

KDD CUP

Making better predictions based on
competitors results

Kuan-Liang Liu, Petr Jarušek
Summer school

What is KDD?

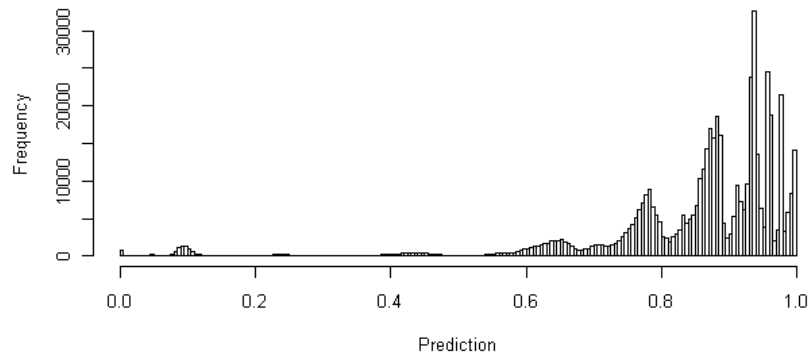
- Online competition
- Goal: develop a learning model based on training datasets (comes from tutor systems)
- Teams submit their predictions

OUR GOAL:

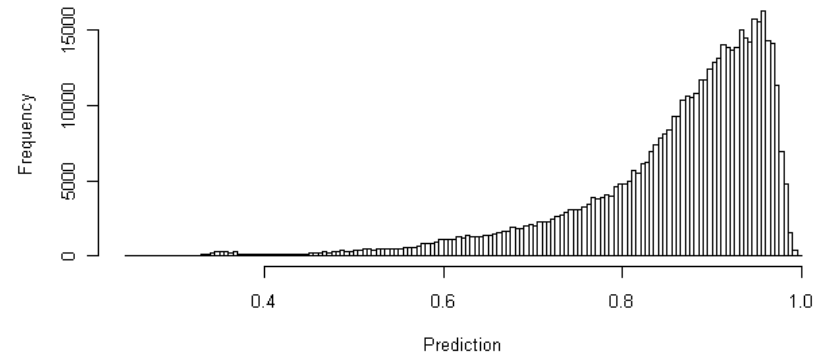
- Combine results of top 20 teams to obtain better results in results list

