

Supplementary Materials for:
**The Influence of Feature Representation of Text
on the Performance of Document Classification**

February 20, 2019

Sanda Martinčić-Ipšić, Tanja Miličić,
smarti@uniri.hr
University of Rijeka, Department of Informatics, Rijeka, Croatia
Ljupčo Todorovski
Ljupco.Todorovski@fu.uni-lj.si
University of Ljubljana, Faculty of Administration
Institute Jožef Stefan, Department of Knowledge Technologies,
Ljubljana Slovenia

1 Results for Reuters8 data set

Accuracy	Per class Precision	Per class Recall	Per class F1	Per class AUC ROC	Per class all
tf-idf	tf-idf	tf	tf-idf	tf-idf	tf
tf	tf	tf-idf	tf	tf	tf-idf
word2vec_75	doc2vec_25	doc2vec_100	doc2vec_25	doc2vec_25	doc2vec_100
word2vec_100	word2vec_100	doc2vec_25	word2vec_100	word2vec_100	doc2vec_25
word2vec_50	word2vec_25	doc2vec_50	word2vec_50	word2vec_50	doc2vec_50
word2vec_25	word2vec_50	doc2vec_75	word2vec_75	word2vec_75	doc2vec_75
doc2vec_25	word2vec_75	word2vec_100	doc2vec_50	doc2vec_1000	word2vec_100
doc2vec_50	doc2vec_50	word2vec_25	word2vec_25	doc2vec_50	word2vec_25
doc2vec_75	doc2vec_75	word2vec_50	doc2vec_75	word2vec_25	word2vec_50
doc2vec_100	doc2vec_100	word2vec_75	doc2vec_100	word2vec_100	word2vec_75
doc2vec_1000	doc2vec_1000	doc2vec_1000	doc2vec_1000	doc2vec_200	doc2vec_1000
doc2vec_200	doc2vec_200	doc2vec_200	doc2vec_200	doc2vec_500	doc2vec_200
doc2vec_500	doc2vec_500	doc2vec_500	doc2vec_500	doc2vec_75	doc2vec_500
gow_histogram	gow_avg	gow_avg	gow_avg	gow_histogram	gow_avg
gow_quantiles	gow_histogram	gow_histogram	gow_histogram	gow_quantiles	gow_histogram
gow_avg	gow_quantiles	gow_quantiles	gow_quantiles	gow_avg	gow_quantiles

Table 1: Ranks of document representation models for Reuters8 dataset in terms of accuracy, per class precision, per class recall, per class F1 and per class AUC ROC

Features	Class	Precision	Recall	F1	AUC ROC	Accuracy
tf	Acq	0.95	0.94	0.95	0.99	89.43%
	Crude	0.81	0.81	0.81	0.99	
	Earn	0.99	0.96	0.97	1.00	
	Grain	0.82	0.83	0.82	0.99	
	Interest	0.47	0.73	0.57	0.98	
	Money-fx	0.83	0.67	0.75	0.98	
	Trade	0.47	0.71	0.57	0.98	
tf-idf	Acq	0.95	0.97	0.96	1.00	91.20%
	Crude	0.89	0.79	0.84	0.99	
	Earn	0.99	0.96	0.98	1.00	
	Grain	0.87	0.91	0.89	1.00	
	Interest	0.49	0.86	0.62	0.99	
	Money-fx	0.93	0.70	0.80	0.99	
	Trade	0.44	0.76	0.56	0.98	

Table 2: Results for 8 classes of Reuters8 data set for bag-of-words document representation models

