3 Supplemental Materials

Table 1: Top 15 words and pairs

Top 15 words for S				Top 15 pairs for S^{10}			
w	P_{TP}	P_{TN}	S	w_i, w_j	P_{TP}	P_{TN}	S^{10}
interact	0.76	0.12	0.64	with,interact	0.31	0.03	0.28
bind	0.63	0.14	0.49	interact,protein	0.21	0.02	0.19
domain	0.52	0.08	0.44	with,protein	0.25	0.12	0.13
complex	0.46	0.15	0.31	with,domain	0.14	0.01	0.13
between	0.01	0.29	0.28	interact,domain	0.13	0.01	0.12
with	0.9	0.65	0.25	bind,protein	0.15	0.03	0.12
activ	0.56	0.32	0.24	interact,between	0.12	0.01	0.11
yeast	0.28	0.04	0.24	protein–domain	0.12	0.01	0.11
between	0.38	0.16	0.22	bind–domain	0.11	0.01	0.1
associ	0.35	0.13	0.22	bind–with	0.11	0.01	0.1
protein	0.86	0.64	0.22	with–complex	0.12	0.02	0.1
region	0.26	0.06	0.2	associ–with	0.14	0.05	0.09
suggest	0.45	0.25	0.2	thi–interact	0.09	0.01	0.08
function	0.48	0.28	0.2	with–activ	0.1	0.04	0.06
regul	0.38	0.19	0.19	activ–protein	0.1	0.04	0.06

Table 2: IAS Results

	Run 1 (SVM)	Run 2 (VTT)	Run 3 (SVD/Int)	Mean*	StDev*	Median*
Total Predictions		750				
Total Positive		375				
Total Negative		375				
True Positives (TP)	330	295	300			
False Positives(FP)	186	118	143			
True Negatives (TN)	189	257	232			
False Negatives (FN)	45	80	75			
Precision	0.64	0.71	0.68	0.66	0.08	0.68
Recall	0.88	0.79	0.8	0.76	0.19	0.85
Accuracy	0.69	0.74	0.71	0.67	0.06	0.67
F-Score	0.74	0.75	0.73	0.69	0.10	0.72
FP Rate	0.5	0.32	0.38			
TP Rate	0.88	0.79	0.8			
Error Rate	0.31	0.26	0.29			
AUC	0.8	0.76	0.71	0.74	0.07	0.75
	*calculated from 51 runs submitted by 19 teams					

Table 3: Top 10 stemmed word features from training sentences

$rank$	feature
1	with
2	protein
3	cell
4	interact
5	bind
6	activ
7	express
8	complex
9	dna
10	human

Table 4: ISS Results

	Average	Run 1	Run 2	Run 3
Predicted Passages	6213.54	18385	18371	18371
Pred. Unique Passages	3429.65	5156	5270	5252
Matches	207.46	360	376	387
Unique Matches	128.61	131	145	156
Fraction Correct	0.047	0.02	0.021	0.021
Frac. Unique Correct	0.047	0.025	0.028	0.03
Mean Rec. Rank	0.557	0.659	0.625	0.642

Table 5: Summary of results of IPS task, including our three submissions and the central tendency values for all submissions. Results are shown both for the identification of protein-protein interaction pairs (PPIN) and single interactors (PN), as well as for the set of all articles (All) and the subset of articles containing exclusively SwissProt interaction pairs (SP).

		PN All			PN SP		
		Precision	Recall	F-score	Precision	Recall	F-score
Articles	Run 1	0.125	0.419	0.173	0.141	0.444	0.192
	Run 2	0.125	0.419	0.173	0.141	0.444	0.192
	Run 3	0.127	0.418	0.174	0.143	0.442	0.193
	Mean	0.159	0.287	0.175	0.174	0.306	0.192
	Stdv	0.08	0.125	0.068	0.085	0.131	0.076
	Median	0.155	0.278	0.173	0.171	0.306	0.192
Articles with Predictions	Run 1	0.161	0.54	0.222	0.178	0.561	0.242
	Run 2	0.161	0.54	0.222	0.178	0.561	0.242
	Run 3	0.164	0.543	0.225	0.183	0.565	0.246
	Mean	0.228	0.379	0.244	0.248	0.405	0.267
	Stdv	0.137	0.131	0.11	0.143	0.138	0.119
	Median	0.21	0.385	0.236	0.233	0.414	0.257
Overall Interactor	Run 1	0.122	0.424	0.189	0.088	0.446	0.147
	Run 2	0.122	0.424	0.189	0.088	0.446	0.147
	Run 3	0.123	0.423	0.19	0.089	0.445	0.148
	Mean	0.203	0.2648	0.198	0.15	0.283	0.171
	Stdv	0.129	0.123	0.08	0.096	0.129	0.076
	Median	0.182	0.252	0.192	0.134	0.272	0.168

