Full Tables for the Report Text Classification using String Kernels

January 2015

Topic	Kernel	Length	Weight Decay	F1	Precision	Recall
acq	Sub-Sequence	3	0.5000	0.8214	0.7419	0.9200
acq	Sub-Sequence	4	0.5000	0.8148	0.7586	0.8800
acq	Sub-Sequence	5	0.5000	0.8302	0.7857	0.8800
acq	Sub-Sequence	6,7	0.5000	0.8627	0.8462	0.8800
acq	Sub-Sequence	8	0.5000	0.8163	0.8333	0.8000
acq	Sub-Sequence	10	0.5000	0.6842	1.0000	0.5200
acq	Sub-Sequence	12	0.5000	0.1481	1.0000	0.0800
acq	Sub-Sequence	14	0.5000	0.0000	1	0.0000
acq	n-grams kernel	3	None	0.7778	0.7241	0.8400
acq	n-grams kernel	4	None	0.8000	0.8000	0.8000
acq	n-grams kernel	5	None	0.8302	0.7857	0.8800
acq	n-grams kernel	6	None	0.8462	0.8148	0.8800
acq	n-grams kernel	7,8	None	0.8627	0.8462	0.8800
acq	n-grams kernel	10	None	0.8095	1.0000	0.6800
acq	n-grams kernel	12	None	0.6111	1.0000	0.4400
acq	n-grams kernel	14	None	0.1481	1.0000	0.0800
acq	Word Kernel (WK)	None	None	**0.9020**	0.8846	0.9200
corn	Sub-Sequence	3	0.5000	0.5714	1.0000	0.4000
corn	Sub-Sequence	4	0.5000	0.8235	1.0000	0.7000
corn	Sub-Sequence	5	0.5000	**0.8889**	1.0000	0.8000
corn	Sub-Sequence	6	0.5000	0.8235	1.0000	0.7000
corn	Sub-Sequence	7,8	0.5000	0.7500	1.0000	0.6000
corn	Sub-Sequence	10	0.5000	0.3333	1.0000	0.2000
corn	Sub-Sequence	12,14	0.5000	0.0000	1	0.0000
corn	n-grams kernel	3	None	0.8235	1.0000	0.7000
corn	n-grams kernel	4-6	None	**0.8889**	1.0000	0.8000
corn	n-grams kernel	7,8	None	0.7500	1.0000	0.6000
corn	n-grams kernel	10	None	0.5714	1.0000	0.4000
corn	n-grams kernel	12	None	0.3333	1.0000	0.2000
corn	n-grams kernel	14	None	0.0000	1	0.0000
corn	Word Kernel (WK)	None	None	**0.8889**	1.0000	0.8000

Table 1: Varying values for length parameter. Classification of documents of acq and corn topics.

Topic	Kernel	Length	Weight Decay	F1	Precision	Recall
crude	Sub-Sequence	3	0.5000	0.7692	0.9091	0.6667
crude	Sub-Sequence	4	0.5000	0.8148	0.9167	0.7333
crude	Sub-Sequence	5,6	0.5000	0.8462	1.0000	0.7333
crude	Sub-Sequence	7	0.5000	0.8000	1.0000	0.6667
crude	Sub-Sequence	8	0.5000	0.6364	1.0000	0.4667
crude	Sub-Sequence	10-14	0.5000	0.3333	1.0000	0.2000
crude	n-grams kernel	3,4	None	**0.8889**	1.0000	0.8000
crude	n-grams kernel	5,6	None	0.8462	1.0000	0.7333
crude	n-grams kernel	7	None	0.8000	1.0000	0.6667
crude	n-grams kernel	8	None	0.6957	1.0000	0.5333
crude	n-grams kernel	10	None	0.6364	1.0000	0.4667
crude	n-grams kernel	12,14	None	0.3333	1.0000	0.2000
crude	Word Kernel (WK)	None	None	0.8000	1.0000	0.6667
earn	Sub-Sequence	3	0.5000	**0.8406**	1.0000	0.7250
earn	Sub-Sequence	4,5	0.5000	0.8286	0.9667	0.7250
earn	Sub-Sequence	6	0.5000	0.8116	0.9655	0.7000
earn	Sub-Sequence	7	0.5000	0.7761	0.9630	0.6500
earn	Sub-Sequence	8	0.5000	0.7500	1.0000	0.6000
earn	Sub-Sequence	10	0.5000	0.6207	1.0000	0.4500
earn	Sub-Sequence	12	0.5000	0.5965	1.0000	0.4250
earn	Sub-Sequence	14	0.5000	0.5714	1.0000	0.4000
earn	n-grams kernel	3-5	None	**0.8406**	1.0000	0.7250
earn	n-grams kernel	6	None	0.8235	1.0000	0.7000
earn	n-grams kernel	7	None	0.8060	1.0000	0.6750
earn	n-grams kernel	8	None	0.7692	1.0000	0.6250
earn	n-grams kernel	10	None	0.6667	1.0000	0.5000
earn	n-grams kernel	12,14	None	0.5965	1.0000	0.4250
earn	Word Kernel (WK)	None	None	0.8235	1.0000	0.7000

Table 2: Varying values for length parameter. Classification of documents of crude and earn topics.

Topic	Kernel	Length	Weight Decay	F1	Precision	Recall
acq	Sub-Sequence	5	0.01-0.5	0.8302	0.7857	0.8800
acq	Sub-Sequence	5	0.7000	0.7925	0.7500	0.8400
acq	Sub-Sequence	5	0.9000	0.7451	0.7308	0.7600
acq	n-grams kernel	5	None	0.8302	0.7857	0.8800
acq	Word Kernel (WK)	None	None	**0.9020**	0.8846	0.9200
corn	Sub-Sequence	5	0.01-0.5	0.8889	1.0000	0.8000
corn	Sub-Sequence	5	0.7000	0.7500	1.0000	0.6000
corn	Sub-Sequence	5	0.9000	0.3333	1.0000	0.2000
corn	n-grams kernel	5	None	0.8889	1.0000	0.8000
corn	Word Kernel (WK)	None	None	0.8889	1.0000	0.8000
crude	Sub-Sequence	5	0.01-0.5	0.8462	1.0000	0.7333
crude	Sub-Sequence	5	0.7000	0.7692	0.9091	0.6667
crude	Sub-Sequence	5	0.9000	0.6087	0.8750	0.4667
crude	n-grams kernel	5	None	0.8462	1.0000	0.7333
crude	Word Kernel (WK)	None	None	0.8000	1.0000	0.6667
earn	Sub-Sequence	5	0.01-0.7	0.8286	0.9667	0.7250
earn	Sub-Sequence	5	0.9000	0.7500	1.0000	0.6000
earn	n-grams kernel	5	None	0.8406	1.0000	0.7250
earn	Word Kernel (WK)	None	None	0.8235	1.0000	0.7000

Table 3: Varying values for the Weight Decay (λ) parameter.

4