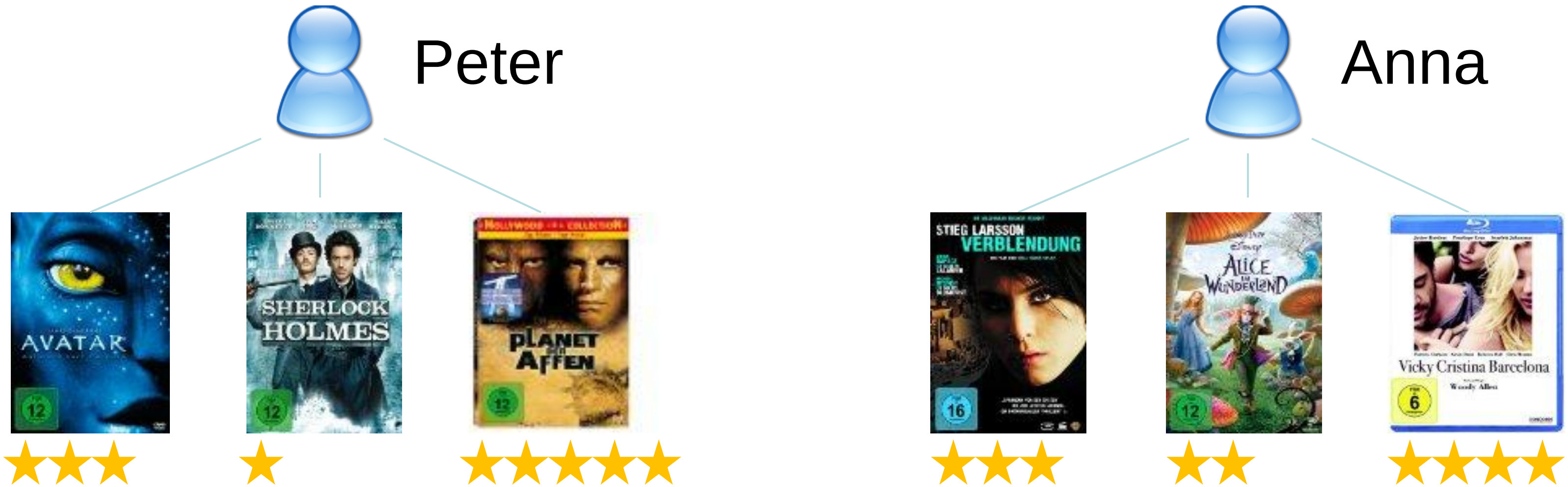
Graph-Based Model-Selection Framework for Large Ensembles

Krisztian Buza, Alexandros Nanopoulos, Lars Schmidt-Thieme
{buza, nanopoulos, schmidt-thieme}@ismll.de

Information Systems and Machine Learning Lab (ISMLL)
University of Hildesheim, Germany

Motivation
Popular Regression Problem: Rating Prediction



- What films should be recommended to each user?

Motivation

- Recommender systems:
  - predict which items users intend to buy, watch, …
  - statistical prediction of the ratings

User	Item	Rating
Peter	Avatar	★★★
Peter	Sherlock H.	★
Peter	Planet of A.	★★★★★
Anna	Verblendung	★★★
Anna	Alice in W.	★★
Anna	Vickey, C., B.	★★★★★
Peter	Benjamin B.	?
Peter	Australia	?
Anna	Benjamin B.	?
Anna	Australia	?
…	…	?