|		PPI All			PPI SP		
		Precision	Recall	F-score	Precision	Recall	F-score
Overall	Run 1	0.033	0.221	0.057	0.035	0.235	0.061
	Run 2	0.033	0.221	0.057	0.035	0.235	0.061
	Run 3	0.034	0.221	0.059	0.036	0.235	0.062
	Mean	0.102	0.115	0.085	0.094	0.106	0.078
	Stdv	0.094	0.076	0.055	0.088	0.07	0.051
	Median	0.065	0.118	0.077	0.061	0.11	0.071
Average	Run 1	0.051	0.275	0.072	0.056	0.285	0.077
	Run 2	0.051	0.275	0.072	0.056	0.285	0.077
	Run 3	0.052	0.278	0.073	0.057	0.288	0.078
	Mean	0.116	0.2	0.113	0.106	0.186	0.113
	Stdv	0.104	0.106	0.084	0.095	0.1	0.084
	Median	0.081	0.216	0.084	0.076	0.196	0.084
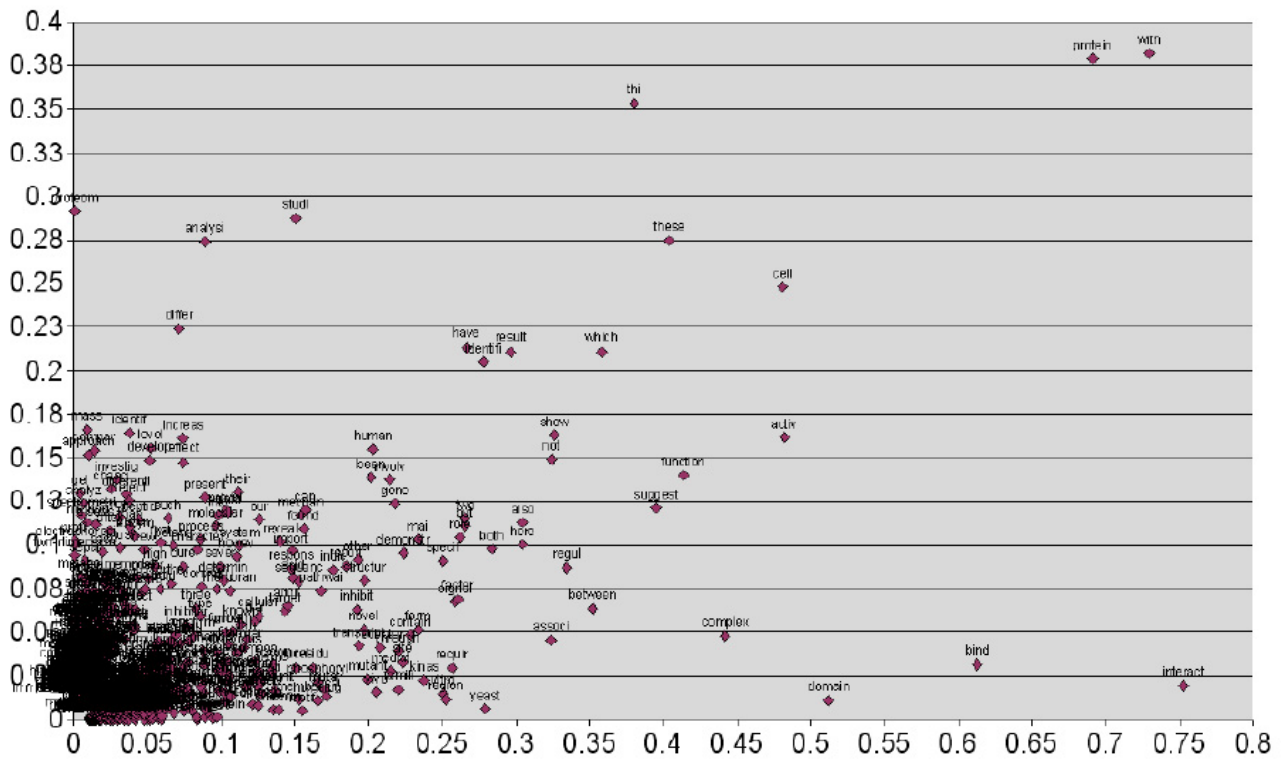
Figure 3: Top 1000 words in the space of P_{TP} (horizontal axis) and $P_{TN}(vertic laxis)$. Words with high S tend to be almost exclusively associated with positive or negative abstracts.
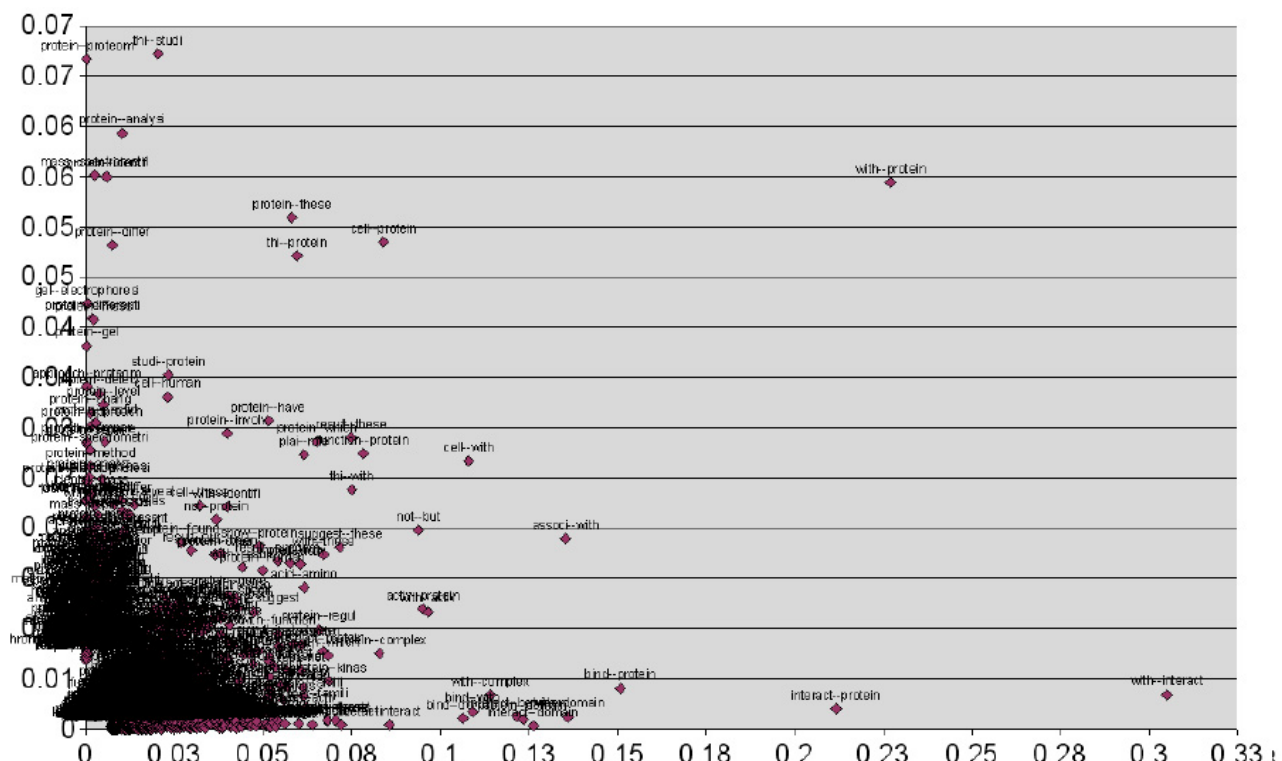