Teams' predictions are various

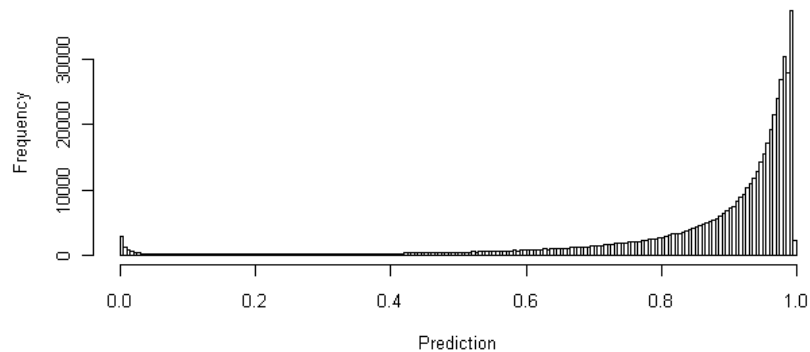
Team: nm3456



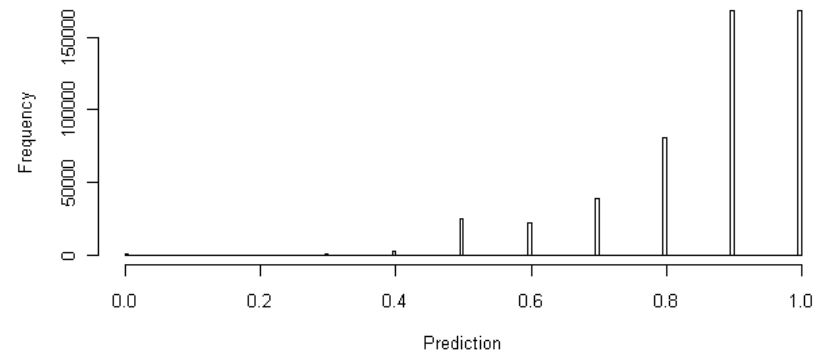
Team: R26am



Team: rofu_3739

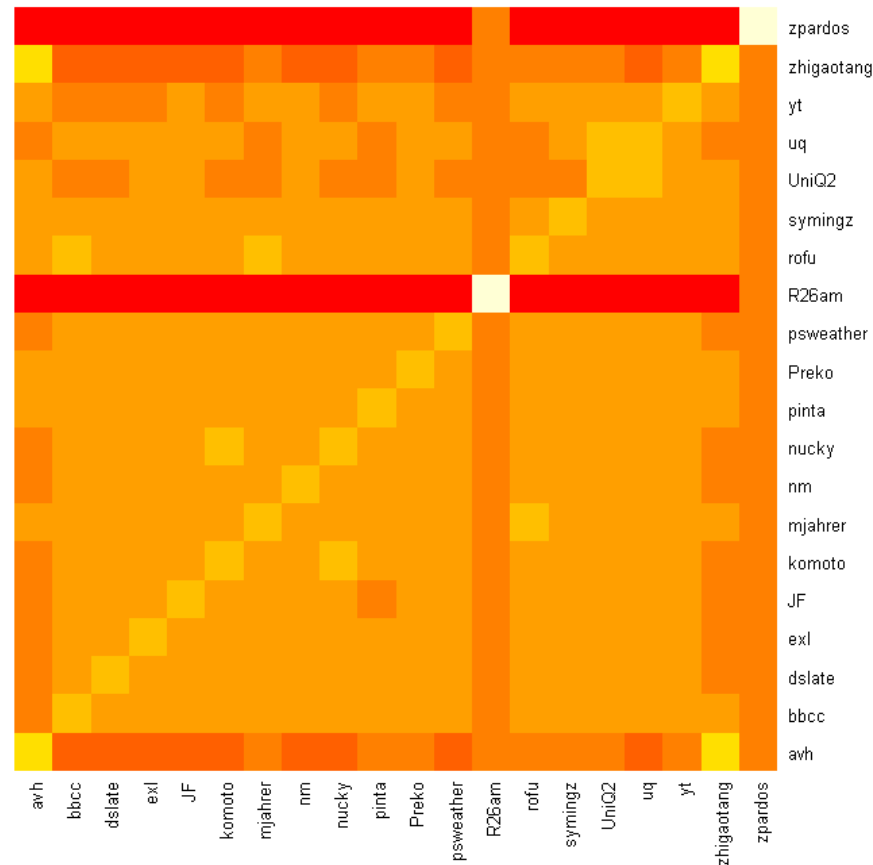


Team: zhigaotang_3149



Correlations between top 20 teams' predictions

Pearson correlation for predictions of top 20 teams