Features	Class	Precision	Recall	F1	AUC ROC	Accuracy
25	Acq	0.95	0.93	0.94	0.99	87.86%
	Crude	0.58	0.80	0.67	0.98	
	Earn	0.98	0.96	0.97	1.00	
	Grain	0.81	0.82	0.81	0.99	
	Interest	0.42	0.80	0.55	0.98	
	Money-fx	0.84	0.66	0.74	0.98	
	Trade	0.64	0.58	0.61	0.97	
50	Acq	0.95	0.93	0.94	0.99	88.35%
	Crude	0.64	0.79	0.71	0.98	
	Earn	0.98	0.96	0.97	1.00	
	Grain	0.79	0.87	0.82	0.99	
	Interest	0.49	0.75	0.59	0.98	
	Money-fx	0.78	0.67	0.72	0.98	
	Trade	0.73	0.60	0.66	0.97	
75	Acq	0.96	0.93	0.94	0.99	88.87%
	Crude	0.63	0.83	0.72	0.98	
	Earn	0.98	0.97	0.98	1.00	
	Grain	0.83	0.83	0.83	0.99	
	Interest	0.49	0.77	0.60	0.98	
	Money-fx	0.80	0.68	0.74	0.98	
	Trade	0.69	0.60	0.64	0.97	
100	Acq	0.95	0.93	0.94	0.99	88.86%
	Crude	0.64	0.80	0.71	0.98	
	Earn	0.98	0.96	0.97	1.00	
	Grain	0.81	0.83	0.82	0.99	
	Interest	0.50	0.79	0.61	0.98	
	Money-fx	0.83	0.68	0.75	0.98	
	Trade	0.63	0.60	0.61	0.97	

Table 3: Results for 8 classes of Reuters8 data set for word2vec document representation models with feature vectors of size 25, 50, 75 and 100

Features	Class	Precision	Recall	F1	AUC ROC	Accuracy
25	Acq	0.91	0.92	0.91	0.99	86.17%
	Crude	0.68	0.79	0.73	0.99	
	Earn	0.98	0.90	0.94	0.99	
	Grain	0.85	0.87	0.86	1.00	
	Interest	0.37	0.77	0.50	0.97	
	Money-fx	0.80	0.67	0.73	0.98	
	Trade	0.42	0.60	0.49	0.98	
50	Acq	0.89	0.87	0.88	0.99	82.11%
	Crude	0.54	0.81	0.65	0.98	
	Earn	0.97	0.83	0.90	0.99	
	Grain	0.81	0.83	0.82	0.99	
	Interest	0.33	0.77	0.46	0.97	
	Money-fx	0.64	0.67	0.66	0.97	
	Trade	0.34	0.61	0.43	0.96	
75	Acq	0.85	0.80	0.82	0.98	77.33%
	Crude	0.40	0.82	0.54	0.98	
	Earn	0.98	0.77	0.86	0.98	
	Grain	0.77	0.85	0.81	0.99	
	Interest	0.14	0.82	0.24	0.97	
	Money-fx	0.61	0.64	0.63	0.97	
	Trade	0.10	0.64	0.17	0.95	
100	Acq	0.82	0.75	0.79	0.97	73.35%
	Crude	0.34	0.80	0.48	0.96	
	Earn	0.97	0.72	0.83	0.97	
	Grain	0.58	0.79	0.67	0.97	
	Interest	0.11	0.88	0.19	0.95	
	Money-fx	0.55	0.64	0.59	0.96	
	Trade	0.07	0.60	0.12	0.96	

Table 4: Results for 8 classes of Reuters8 data set for doc2vec document representation models with feature vectors of sizes 25, 50, 75 and 100

Features	Class	Precision	Recall	F1	AUC ROC	Accuracy
200	Acq	0.77	0.69	0.73	0.97	66.77%
	Crude	0.10	0.90	0.17	0.94	
	Earn	0.98	0.65	0.78	0.96	
	Grain	0.51	0.75	0.61	0.98	
	Interest	0.02	1.00	0.03	0.95	
	Money-fx	0.26	0.62	0.36	0.95	
	Ship	0.00	0.00	0.00	0.96	
	Trade	0.15	0.81	0.25	0.96	
500	Acq	0.77	0.68	0.72	0.97	65.04%
	Crude	0.03	0.86	0.06	0.95	
	Earn	0.98	0.63	0.76	0.96	
	Grain	0.31	0.96	0.47	0.99	
	Interest	0.03	0.80	0.06	0.96	
	Money-fx	0.28	0.63	0.39	0.96	
	Ship	0.00	0.00	0.00	0.96	
	Trade	0.07	0.89	0.13	0.96	
1000	Acq	0.82	0.69	0.75	0.98	68.91%
	Crude	0.08	0.88	0.15	0.96	
	Earn	0.97	0.68	0.80	0.97	
	Grain	0.38	0.95	0.54	0.99	
	Interest	0.11	0.78	0.19	0.96	
	Money-fx	0.40	0.61	0.48	0.96	
	Ship	0.03	1.00	0.07	0.97	
	Trade	0.24	0.88	0.38	0.96	