User	Item	Rating
Peter	Avatar	★★★
Peter	Sherlock H.	★
Peter	Planet of A.	★★★★★
Anna	Verblendung	★★★
Anna	Alice in W.	★★
Anna	Vickey, C., B.	★★★★★
Peter	Benjamin B.	★★★★
Peter	Australia	★★
Anna	Benjamin B.	★★★
Anna	Australia	★★★
…	…	…

Motivation

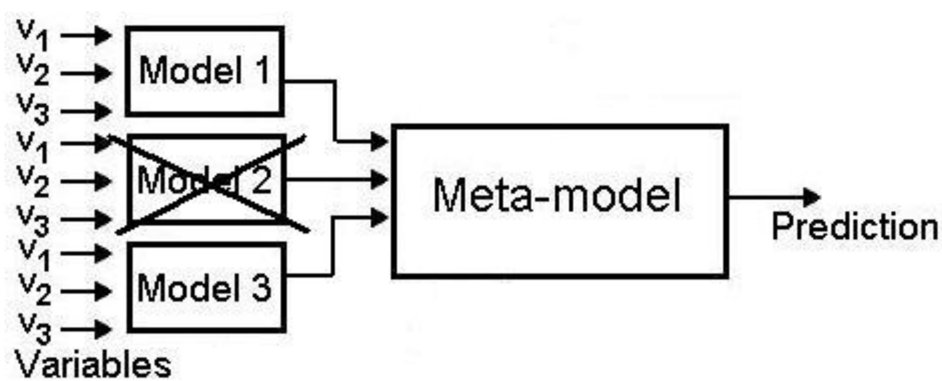
- Many different regression algorithms
  - Various methods (k-NN, SVM, matrix factorisation, etc.)
- Combination of regression models
  - Simple method: averaging
  - More advanced method: stacking
  - Different models compensate each other's errors
- Special setting
  - Large number (≈100+) of regression models are present
- Compensation effect
  - can be depressed if many models have similar error characteristic
    → **selection of the right models** is crucial

Outline

- Contribution
- Graph-Based Ensemble Framework
- Ensemble Techniques
  - Basic, EarlyStop, RegOpt, GraphOpt
- Experiments
- Conclusion

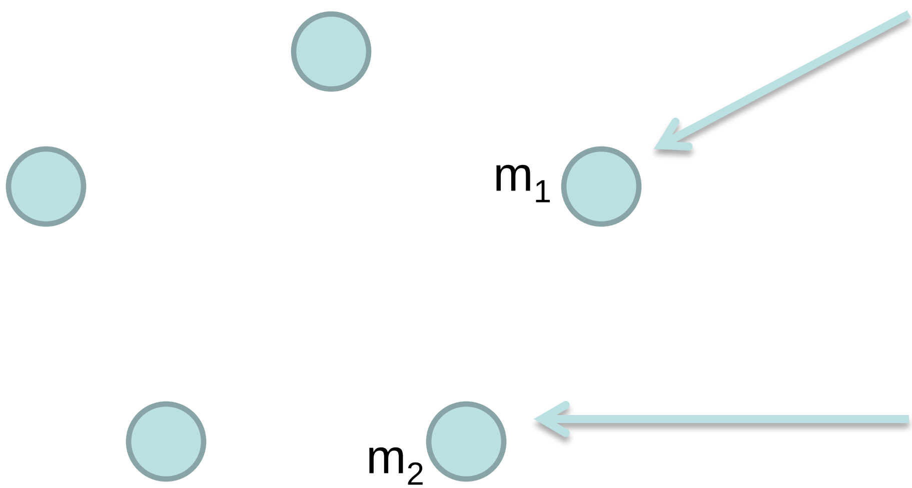
Contribution

- Graph-based ensemble framework for **model selection**. This
  - works according to the stacking schema
  - is generic enough to describe a wide range of model selection strategies varying from meta-filter to meta-wrapper approaches
  - allows to deploy our ensemble techniques *Basic, EarlyStop, RegOpt, GraphOpt*



Graph-Based Ensemble Framework

- Model-pair graph
  - each vertex corresponds to a model (in our case: model = recommender system)