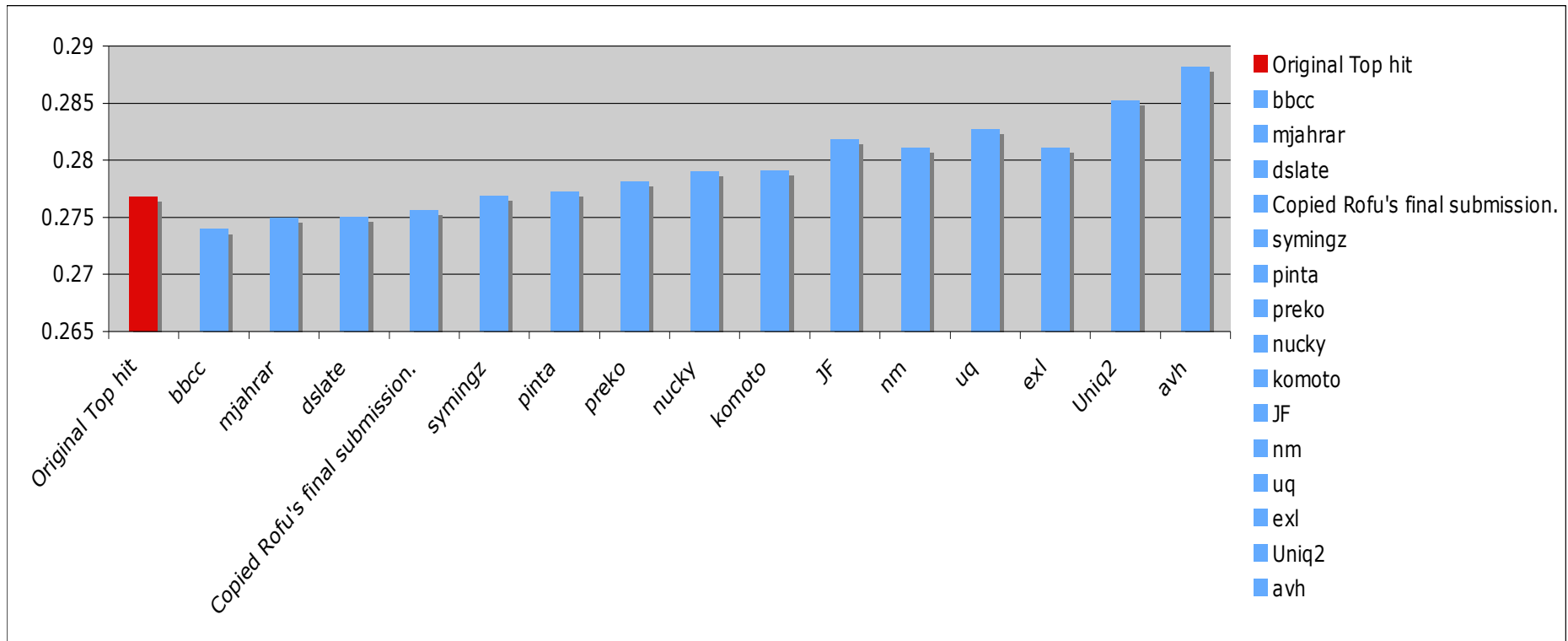
Our method

- Method 1: take the best team and combine it subsequently with others
- Method 2: take the best team and combine it subsequently with 2th, 2th and 3th, 2th and 3th and 4th ...
- Both: Weight teams values according to the obtained scores on leaderboard