Table 5: Results for 8 classes of Reuters8 data set for doc2vec document representation models with feature vectors of sizes 200, 500 and 1000

Features	Class	Precision	Recall	F1	AUC ROC	Accuracy
Average	Acq	0.71	0.50	0.58	0.81	58.31%
	Crude	0.05	0.28	0.09	0.76	
	Earn	0.88	0.71	0.79	0.93	
	Grain	0.13	0.29	0.18	0.76	
	Interest	0.09	0.43	0.15	0.70	
	Money-fx	0.14	0.33	0.20	0.71	
	Ship	0.02	0.29	0.04	0.77	
	Trade	0.14	0.20	0.16	0.82	
Histogram	Acq	0.76	0.48	0.59	0.82	59.66%
	Crude	0.04	0.44	0.07	0.78	
	Earn	0.90	0.72	0.80	0.94	
	Grain	0.05	0.57	0.10	0.77	
	Interest	0.08	0.48	0.14	0.72	
	Money-fx	0.15	0.39	0.22	0.71	
	Ship	0.01	0.50	0.02	0.80	
	Trade	0.11	0.27	0.16	0.83	
Quantiles	Acq	0.72	0.48	0.57	0.81	58.50%
	Crude	0.06	0.41	0.11	0.78	
	Earn	0.89	0.72	0.80	0.94	
	Grain	0.08	0.24	0.12	0.77	
	Interest	0.08	0.46	0.14	0.71	
	Money-fx	0.13	0.35	0.19	0.71	
	Ship	0.01	0.50	0.02	0.79	
	Trade	0.10	0.18	0.13	0.83	

Table 6: Results for 8 classes of Reuters8 data set for graph-of-word representation models using averaging, histograms and quantiles for aggregations of local measures

2 Results WebKB data set

Accuracy		Per class Precision		Per class Recall		Per class F1		Per class AUC ROC		Per class all	
word2vec_75	1	tf	4.5	tf	3	tf	4	tf	5	tf	2
word2vec_50	2	tf-idf	4.5	tf-idf	3	tf-idf	4	tf-idf	5	tf-idf	2
word2vec_100	3	doc2vec_25	4.5	doc2vec_100	3	doc2vec_25	4	doc2vec_25	5	doc2vec_50	2
word2vec_25	4	doc2vec_50	4.5	doc2vec_50	3	word2vec_100	4	doc2vec_50	5	doc2vec_100	10
tf-idf	5	word2vec_100	4.5	doc2vec_75	3	word2vec_25	4	doc2vec_75	5	doc2vec_1000	10
tf	6	word2vec_25	4.5	doc2vec_1000	11	word2vec_50	4	word2vec_100	5	doc2vec_200	10
doc2vec_25	7	word2vec_50	4.5	doc2vec_200	11	word2vec_75	4	word2vec_25	5	doc2vec_25	10
doc2vec_50	8	word2vec_75	4.5	doc2vec_25	11	doc2vec_50	8	word2vec_50	5	doc2vec_500	10
doc2vec_75	9	doc2vec_100	12.5	doc2vec_500	11	doc2vec_100	9.5	word2vec_75	5	doc2vec_75	10
doc2vec_100	10	doc2vec_1000	12.5	gow_avg	11	doc2vec_75	9.5	doc2vec_100	10.5	gow_avg	10
doc2vec_200	11	doc2vec_200	12.5	gow_histogram	11	doc2vec_200	12	doc2vec_1000	10.5	gow_histogram	10
gow_histogram	12	doc2vec_500	12.5	gow_quantiles	11	gow_avg	12	doc2vec_200	12.5	gow_quantiles	10
gow_quantiles	13	doc2vec_75	12.5	word2vec_100	11	gow_histogram	12	doc2vec_500	12.5	word2vec_100	10
gow_avg	14	gow_avg	12.5	word2vec_25	11	doc2vec_1000	15	gow_avg	15	word2vec_25	10
doc2vec_1000	15	gow_histogram	12.5	word2vec_50	11	doc2vec_500	15	gow_histogram	15	word2vec_50	10
doc2vec_500	16	gow_quantiles	12.5	word2vec_75	11	gow_quantiles	15	gow_quantiles	15	word2vec_75	10

Table 7: Ranks of document representation models for WebKB data set in terms of accuracy, per class precision, per class recall, per class F1 and per class AUC ROC