Model 1

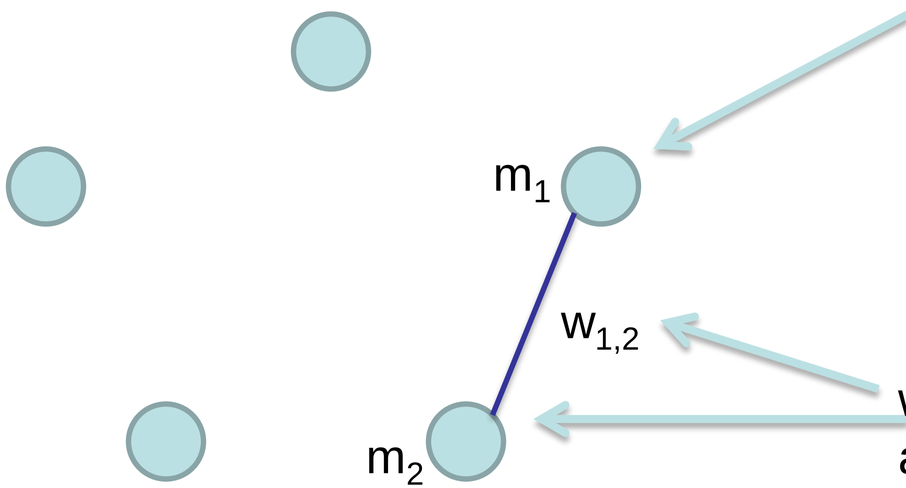
User	Item	Prediction
Peter	Benjamin B.	★★★★
Peter	Australia	★★
Anna	Benjamin B.	★★★
Anna	Australia	★★★
…	…	…

Model 2

User	Item	Prediction
Peter	Benjamin B.	★★★
Peter	Australia	★★★
Anna	Benjamin B.	★★
Anna	Australia	★
…	…	…

Graph-Based Ensemble Framework

- Model-pair graph
  - The weight of an edge reflects the mutual error compensation power of the models connected by that edge



Model 1

User	Item	Prediction
Peter	Benjamin B.	★★★★
Peter	Australia	★★
[illegible]	[illegible]	[illegible]

Avg. of Model 1 and Model 2

User	Item	Prediction
Peter	Benjamin B.	3.5
Peter	Australia	2.5
Anna	Benjamin B.	2.5
Anna	Australia	2
…	…	…

Model 2

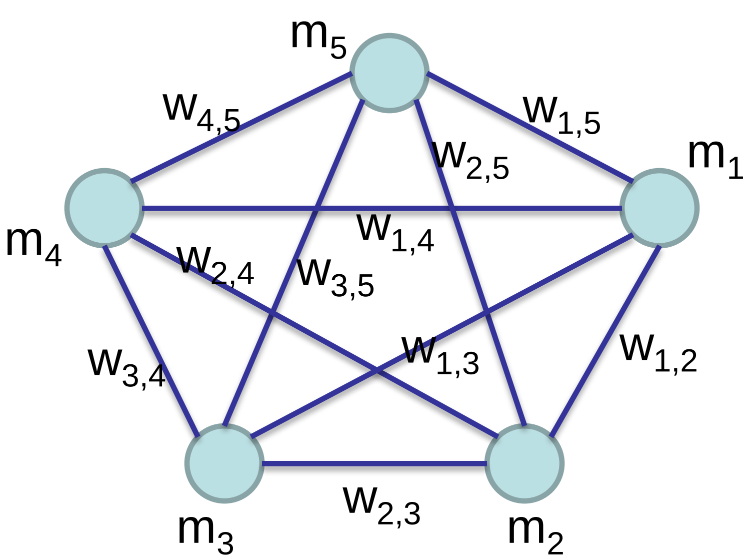
User	Item	Prediction
Peter	Benjamin B.	★★★
Peter	Australia	★★★
Anna	Benjamin B.	★★
Anna	Australia	★
…	…	…

$w_{1,2}$ = *error* (e.g. RMSE) of the average of the both models.

(Weights are calculated using labeled **train** data.)

