

# Collaborative Filtering Applied to Educational Data Mining KDD Cup 2010

presentation by: Andreas Töscher

July 25<sup>th</sup>, 2010



# BigChaos @ KDD



Michael Jahrer, Andreas Töschler from *commendo research*

# Dataset

- Algebra 2008-2009 (9 million lines)
- Bridge to Algebra 2008-2009 (20 million lines)
- Predict correct first attempt
- Similarities to collaborative filtering (CF)

		<b>KDD Cup</b>			
		Student			
Step		1		1	1
			1	0	0
		1		0	1

		<b>CF</b>			
		User			
Item		5		1	5
		4	1	3	2
			2		

# Used Information

- Student and step
- Unit and section parsed from the provided hierarchy string
- Problem name
- Knowledge components (KC)
- Opportunity count
- Problem view

# Solution Overview

- Algorithms
  - K Nearest Neighbor (KNN)
  - Singular Value Decomposition (SVD)
  - Factor Model 1 (FM1)
  - Factor Model 2 (FM2)
  - Factor Model 3 (FM3)
  - Group Factor Model (GFM)
- Neural Network for blending
- 8-fold cross validation

# Algorithms

# K Nearest Neighbor (KNN)

- Prediction is based on students with similar historic results
- Pearson correlation between students
- Use the K most similar students for prediction

## Problems

- Not all student correlations are defined with equal quality, so we use a correlation shrinkage
- Some steps are answered by no other students

$$\bar{\rho}_{s_1 s_2} = \frac{|\mathbb{I}_{s_1 s_2}| \cdot \rho_{s_1 s_2}}{|\mathbb{I}_{s_1 s_2}| + \alpha}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

$$\tilde{\rho}_{s_1 s_2} = \sigma(\delta \cdot \bar{\rho}_{s_1 s_2} + \gamma)$$

$$\tilde{c}_{is} = \frac{\sum_{\tilde{s} \in \mathbb{S}_i(s; K)} \tilde{\rho}_{s \tilde{s}} c_{i \tilde{s}}}{\sum_{\tilde{s} \in \mathbb{S}_i(s; K)} |\tilde{\rho}_{s \tilde{s}}|}$$

$$\hat{c}_{is} = \frac{\tilde{c}_{is} \sum_{\tilde{s} \in \mathbb{S}_i(s; K)} |\tilde{\rho}_{s \tilde{s}}| + \mu_s \beta}{\sum_{\tilde{s} \in \mathbb{S}_i(s; K)} |\tilde{\rho}_{s \tilde{s}}| + \beta}$$

- $\mathbb{I}_{s_1 s_2}$  ... set of steps commonly answered by student  $s_1$  and student  $s_2$
- $\hat{c}_{is}$  ... prediction for student  $s$  on step  $i$
- $\rho_{s_1 s_2}$  ... Pearson correlation between student  $s_1$  and student  $s_2$
- $\mathbb{S}_i(s; K)$  ... to student  $s$  the  $K$  most similar students who answered step  $i$
- $\alpha, \beta, \gamma, \delta$  ... meta parameters

## KNN results

Dataset	RMSE	Meta parameters
Algebra 2008-2009	0.3257	$K = 41, \alpha = 12.9, \beta = 1.5,$ $\delta = 6.2, \gamma = -1.9$
B to A 2008-2009	0.3049	$K = 41, \alpha = 12.9, \beta = 1.5,$ $\delta = 6.2, \gamma = -1.9$

- We have not used additional information  $\implies$  room for improvement
- Using step correlations does not work well

# Singular Value Decomposition (SVD)

- Sparse student by step matrix  $\mathbf{C} = [c_{is}]$ , containing the correct first attempt values
- Represent a student  $s$  by a  $N$  dimensional feature vector  $\mathbf{b}_s$
- Represent a step  $i$  by a  $N$  dimensional feature vector  $\mathbf{a}_i$

$$\hat{c}_{is} = \mathbf{a}_i^T \cdot \mathbf{b}_s$$

# Results

- Parameters trained by stochastic gradient descent with L2 regularization  $\lambda$
- Does not model biases and additional information