Features	Class	Precision	Recall	F1	AUC ROC	Accuracy
tf	Course	0.61	0.75	0.67	0.95	70.44%
	Department	0.58	0.95	0.72	0.96	
	Faculty	0.64	0.67	0.66	0.94	
	Other	0.86	0.71	0.78	0.88	
	Project	0.03	0.75	0.06	0.88	
	Staff	0.00	0.00	0.00	0.83	
	Student	0.72	0.67	0.69	0.92	
tf-idf	Course	0.68	0.77	0.72	0.96	70.93%
	Department	0.72	0.81	0.76	0.97	
	Faculty	0.62	0.70	0.66	0.93	
	Other	0.83	0.73	0.78	0.87	
	Project	0.18	0.56	0.27	0.86	
	Staff	0.00	0.00	0.00	0.83	
	Student	0.73	0.65	0.69	0.91	

Table 8: Results for 7 classes of WebKB data set for bag-of-words document representation models

Features	Class	Precision	Recall	F1	AUC ROC	Accuracy	Features	Class	Precision	Recall	F1	AUC ROC	Accuracy
25	Course	0.73	0.82	0.77	0.97	71.65%	25	Course	0.62	0.80	0.70	0.97	69.05%
	Department	0.47	0.61	0.53	0.94			Department	0.44	0.76	0.56	0.96	
	Faculty	0.65	0.65	0.65	0.93			Faculty	0.61	0.63	0.62	0.91	
	Other	0.86	0.75	0.80	0.89			Other	0.88	0.68	0.77	0.85	
	Project	0.21	0.62	0.31	0.90			Project	0.18	0.75	0.29	0.89	
	Staff	0.00	0.00	0.00	0.80			Staff	0.00	0.00	0.00	0.80	
	Student	0.67	0.64	0.65	0.91			Student	0.60	0.69	0.64	0.90	
50	Course	0.70	0.80	0.75	0.97	72.32%	50	Course	0.52	0.84	0.64	0.97	65.11%
	Department	0.47	0.63	0.54	0.94			Department	0.22	0.89	0.36	0.97	
	Faculty	0.67	0.69	0.68	0.93			Faculty	0.50	0.62	0.56	0.90	
	Other	0.87	0.75	0.81	0.89			Other	0.90	0.63	0.74	0.84	
	Project	0.20	0.50	0.29	0.91			Project	0.01	1.00	0.02	0.88	
	Staff	0.00	0.00	0.00	0.86			Staff	0.00	0.00	0.00	0.70	
	Student	0.69	0.67	0.68	0.92			Student	0.55	0.67	0.60	0.90	
75	Course	0.69	0.82	0.75	0.97	72.74%	75	Course	0.43	0.83	0.57	0.96	62.57%
	Department	0.50	0.62	0.55	0.94			Department	0.11	0.80	0.20	0.96	
	Faculty	0.68	0.72	0.70	0.94			Faculty	0.48	0.68	0.56	0.91	
	Other	0.87	0.75	0.80	0.90			Other	0.90	0.59	0.71	0.83	
	Project	0.22	0.58	0.32	0.92			Project	0.01	1.00	0.02	0.86	
	Staff	0.00	0.00	0.00	0.84			Staff	0.00	0.00	0.00	0.70	
	Student	0.70	0.66	0.68	0.92			Student	0.51	0.69	0.59	0.90	
100	Course	0.70	0.82	0.75	0.97	71.96%	100	Course	0.36	0.85	0.50	0.95	60.69%
	Department	0.47	0.55	0.51	0.94			Department	0.14	0.83	0.24	0.95	
	Faculty	0.65	0.69	0.67	0.94			Faculty	0.41	0.68	0.51	0.90	
	Other	0.86	0.75	0.80	0.90			Other	0.94	0.56	0.71	0.82	
	Project	0.23	0.52	0.32	0.91			Project	0.00	0.00	0.00	0.84	
	Staff	0.00	0.00	0.00	0.84			Staff	0.00	0.00	0.00	0.63	
	Student	0.70	0.66	0.68	0.92			Student	0.41	0.73	0.52	0.88	

Table 9: Results for 7 classes of WebKB data set for word2vec document representation models with feature vectors of sizes 25, 50, 75 and 100