Graph-Based Ensemble Framework

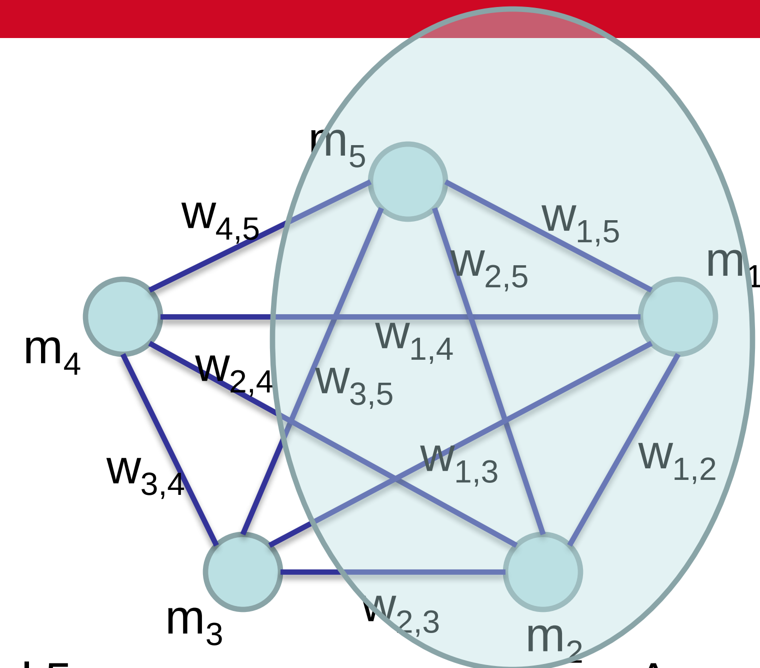
- Method (simplified)
  – Search for a subgraph representing models with optimal error compensation



1. Build graph (using ***error*** function)
2. Let M denote the optimal set of models found till now (initially: M={})
3. Process edges in order of their weights (begin with the best, w.r.t. ***restrictions***)
   – **if** ***score***(M + {vertices of current egde}) better than *score*(M) **then** M := M + {vertices of current edge)
4. Stacking of models in M (***meta model type***)

Basic

- ***Error function***: RMSE
- ***Score function***: f_{avg}
- ***Meta-model type***: multivariate linear regression
- ***Restrictions*** of edge examination: none



Model 1

User	Item	Pred.
u1	i1	4
u1	i2	2
u2	i1	3
u2	i2	3
…	…	…

Model 2

User	Item	Pred.
u1	i1	3
u1	i2	3
u2	i1	2
u2	i2	1
…	…	…

Model 5

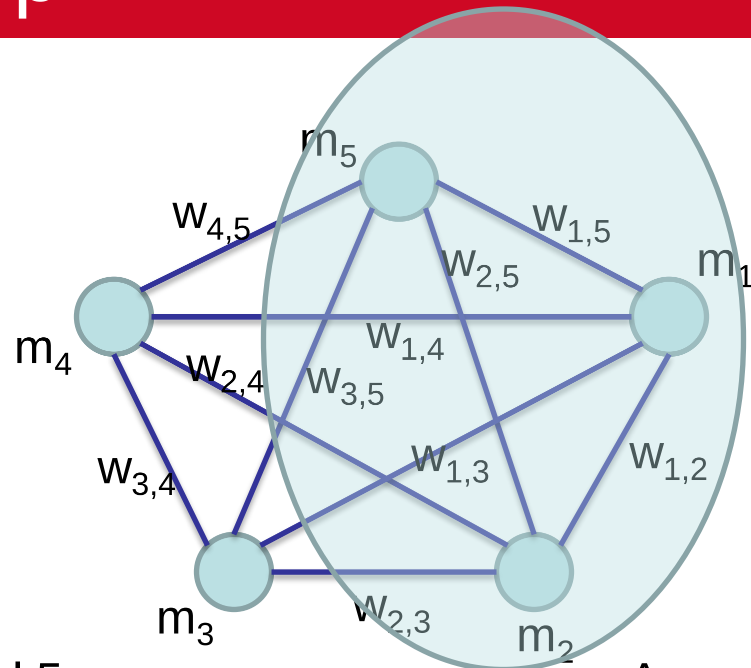
User	Item	Pred.
u1	i1	5
u1	i2	2
u2	i1	4
u2	i2	3
…	…	…