Figure 4: Top 1800 word pairs (in 10 word wondows) in the space of $P_{TP}(w_i, w_j)$ and $P_{TN}(w_i, w_j)$. Word pairs with high S^{10} tend to be almost exclusively associated with positive or negative abstracts.
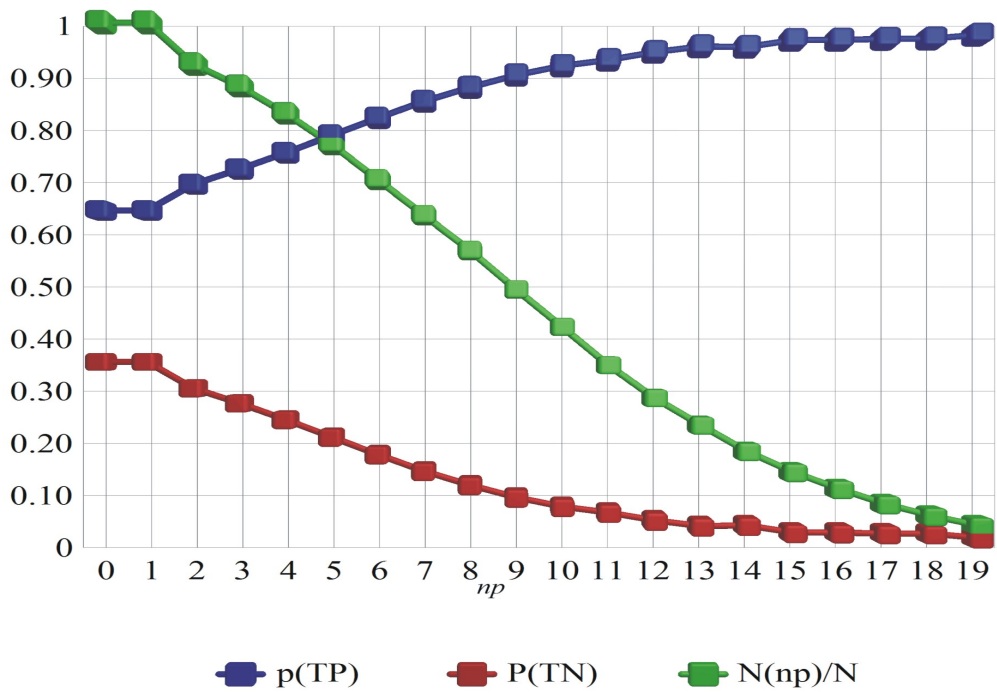
Figure 5: Probability that an abstract, from the learning set, with np or more protein mentions is positive (blue). Same for negative (red). Proportion of abstracts with np or more protein mentions (green).
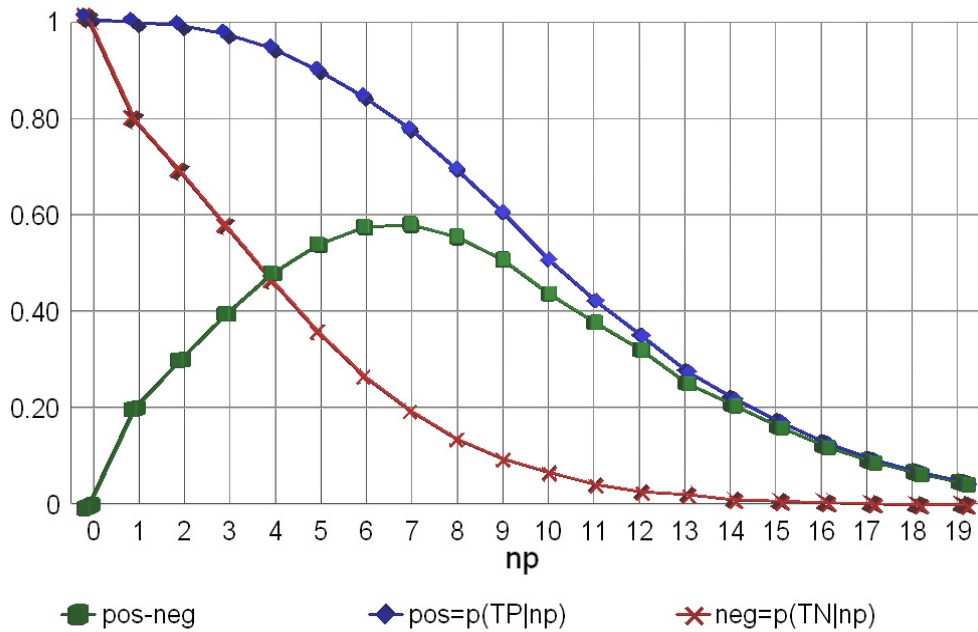