Table 10: Results for 7 classes of WebKB data set for doc2vec document representation models with feature vectors of sizes 25, 50, 75 and 100

Features	Class	Precision	Recall	F1	AUC ROC	Accuracy
200	Course	0.10	0.90	0.18	0.94	52.76%
	Department	0.08	0.75	0.15	0.95	
	Faculty	0.17	0.83	0.29	0.88	
	Other	0.97	0.50	0.66	0.79	
	Project	0.00	0.00	0.00	0.83	
	Staff	0.00	0.00	0.00	0.61	
	Student	0.24	0.71	0.36	0.86	
500	Course	0.02	1.00	0.04	0.92	49.55%
	Department	0.03	1.00	0.05	0.90	
	Faculty	0.09	0.91	0.16	0.89	
	Other	0.99	0.48	0.65	0.78	
	Project	0.00	0.00	0.00	0.79	
	Staff	0.00	0.00	0.00	0.64	
	Student	0.15	0.67	0.25	0.85	
1000	Course	0.04	1.00	0.07	0.94	50.15%
	Department	0.03	1.00	0.05	0.93	
	Faculty	0.08	0.90	0.16	0.89	
	Other	1.00	0.48	0.65	0.80	
	Project	0.00	0.00	0.00	0.80	
	Staff	0.00	0.00	0.00	0.66	
	Student	0.16	0.78	0.26	0.87	

Table 11: Results for 7 classes of WebKB data set for doc2vec document representation models with feature vectors of sizes 200, 500 and 1000

Features	Class	Precision	Recall	F1	AUC ROC	Accuracy
Average	Course	0.06	0.41	0.10	0.68	50.39%
	Department	0.14	0.83	0.24	0.71	
	Faculty	0.15	0.29	0.20	0.72	
	Other	0.81	0.56	0.66	0.73	
	Project	0.00	0.00	0.00	0.68	
	Staff	0.00	0.00	0.00	0.56	
	Student	0.54	0.43	0.48	0.80	
Histogram	Course	0.06	0.61	0.11	0.70	51.48%
	Department	0.14	0.83	0.24	0.76	
	Faculty	0.15	0.46	0.22	0.74	
	Other	0.84	0.54	0.66	0.72	
	Project	0.00	0.00	0.00	0.70	
	Staff	0.00	0.00	0.00	0.57	
	Student	0.52	0.44	0.47	0.80	
Quantiles	Course	0.02	0.43	0.03	0.69	50.58%
	Department	0.11	0.80	0.20	0.73	
	Faculty	0.10	0.47	0.16	0.71	
	Other	0.81	0.54	0.65	0.73	
	Project	0.00	0.00	0.00	0.71	
	Staff	0.00	0.00	0.00	0.53	
	Student	0.59	0.42	0.49	0.80	

Table 12: Results for 7 classes of WebKB data set for graph-of-word representation models using averaging, histograms and quantiles for aggregations of local measures

3 Results for 20Newsgroups data set

Accuracy	Per class Precision	Per class Recall	Per class F1	Per class AUC ROC	Per class all
tf-idf	1	1.5	1.5	1	1
doc2vec_50	2	1.5	1.5	3.5	7.5
doc2vec_25	3	8	8	3.5	7.5
doc2vec_75	4	8	8	3.5	7.5
doc2vec_100	5	8	8	3.5	7.5
tf	6	8	8	9.5	7.5
doc2vec_200	7	8	8	9.5	7.5
word2vec_100	8	8	8	9.5	7.5
doc2vec_75	9	8	8	9.5	7.5
word2vec_50	10	8	8	9.5	7.5
doc2vec_1000	11	8	8	9.5	7.5
word2vec_500	12	8	8	9.5	7.5
word2vec_25	13	8	8	9.5	7.5
gow_quantiles	14	15	15	15	15
gow_histogram	15	15	15	15	15
gow_avg	16	15	15	15	15

Table 13: Ranks of document representation models for 20News data set in terms of accuracy, per class precision, per class recall, per class F1 and per class AUC ROC