Average

User	Item	Pred.
u1	i1	4
u1	i2	2.33
u2	i1	3
u2	i2	2.33
…	…	…

EarlyStop

- ***Error function***: RMSE
- ***Score function***: f_{avg}
- ***Meta-model type***: multivariate linear regression
- ***Restrictions*** of edge examination: **only most promising *n* edges**



Model 1

User	Item	Pred.
u1	i1	4
u1	i2	2
u2	i1	3
u2	i2	3
…	…	…

Model 2

User	Item	Pred.
u1	i1	3
u1	i2	3
u2	i1	2
u2	i2	1
…	…	…

Model 5

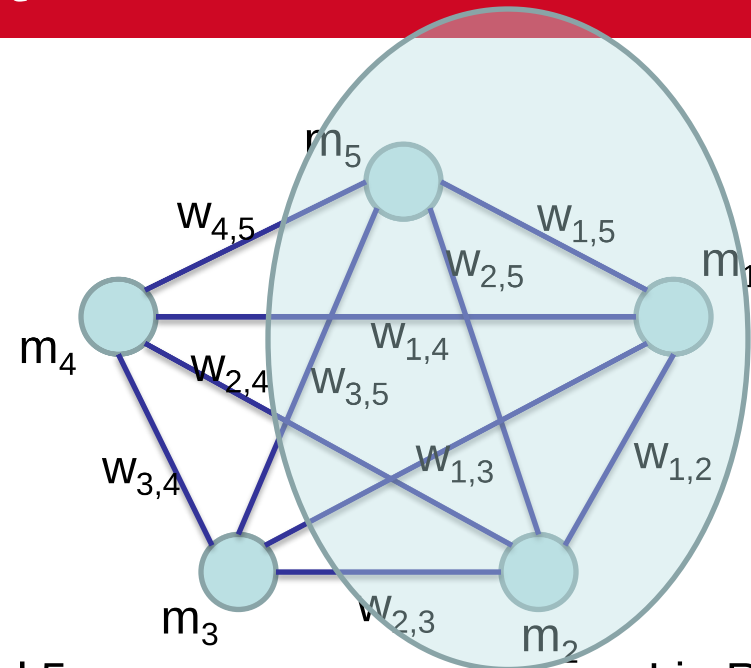
User	Item	Pred.
u1	i1	5
u1	i2	2
u2	i1	4
u2	i2	3
…	…	…

Average

User	Item	Pred.
u1	i1	4
u1	i2	2.33
u2	i1	3
u2	i2	2.33
…	…	…

RegOpt

- ***Error function***: RMSE
- ***Score function***: f_{reg}
- ***Meta-model type***: multivariate linear regression
- ***Restrictions*** of edge examination: **only most promising *n* edges**



Model 1

User	Item	Pred.
u1	i1	4
u1	i2	2
u2	i1	3
u2	i2	3
…	…	…

Model 2

User	Item	Pred.
u1	i1	3
u1	i2	3
u2	i1	2
u2	i2	1
…	…	…

Model 5

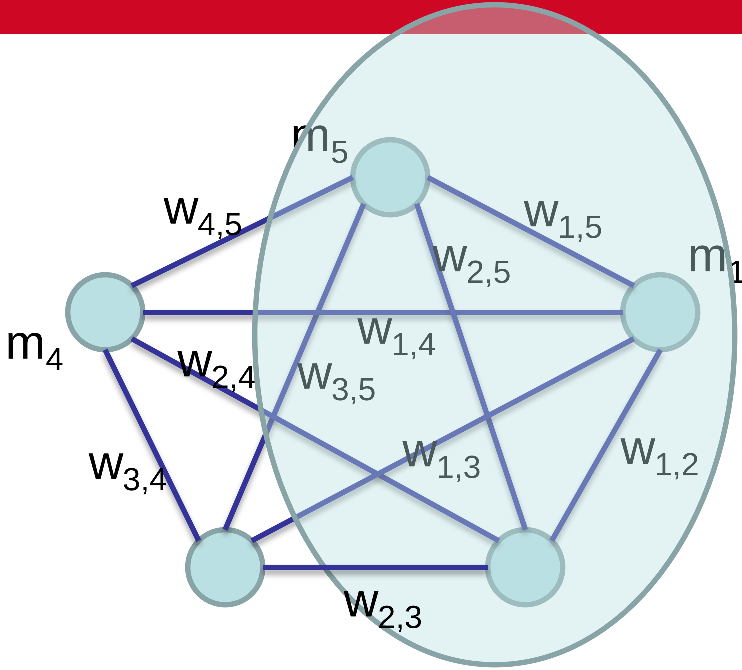
User	Item	Pred.
u1	i1	5
u1	i2	2
u2	i1	4
u2	i2	3
…	…	…