Figure 6: Probability that a positive abstract in the learning set has more than np protein mentions (blue). Same for negative (red). Difference (green).
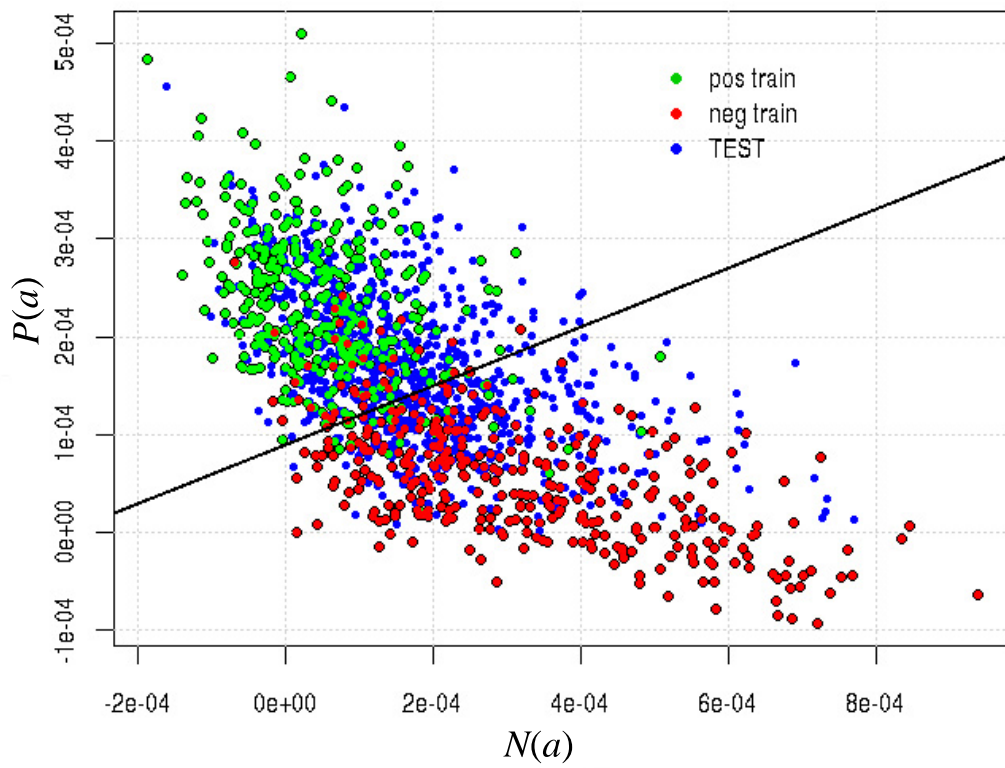
Figure 7: Decision boundary for vector/SVD model on the space of positive and negative scores P and N. Red and green dots represent negative and positive abstracts in the learning data set, respectively. Blue dots represent test set abstracts.
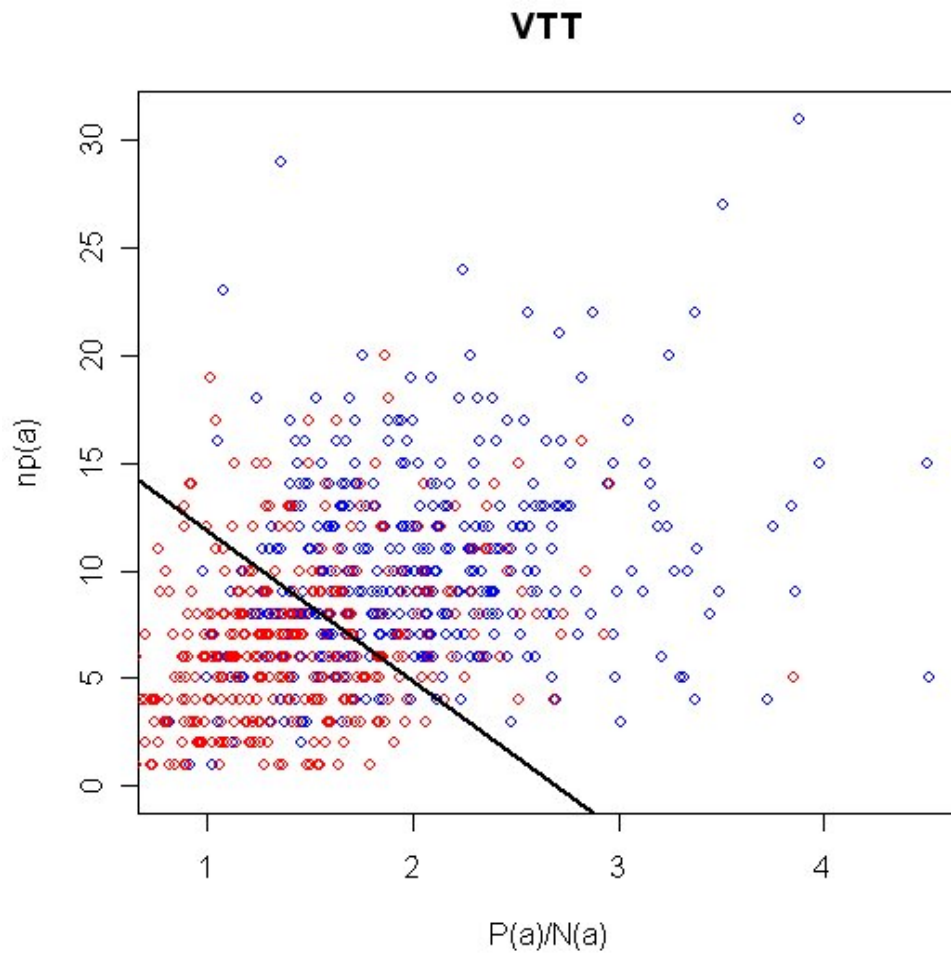