Dataset	RMSE	Meta parameters
Algebra 2008-2009	0.4462	$N = 10, \eta = 0.002, \lambda = 0.02$
B to A 2008-2009	0.3168	$N = 10, \eta = 0.002, \lambda = 0.02$
B to A 2008-2009	0.3159	$N = 20, \eta = 0.002, \lambda = 0.01$
B to A 2008-2009	0.3178	$N = 20, \eta = 0.002, \lambda = 0.03$

## Factor Model 1 (FM1)

$$\hat{c}_{is} = \mu + \hat{\mu}_i + \bar{\mu}_s + \tilde{\mu}_{p(i)} + \acute{\mu}_{x(i)} + \check{\mu}_{u(i)} \\ + \frac{1}{\sqrt{|\mathbb{K}(i, s)|}} \sum_{k \in \mathbb{K}(i, s)} (\check{\mu}_k + \acute{\mu}_{ks}) \\ + \left( \mathbf{a}_i + \frac{1}{\sqrt{|\mathbb{K}(i, s)|}} \sum_{k \in \mathbb{K}(i, s)} \acute{\mathbf{a}}_k \right)^T \cdot \mathbf{b}_s$$

- $\mu$  ... global bias
- $\hat{\mu}_i$  ... bias for step  $i$
- $\bar{\mu}_s$  ... bias for student  $s$
- $\tilde{\mu}_{p(i)}$  ... bias for problem  $p(i)$
- $\acute{\mu}_{x(i)}$  ... bias for section  $x(i)$
- $\check{\mu}_{u(i)}$  ... bias for unit  $u(i)$
- $\mathbb{K}(i, s)$  ... set of knowledge components
- $\acute{\mathbf{a}}_k$  ...  $N$  dimensional KC feature vector

# FM results

- Significant improvement over SVD
- Models: unit, section, knowledge component
- FM2, FM3 and GFM contain further improvements

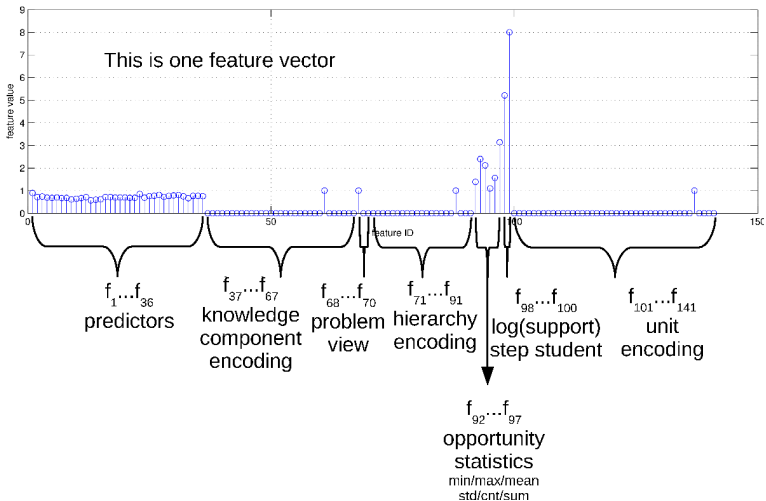
Dataset	RMSE	Meta parameters
Algebra 2008-2009	0.3078	$N = 50, \eta = 0.0005, \lambda = 0.01$
B to A 2008-2009	0.3013	$N = 50, \eta = 0.0005, \lambda = 0.01$

# Blending

# Blending

- 2 layer neural network
  - 80 sigmoid units on the first and second hidden layer
- 36 predictors for *Algebra 2008-2009*
- 37 predictors for *Bridge to Algebra 2008-2009*
- Additional information

# Feature Generation



# Results

Best individual results:

	Cross validation	Leaderboard
Algebra 2008-2009	0.2997	0.2849
Bridge to Algebra 2008-2009	0.2898	0.2837

Neural network blending:

	Cross validation	Leaderboard
Algebra 2008-2009	0.280925	0.277355
Bridge to Algebra 2008-2009	0.288417	0.28073

# Summary

- Ideas from collaborative filtering are suitable for the KDD Cup 2010
- An ensemble of different models is a good approach for a competition
- For sure there is a way to build simpler models with similar or better performance

# Thanks for your attention

The logo for commendo, featuring the word "commendo" in a blue, lowercase, sans-serif font. A light blue swoosh underline is positioned beneath the letters "m", "e", "n", and "d".

commendo research & consulting GmbH  
Austria