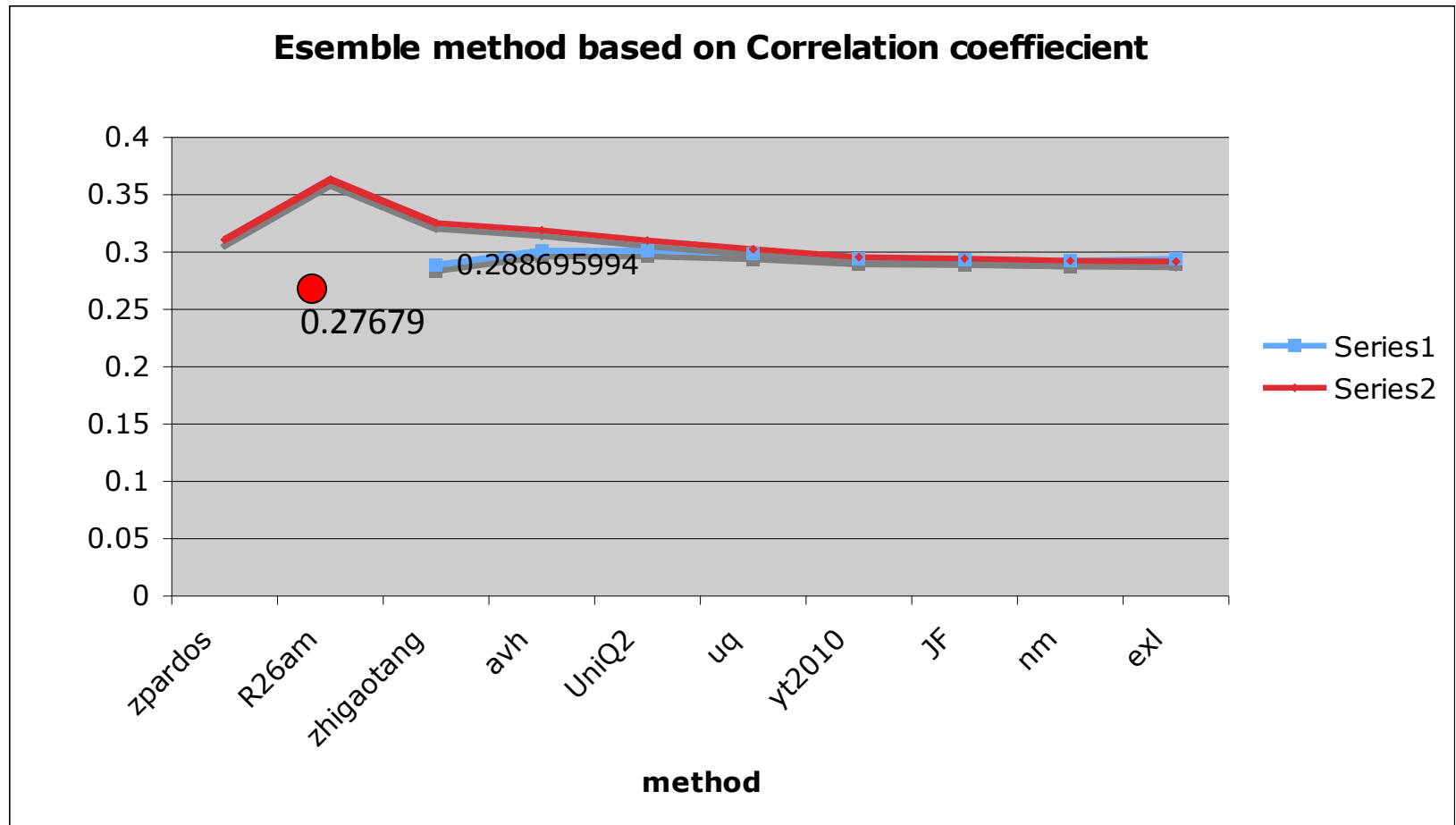
Pair combination to top 1

<u>Overall Rank</u> ▲	<u>Individual/Team Name</u>	<u>Algebra I 2008-2009</u>	<u>Bridge to Algebra 2008-2009</u>	<u>Total Score</u>	<u>Date</u>
1	NCKU-IM	0.274933	0.277428	0.27618	2010-07-15 09:53:44
2	Zhang and Su	0.275893	0.277687	0.27679	2010-06-08 23:06:06
3	Zhang and Su	0.275893	0.277687	0.27679	2010-06-08 23:39:35
4	Zhang and Su	0.275893	0.277691	0.276792	2010-06-08 23:22:22
5	Zhang and Su	0.275898	0.277687	0.276793	2010-06-08 23:35:38
6	Zhang and Su	0.275893	0.277694	0.276793	2010-06-08 22:48:45
7	Zhang and Su	0.275908	0.277687	0.276797	2010-06-08 23:37:58
8	Zhang and Su	0.275893	0.277703	0.276798	2010-06-08 23:30:14
9	Zhang and Su	0.275893	0.277707	0.2768	2010-06-08 23:03:36
10	Zhang and Su	0.275893	0.277712	0.276802	2010-06-08 23:18:58

Pair combine



Gradually combination to top 1



Possible extensions

- Weight teams results according to correlation and to get maximum diverse results dataset
- Focus on teams' strategies, combine these method instead of combining results