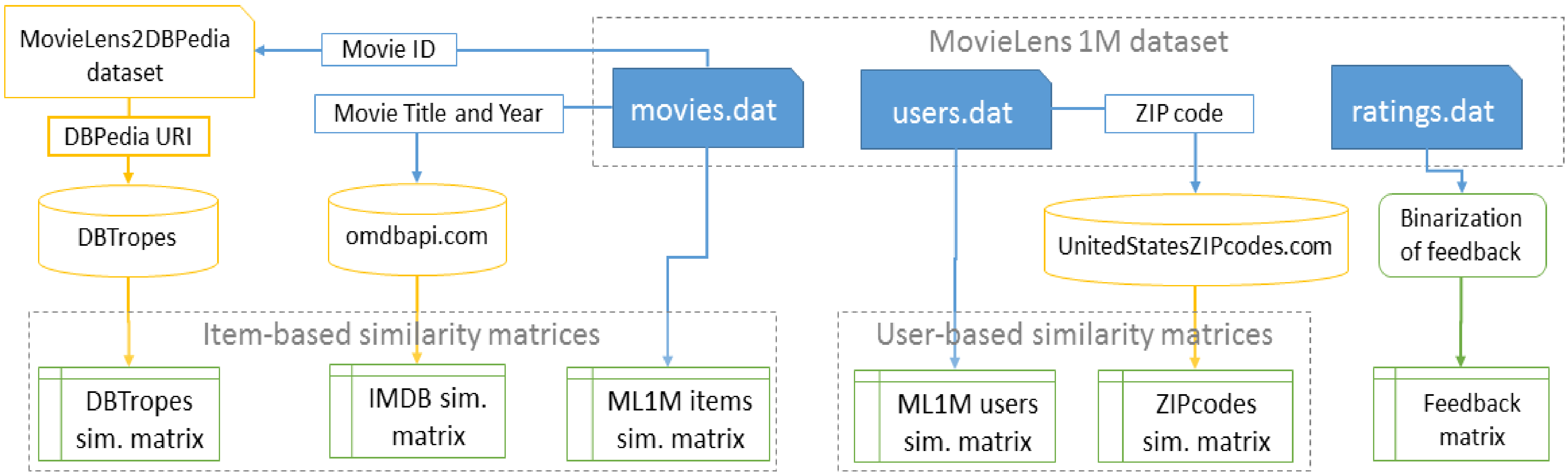
- DBTropes – The Matrix

type	Talking Is a Free Action
comment	*Talking Is a Free Action: When Morpheus does a Barrier-Busting Blow and jumps onto Agent Smith in the bathroom, the latter lies still for a couple if seconds during which Morpheus orders Trinity to get Neo out of danger.*

type	The Chooser of The One
comment	*The Chooser of The One: The Oracle can tell who is or isn't The One.*

type	Beard of Evil
comment	*Beard of Evil / Bald of Evil: Cypher and his pencil-thin goatee.*

Schema of MovieLens1M Dataset Extensions



Coverage:

15% 79% 100% 100% 97%

Evaluation

- Simulate ***cold-start problem*** on extended MovieLens1M
  - Monte-Carlo CV,
    - Hide 75%, 90%, 95% and 98% of interactions (p75, p90, p95, p98)
    - Repeat 10 times, the same train/test cuts for each method
    - *MovieLens1M is extremly dense dataset, even 98% hidden interactions is realistical*
  - Graph-based *(CB recommendation)*
  - WRMF *(rating oriented MF for implicit feedback)*[1]
  - BPR[2]
  - BPR-MCA2 *(based on original MovieLens1M dataset)*
  - BPR-MCA5 *(based on the whole extended dataset)*
  - BPR-MCA5*uniform* *(without learning relevance of similarity matrices)*

[1]Hu et al.: Collaborative Filtering for Implicit Feedback Datasets. *IEEE ICDM 2008*

