Personalised Rating Prediction for New Users Using Latent Factor Models

Yanir Seroussi Fabian Bohnert Ingrid Zukerman

Faculty of Information Technology
Monash University
Melbourne, Australia

Outline

Motivation

- Recommender systems help users deal with information overload
  - Often rely on ratings to model user preferences

Motivation

- Recommender systems help users deal with information overload
  - Often rely on ratings to model user preferences
- Few ratings are available for new users
  - Rating prediction is challenging
  - Recommendation generation is challenging

Motivation

- Recommender systems help users deal with information overload
  - Often rely on ratings to model user preferences
- Few ratings are available for new users
  - Rating prediction is challenging
  - Recommendation generation is challenging
- But background information is often available
  - Demographics
  - User-generated texts

Matrix Factorisation and New Users

- Matrix factorisation methods yield highly accurate and scalable rating predictions
  - ...but tend to perform poorly on new users

Matrix Factorisation and New Users

- Matrix factorisation methods yield highly accurate and scalable rating predictions
  - ...but tend to perform poorly on new users
- New users: the majority of the population
  - ...but a minority in *test samples*

Matrix Factorisation and New Users

- Matrix factorisation methods yield highly accurate and scalable rating predictions
  - ...but tend to perform poorly on new users
- New users: the majority of the population
  - ...but a minority in *test samples*
- We address the new user problem by considering explicit and implicit user attributes in matrix factorisation

Traditional Matrix Factorisation

- The rating matrix $\mathbf{R}_{N \times M}$ is decomposed into user-factors $\mathbf{P}_{D \times N}$ and item-factors $\mathbf{Q}_{D \times M}$:

$$\mathbf{R} \approx \mathbf{P}^T \mathbf{Q}$$

Traditional Matrix Factorisation

- The rating matrix $\mathbf{R}_{N\times M}$ is decomposed into user-factors $\mathbf{P}_{D\times N}$ and item-factors $\mathbf{Q}_{D\times M}$:

$$\mathbf{R} \approx \mathbf{P}^T\mathbf{Q}$$

- The predicted rating for user u and item i is

$$\hat{r}_{ui} = \mathbf{p}_u^T\mathbf{q}_i$$

Traditional Matrix Factorisation

- The rating matrix $\mathbf{R}_{N\times M}$ is decomposed into user-factors $\mathbf{P}_{D\times N}$ and item-factors $\mathbf{Q}_{D\times M}$:

$$\mathbf{R} \approx \mathbf{P}^T\mathbf{Q}$$

- The predicted rating for user u and item i is
- Item biases and user biases can also be considered

$$\hat{r}_{ui} = \mu + b_i + b_u + \mathbf{p}_u^T\mathbf{q}_i$$

Traditional Matrix Factorisation

- The rating matrix $\mathbf{R}_{N\times M}$ is decomposed into user-factors $\mathbf{P}_{D\times N}$ and item-factors $\mathbf{Q}_{D\times M}$:

$$\mathbf{R} \approx \mathbf{P}^T\mathbf{Q}$$

- The predicted rating for user u and item i is
- Item biases and user biases can also be considered
- Ignoring user information improves accuracy for new users
  - ...but yields non-personalised predictions

$$\hat{r}_{ui} = \mu + b_i$$

Matrix Factorisation with User Attributes

- Considering user attributes for new users:
  - Switching model

$$\hat{r}_{ui} = \mu + b_i + \begin{cases} & u \text{ is a new user} \\ b_u + \mathbf{p}_u^T \mathbf{q}_i & \text{otherwise} \end{cases}$$

Matrix Factorisation with User Attributes

- Considering user attributes for new users:
  - Switching model
  - Probabilities of user attributes ($P(a|u)$)

$$\hat{r}_{ui} = \mu + b_i + \begin{cases} \sum_a P(a|u) & u \text{ is a new user} \\ b_u + \mathbf{p}_u^T \mathbf{q}_i & \text{otherwise} \end{cases}$$

Matrix Factorisation with User Attributes

- Considering user attributes for new users:
  - Switching model
  - Probabilities of user attributes ($P(a|u)$)
  - Attribute biases (b_a) and factors ($\mathbf{y}_a$)

$$\hat{r}_{ui} = \mu + b_i + \begin{cases} \sum_a P(a|u)\left(b_a + \mathbf{y}_a^T \mathbf{q}_i\right) & u \text{ is a new user} \\ b_u + \mathbf{p}_u^T \mathbf{q}_i & \text{otherwise} \end{cases}$$

Explicit Demographic Attributes

- Our model is defined for discrete attributes

Explicit Demographic Attributes

- Our model is defined for discrete attributes
- Some demographic characteristics are non-discrete

Explicit Demographic Attributes

- Our model is defined for discrete attributes
- Some demographic characteristics are non-discrete
- Employing too many attributes may result in overfitting

Explicit Demographic Attributes

- Our model is defined for discrete attributes
- Some demographic characteristics are non-discrete
- Employing too many attributes may result in overfitting
- Therefore, we reduce the dimensionality of the attributes with *Principal Component Analysis* (PCA)

Implicit Textual Attributes

- Converting texts into attributes is challenging

Implicit Textual Attributes

- Converting texts into attributes is challenging
- Token frequencies can be used to infer demographics
  - Continuous
  - Hundreds to thousands of tokens are required (overfitting)

