

Preliminary evaluation report

Performance (accuracy) of all groups on training and testing data set both on beetle and seb

	beetle						seb					
	train			test			train			test		
team	5way	3way	2way	5way	3way	2way	5way	3way	2way	5way	3way	2way
baseline	0.727	0.742	0.849	0.552	0.581	0.762	0.761	0.813	0.847	0.445	0.534	0.619
team1	0.202	0.311	0.58	0.181	0.298	0.581	0.268	0.49	0.586	0.301	0.505	0.597
team2	0.379	0.449	0.579	0.356	0.421	0.587	0.352	0.508	0.58	0.334	0.509	0.585
team3	0.466	0.466	0.621	0.513	0.513	0.667	0.475	0.542	0.576	0.48	0.56	0.603
team4	0.207	0.336	0.545	0.162	0.308	0.562	0.201	0.395	0.549	0.197	0.403	0.57
team5	0.588	0.605	0.748	0.604	0.633	0.746	0.469	0.529	0.555	0.518	0.638	0.696
team6	0.665	0.688	0.792	0.785	0.808	0.89	0.617	0.668	0.69	0.506	0.559	0.581
team7	0.681	0.701	0.817	0.792	0.813	0.879	0.837	0.881	0.901	0.936	0.953	0.961

Wencan's Observations:

- The baseline seems be overfitting. It performs well on the training, but very badly on the test sets
- No one beats the baseline on the training of beetle