Features	Class	Precision	Recall	F1	AUC ROC	Accuracy
tf	alt.atheism	0.43	0.54	0.48	0.91	61.7%
	comp.graphics	0.49	0.49	0.49	0.91	
	comp.os.ms-windows.misc	0.61	0.60	0.61	0.92	
	comp.sys.ibm.pc.hardware	0.51	0.54	0.53	0.91	
	comp.sys.mac.hardware	0.52	0.64	0.57	0.93	
	comp.windows.x	0.50	0.62	0.55	0.93	
	misc.forsale	0.77	0.59	0.67	0.97	
	rec.autos	0.62	0.73	0.67	0.94	
	rec.motorcycles	0.77	0.62	0.69	0.97	
	rec.sport.baseball	0.70	0.60	0.65	0.95	
	rec.sport.hockey	0.86	0.76	0.81	0.98	
	sci.crypt	0.63	0.79	0.70	0.94	
	sci.electronics	0.41	0.35	0.38	0.85	
	sci.med	0.52	0.53	0.52	0.90	
	sci.space	0.72	0.73	0.72	0.95	
	soc.religion.christian	0.82	0.57	0.67	0.95	
	talk.politics.guns	0.71	0.54	0.61	0.94	
	talk.politics.mideast	0.77	0.86	0.81	0.98	
	talk.politics.misc	0.44	0.51	0.47	0.92	
	talk.religion.misc	0.11	0.71	0.19	0.90	

Table 14: Results for 20 classes of 20News data set for *tf* variant of bag-of-words document representation models

Features	Class	Precision	Recall	F1	AUC ROC	Accuracy
tf-idf	alt.atheism	0.61	0.63	0.62	0.97	75.6%
	comp.graphics	0.70	0.59	0.64	0.96	
	comp.os.ms-windows.misc	0.68	0.65	0.67	0.96	
	comp.sys.ibm.pc.hardware	0.59	0.64	0.62	0.96	
	comp.sys.mac.hardware	0.74	0.70	0.72	0.98	
	comp.windows.x	0.64	0.74	0.69	0.97	
	misc.forsale	0.80	0.79	0.80	0.99	
	rec.autos	0.83	0.84	0.84	0.99	
	rec.motorcycles	0.87	0.91	0.89	0.99	
	rec.sport.baseball	0.85	0.89	0.87	0.99	
	rec.sport.hockey	0.92	0.94	0.93	1.00	
	sci.crypt	0.81	0.90	0.85	0.99	
	sci.electronics	0.65	0.56	0.60	0.95	
	sci.med	0.76	0.75	0.76	0.97	
	sci.space	0.87	0.90	0.88	0.99	
	soc.religion.christian	0.87	0.67	0.76	0.98	
	talk.politics.guns	0.89	0.66	0.76	0.98	
	talk.politics.mideast	0.82	0.94	0.87	0.99	
	talk.politics.misc	0.52	0.53	0.52	0.94	
	talk.religion.misc	0.22	0.88	0.36	0.94	

Table 15: Results for 20 classes of 20News data set for *tf-idf* variant of bag-of-words document representation models

Features	Class	Precision	Recall	F1	AUC ROC	Accuracy
25	alt.atheism	0.51	0.45	0.48	0.93	52.94%
	comp.graphics	0.39	0.38	0.38	0.92	
	comp.os.ms-windows.misc	0.41	0.46	0.43	0.92	
	comp.sys.ibm.pc.hardware	0.36	0.43	0.39	0.93	
	comp.sys.mac.hardware	0.38	0.40	0.39	0.92	
	comp.windows.x	0.60	0.51	0.55	0.95	
	misc.forsale	0.76	0.63	0.69	0.97	
	rec.autos	0.42	0.46	0.44	0.93	
	rec.motorcycles	0.66	0.53	0.59	0.96	
	rec.sport.baseball	0.57	0.57	0.57	0.96	
	rec.sport.hockey	0.65	0.70	0.68	0.98	
	sci.crypt	0.71	0.81	0.76	0.97	
	sci.electronics	0.34	0.36	0.35	0.88	
	sci.med	0.56	0.52	0.54	0.93	
	sci.space	0.62	0.61	0.61	0.95	
	soc.religion.christian	0.73	0.63	0.68	0.97	
	talk.politics.guns	0.62	0.51	0.56	0.95	
	talk.politics.mideast	0.70	0.77	0.73	0.98	
	talk.politics.misc	0.36	0.47	0.41	0.91	
	talk.religion.misc	0.19	0.41	0.26	0.91	
50	alt.atheism	0.54	0.46	0.50	0.94	56.77%
	comp.graphics	0.49	0.43	0.46	0.93	
	comp.os.ms-windows.misc	0.51	0.53	0.52	0.93	
	comp.sys.ibm.pc.hardware	0.40	0.49	0.44	0.94	
	comp.sys.mac.hardware	0.39	0.45	0.42	0.92	
	comp.windows.x	0.63	0.57	0.59	0.95	
	misc.forsale	0.79	0.63	0.70	0.97	
	rec.autos	0.50	0.50	0.50	0.94	
	rec.motorcycles	0.66	0.53	0.59	0.96	
	rec.sport.baseball	0.64	0.64	0.64	0.97	
	rec.sport.hockey	0.66	0.75	0.70	0.98	
	sci.crypt	0.74	0.83	0.78	0.98	
	sci.electronics	0.36	0.40	0.38	0.90	
	sci.med	0.60	0.60	0.60	0.94	
	sci.space	0.70	0.67	0.69	0.96	
	soc.religion.christian	0.74	0.64	0.69	0.98	
	talk.politics.guns	0.70	0.56	0.62	0.96	
	talk.politics.mideast	0.73	0.75	0.74	0.98	
	talk.politics.misc	0.34	0.45	0.39	0.92	
	talk.religion.misc	0.16	0.47	0.24	0.92	