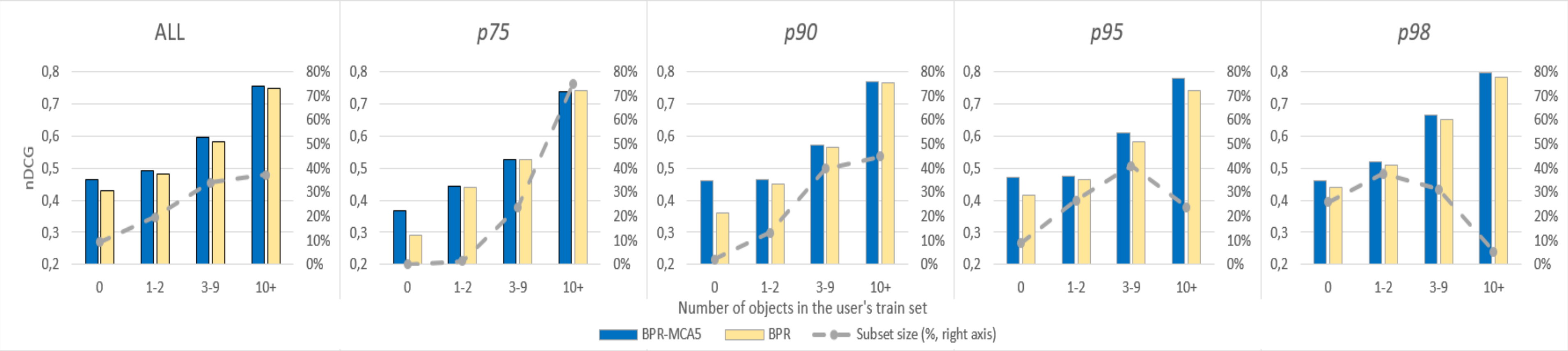
[2]Rendle et al.: BPR: Bayesian Personalized Ranking from Implicit Feedback, *UAI 2009*

Results - Methods

Method	nDCG			
	p75	p90	p95	p98
Graph-based	*0.4863	*0.4930	*0.4913	*0.4853
WRMF	*0.6634	*0.5854	*0.5421	*0.5124
BPR	*0.6848	*0.6351	*0.5739	*0.5490
BPR-MCA2	**0.6872**	*0.6361	*0.5732	*0.5510
BPR-MCA5	*0.6836	**0.6425**	**0.6016**	**0.5645**
BPR-MCA5uniform	*0.6846	*0.6336	*0.5757	*0.5516

- BPR-MCA5 dominates in all but p75 scenarios
  - *Cold-start in p75 not difficult enough to justify content-based extensions*
  - *Content-based component increased on signifficance with more complex cold-start scenarios*
- Both ***extended datasets*** and ***learning relevance*** of the similarity matrices are relevant
- Graph-based and rating-based MF are inferior to BPR

Results – Train Set Size



- BPR-MCA5 improves mainly results for users with only a few known interactions (0-2)
- However, in thougher cold-start scenarios (p95, p98) signifficant improvements are generated for all users

=> *Improvement cannot be contributed solely to the improved prediction for users without any interactions*

Conclusions, Future Work

Key outcomes

- Proposed BPR-MCA algorithm improved performance in ranking task on MovieLens1M under cold-start conditions
- Extension of the MovieLens1M dataset
  - Find & evaluate potentially relevant data sources seems as a key task

Future work, Open Problems

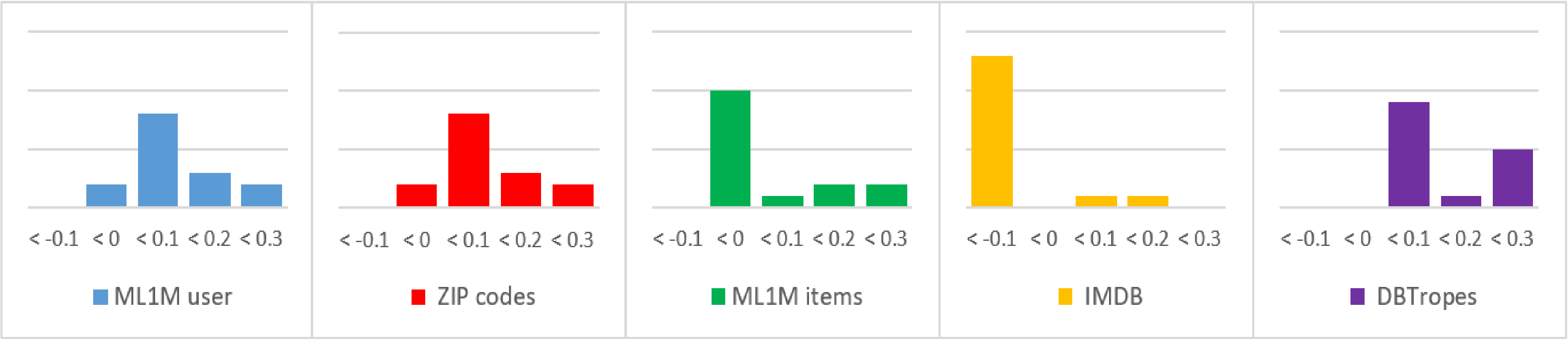
- Incorporate Mutlimodal Implicit Feedback (beyond binary feedback)
- Incorporate Diversity / Novelty / Temporal dependence…
- Further evaluation scenarios, datasets
  - Positive-only similarity weights?
  - Tradeoff between the ammount of feedback vs. content-based similarity?
  - Further „intrinsic“ datasets such as DBTropes, e.g., visual descriptors
- On-line deployment and evaluation

Thank you!

Questions, comments?