Lin.Reg

User	Item	Pred.
u1	i1	4.1
u1	i2	2.49
u2	i1	3.2
u2	i2	2.31
…	…	…

GraphOpt

- ***Error function***: RMSE
- ***Score function***: f_{gopt}
  (operates exclusively on the graph)
- ***Meta-model type***:
  multivariate linear regression
- ***Restrictions*** of edge examination: **only most promising *n* edges**



---

f_{gopt}

Require: Modelset M, Graph g

1: SumW $\leftarrow 0$

2: **for** $(\forall\{m_i, m_j\} | m_i, m_j \in M)$ **do** SumW $\leftarrow$ SumW $+\ g$.edgeWeight$(\{m_i, m_j\})$

3: **return** $\frac{\text{SumW}}{(M.\text{size})^2 * \ln(M.\text{size})}$

Experiments

Datasets	AusDM-S	AusDM-M	AusDM-L
Cases	15000	20000	50000
Models	200	250	1151
Performance (avg. RMSE, 10 folds)			
SVM-stacking best 20	871.97	872.38	876.68
Basic	869.68 (9)	868.42 (10)	871.88 (10)
EarlyStop	869.79 (10)	868.59 (10)	872.61 (10)
RegOpt	868.81 (10)	867.88 (10)	871.41 (10)
GraphOpt	870.49 (7)	868.33 (10)	870.53 (10)

Ensembling Challenge (RMSE task) of the Australian Data Mining Conference 2009
http://www.tiberius.biz/ausdm09/

Conclusion

- Large ensembles:
  selection of the right models is crucial
- We introduced a generic Graph-Based Ensemble Framework that allows to deploy various ensemble techniques
  - Basic, EarlyStop, RegOpt, GraphOpt
- Experimental evaluation:
  our approach outperforms stacking

Execution Times

- Scale-up compared to Basic
  - EarlyStop: 3.3
  - RegOpt: 1.4
  - GraphOpt: 1.6

Graph-Based Ensemble Framework

- Model-pair graph
  - each vertex corresponds to a model (in our case: model = recommender system)

Recommender System 1

User	Item	Prediction
Peter	Benjamin B.	★★★★
Peter	Australia	★★
Anna	Benjamin B.	★★★
Anna	Australia	★★★
…	…	…

Avg of Model 1 and Model 2

User	Item	Prediction
Peter	Benjamin B.	3.5
Peter	Australia	2.5
Anna	Benjamin B.	2.5
Anna	Australia	2
…	…	…

Recommender System 2

User	Item	Prediction
Peter	Benjamin B.	★★★
Peter	Australia	★★★
Anna	Benjamin B.	★★
Anna	Australia	★
…	…	…

Graph-Based Ensemble Framework

- In this framework an ensemble technique can be described by a particular choice of the
  - **Error function** (used when building the graph)
  - **Score function** used to evaluate subsets
  - **Meta model type** used at stacking
  - **Restrictions of edge** examination (we can restrict the examination to the most promising edges)
- Our ensemble techniques (Basic, EarlyStop, RegOpt, GraphOpt) are realised as different choices of the above meta-parameters