Implicit Textual Attributes

- Converting texts into attributes is challenging
- Token frequencies can be used to infer demographics
  - Continuous
  - Hundreds to thousands of tokens are required (overfitting)
- We use *Latent Dirichlet Allocation* (LDA) to model users:
  - Traditionally used to find document topic distributions
  - Each user is represented by a single document
  - The document's "topic" distribution is the user's attribute distribution

Experimental Setup and Datasets

- Measure: Normalised root mean square error (NRMSE)

Experimental Setup and Datasets

- Measure: Normalised root mean square error (NRMSE)
- Protocol:
  - GivenX: all target users have X training ratings
  - Ten-fold cross validation on users

Experimental Setup and Datasets

- Measure: Normalised root mean square error (NRMSE)
- Protocol:
  - GivenX: all target users have X training ratings
  - Ten-fold cross validation on users
- Datasets:

Experimental Setup and Datasets

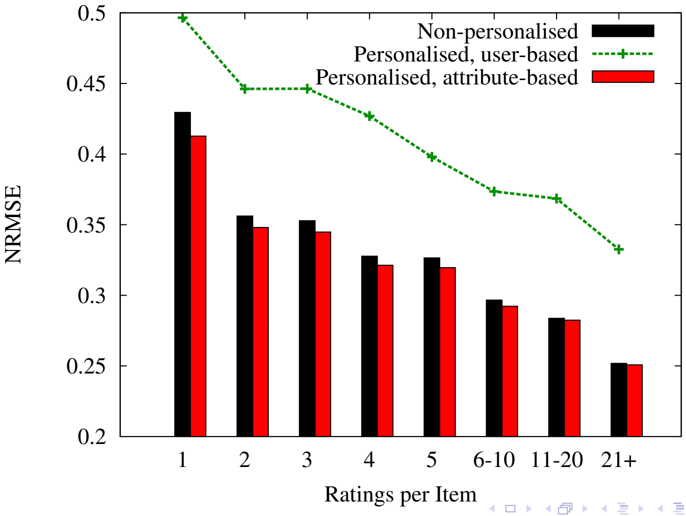
- Measure: Normalised root mean square error (NRMSE)
- Protocol:
  - GivenX: all target users have X training ratings
  - Ten-fold cross validation on users
- Datasets:
  - MovieLens100K:
    - 100K movie ratings
    - Sparsity: 93.695%
    - User demographics (age, gender, occupation and location)

Experimental Setup and Datasets

- Measure: Normalised root mean square error (NRMSE)
- Protocol:
  - GivenX: all target users have X training ratings
  - Ten-fold cross validation on users
- Datasets:
  - MovieLens100K:
    - 100K movie ratings
    - Sparsity: 93.695%
    - User demographics (age, gender, occupation and location)
  - IMDb1M:
    - 67K rated movie reviews
    - Sparsity: 99.988%
    - 200K message board posts

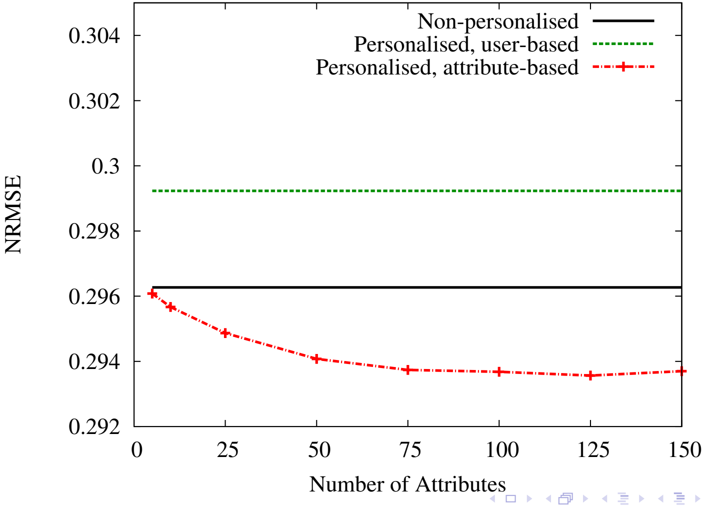
Matrix Factorisation with Demographic User Attributes

Protocol: Given1; Dataset: MovieLens100K



Matrix Factorisation with Textual User Attributes

Protocol: Given1; Dataset: IMDb1M



Impact of Our Results

- Hard to predict impact on recommendation quality

Impact of Our Results

- Hard to predict impact on recommendation quality
- Seemingly small NRMSE improvements. However:

Impact of Our Results

- Hard to predict impact on recommendation quality
- Seemingly small NRMSE improvements. However:
  - Small improvements matter (Koren, 2008)

Impact of Our Results

- Hard to predict impact on recommendation quality
- Seemingly small NRMSE improvements. However:
  - Small improvements matter (Koren, 2008)
  - Our improvements are four times larger on high predictions (8+/10) than the average improvements

Conclusion

- Main contributions:
  - Matrix factorisation with user attributes
  - Demographic and textual user attributes
  - Personalised, more accurate rating predictions for new users

Conclusion

- Main contributions:
  - Matrix factorisation with user attributes
  - Demographic and textual user attributes
  - Personalised, more accurate rating predictions for new users
- Future work:
  - Integrate sentiment analysis
  - Extend our model to consider item attributes
  - Evaluate our model as part of a recommender system