Supplementary materials: *http://www.ksi.mff.cuni.cz/~peska/BPR_MCA*
Slides: *https://www.slideshare.net/LadislavPeska/linking-content-information-with-bayesian-personalized-ranking-via-multiple-content-alignments*

Results – Weights of Similarity Matrices

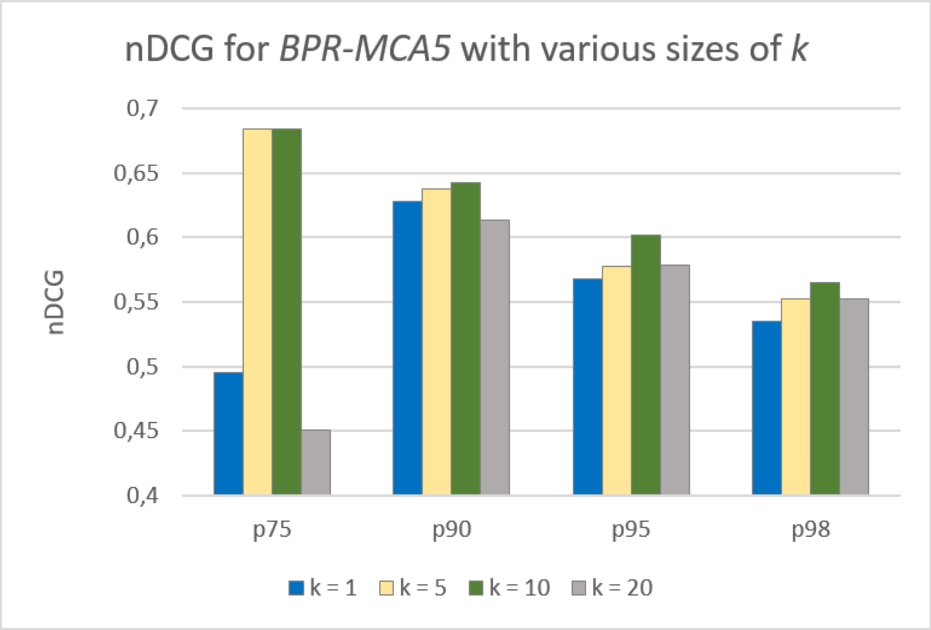


- Similarity based on „external“ movie features seems to have rather negative correlation with intrinsic user-based similarity
- Both user-based similarity matrices seems relevant
- Despite the low coverage, DBTropes-based similarity receives highest weights in average

Results – Size of *k* Parameter

- Parameter *k* sets the size of keeped nearest neighbors for each similarity matrix.

- Optimal value for MovieLens1M dataset seems to be between 5 and10.



$$\boldsymbol{\mu}_i = \boldsymbol{\mu}_i + \eta\left(x \cdot (\mathbf{v}_j - \mathbf{v}_k) - \lambda_R\left(\boldsymbol{\mu}_i + \sum_{p=1}^{n_S}\sum_{\bar{\imath}=1}^{n} \omega_p \mathbf{S}^{U}_{p,i,\bar{\imath}}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_{\bar{\imath}})\right)\right)$$

$$\mathbf{v}_j = \mathbf{v}_j + \eta\left(x \cdot \boldsymbol{\mu}_i - \lambda_R\left(\mathbf{v}_j + \sum_{q=1}^{m_S}\sum_{\bar{\jmath}=1}^{m} \omega_q \mathbf{S}^{O}_{q,j,\bar{\jmath}}(\mathbf{v}_j - \mathbf{v}_{\bar{\jmath}})\right)\right)$$

$$\mathbf{v}_k = \mathbf{v}_k + \eta\left(x \cdot (-\boldsymbol{\mu}_i) - \lambda_R\left(\mathbf{v}_j + \sum_{q=1}^{m_S}\sum_{\bar{k}=1}^{m} \omega_q \mathbf{S}^{O}_{q,k,\bar{k}}(\mathbf{v}_k - \mathbf{v}_{\bar{k}})\right)\right)$$

$$\omega_p = \omega_p - \eta\left(\sum_{\bar{\imath}=1}^{n} \mathbf{S}^{U}_{p,i,\bar{\imath}}\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_{\bar{\imath}}\|^2 + \lambda_w(\omega_p - \omega_0)\right)$$

$$\omega_q = \omega_q - \eta\left(\sum_{\bar{\jmath}=1}^{m}\left(\mathbf{S}^{O}_{q,j,\bar{\jmath}}\left\|\mathbf{v}_j - \mathbf{v}_{\bar{\jmath}}\right\|^2 + \mathbf{S}^{O}_{q,k,\bar{\jmath}}\left\|\mathbf{v}_k - \mathbf{v}_{\bar{\jmath}}\right\|^2\right) + 2\lambda_w(\omega_p - \omega_0)\right)$$