Readme
Yao Yan
1669598

Expt_id	Rare thres	Feat thres	Training accuracy	Test accuracy	# of features	# of kept features	Running time
1-1	1	1	0.8133947600	0.6757860351	49204	21520	3m59.285s
1-3	1	3	0.8105530559	0.6757860351	49204	14701	2m53.094s
2-3	2	3	0.8232546123	0.6872192731	50965	14948	4m21.366s
3-5	3	5	0.8377860541	0.6986525112	51772	12604	5m27.570s
5-10	5	10	0.8440937762	0.7207023274	52552	10410	4m58.920s

The conclusion: once I increase the threshold for rare words and features, less features are used for training, and meanwhile there is an increase in the train and test accuracy. But this trend is not unlimited, a decent amount features are needed to guarantee the model's performance.