Motivation

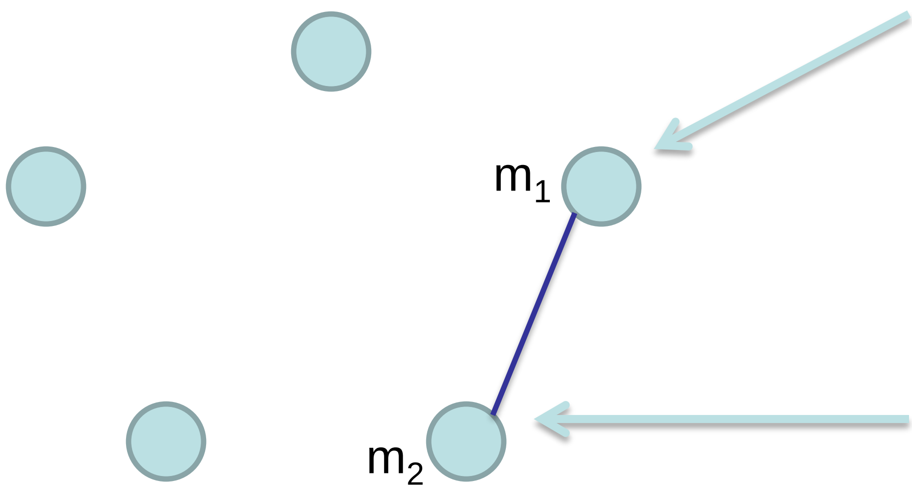
- Selection of the right models is crucial

Table 1. Performance (Root Mean Squared Error) improvement w.r.t. best individual model using simple ensemble schemes on the AusDM-S dataset. (10 fold cross validation, averaged results, in each fold the best/worst model(s) were selected based on the performances on the train subset.)

Method	RMSE-improvement
Average over all models	2.40
Average over the best 10 models	8.72
Average over the worst 10 models	−20.84

Graph-Based Ensemble Framework

- Model-pair graph
  - The weight of an edge reflects the mutual error compensation power of the models connected by that edge



Model 1

User	Item	Prediction
Peter	Benjamin B.	★★★★
Peter	Australia	★★
Anna	Benjamin B.	★★★
Anna	Australia	★★★
…	…	…

Model 2

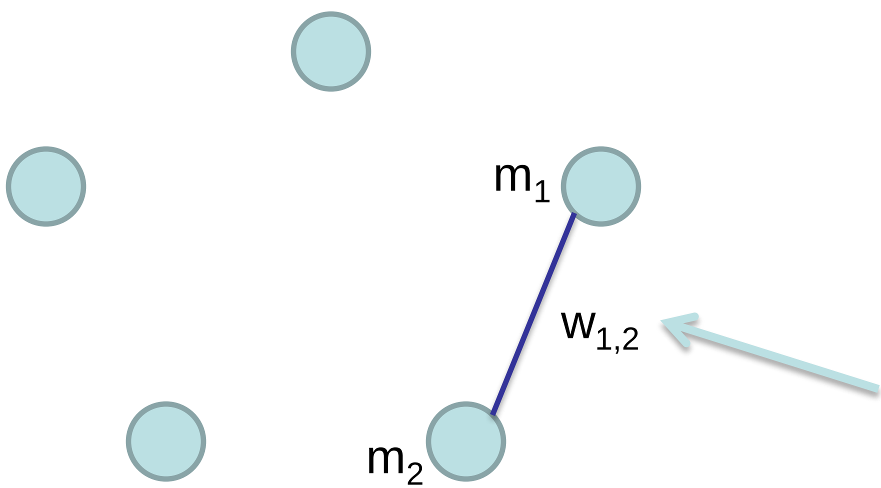
User	Item	Prediction
Peter	Benjamin B.	★★★
Peter	Australia	★★★
Anna	Benjamin B.	★★
Anna	Australia	★
…	…	…

Graph-Based Ensemble Framework

- Model-pair graph
  - The weight of an edge reflects the mutual error compensation power of the models connected by that edge

Avg. of Model 1 and Model 2

User	Item	Prediction
Peter	Benjamin B.	3.5
Peter	Australia	2.5
Anna	Benjamin B.	2.5
Anna	Australia	2
…	…	…



$w_{1,2}$ = *error* (e.g. RMSE) of the average of the both models.

(Weights are calculated using labeled **train** data.)