Figure 8: Decision boundary for VTT method on the space of the ratio of positive to negative scores $P(a)/N(a)$ and the number of protein mentions $np(a)$. Red and blue circles represent negative and positive abstracts in the test data set, respectively.
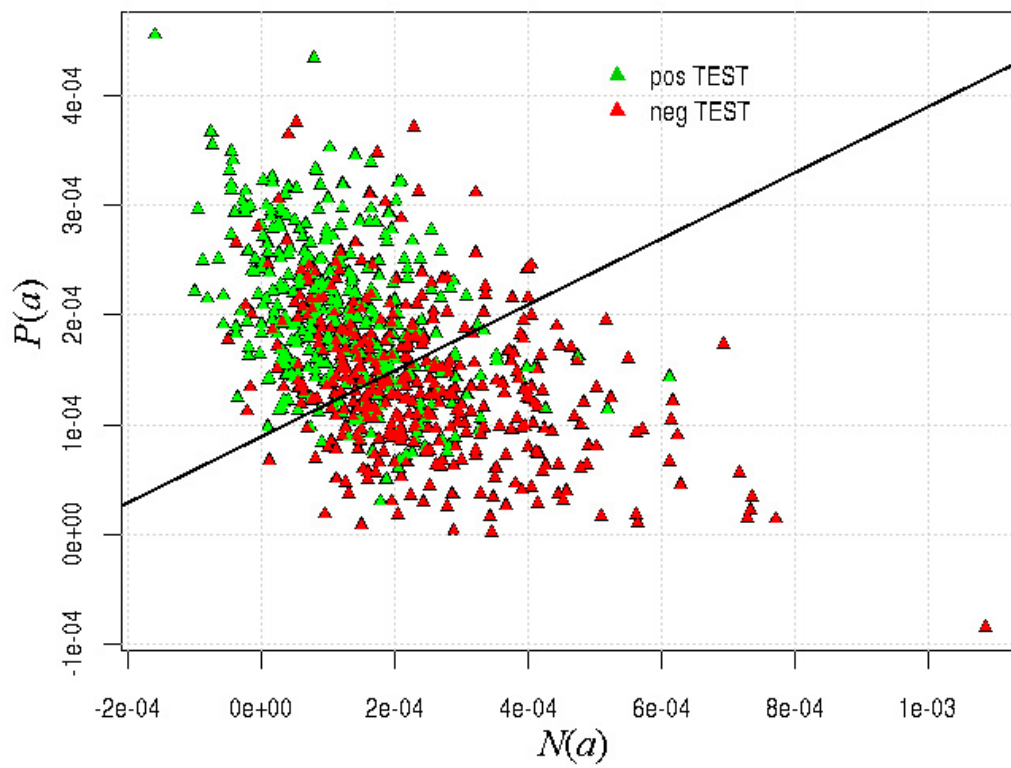
Figure 9: Decision boundary for vector/SVD model on the space of positive and negative scores P and N. Red and green triangles represent negative and positive abstracts in the test data set, respectively.