Table 16: Results for 20 classes of 20News data set for word2vec document representation models with feature vectors of sizes 25 and 50

Features	Class	Precision	Recall	F1	AUC ROC	Accuracy
75	alt.atheism	0.53	0.46	0.49	0.94	57.4%
	comp.graphics	0.49	0.44	0.47	0.93	
	comp.os.ms-windows.misc	0.53	0.53	0.53	0.93	
	comp.sys.ibm.pc.hardware	0.39	0.51	0.44	0.94	
	comp.sys.mac.hardware	0.39	0.44	0.42	0.92	
	comp.windows.x	0.63	0.59	0.60	0.96	
	misc.forsale	0.79	0.66	0.72	0.97	
	rec.autos	0.49	0.50	0.50	0.94	
	rec.motorcycles	0.67	0.57	0.62	0.96	
	rec.sport.baseball	0.68	0.58	0.63	0.97	
	rec.sport.hockey	0.69	0.77	0.73	0.98	
	sci.crypt	0.73	0.79	0.76	0.98	
	sci.electronics	0.37	0.40	0.39	0.90	
	sci.med	0.58	0.61	0.60	0.94	
	sci.space	0.70	0.67	0.69	0.96	
	soc.religion.christian	0.79	0.66	0.72	0.98	
	talk.politics.guns	0.70	0.55	0.62	0.96	
	talk.politics.mideast	0.75	0.75	0.75	0.98	
	talk.politics.misc	0.35	0.51	0.42	0.92	
	talk.religion.misc	0.14	0.48	0.22	0.92	
100	alt.atheism	0.55	0.47	0.51	0.94	57.9%
	comp.graphics	0.48	0.46	0.47	0.93	
	comp.os.ms-windows.misc	0.54	0.53	0.54	0.93	
	comp.sys.ibm.pc.hardware	0.41	0.50	0.45	0.94	
	comp.sys.mac.hardware	0.40	0.46	0.43	0.92	
	comp.windows.x	0.63	0.60	0.62	0.96	
	misc.forsale	0.80	0.67	0.73	0.97	
	rec.autos	0.52	0.52	0.52	0.95	
	rec.motorcycles	0.66	0.57	0.61	0.96	
	rec.sport.baseball	0.65	0.57	0.61	0.97	
	rec.sport.hockey	0.70	0.76	0.73	0.98	
	sci.crypt	0.74	0.80	0.77	0.98	
	sci.electronics	0.38	0.41	0.40	0.90	
	sci.med	0.59	0.64	0.61	0.94	
	sci.space	0.72	0.66	0.69	0.96	
	soc.religion.christian	0.77	0.67	0.72	0.98	
	talk.politics.guns	0.70	0.56	0.62	0.96	
	talk.politics.mideast	0.75	0.75	0.75	0.98	
	talk.politics.misc	0.36	0.47	0.41	0.92	
	talk.religion.misc	0.17	0.51	0.25	0.92	

Table 17: Results for 20 classes of 20News data set for word2vec document representation models with feature vectors of sizes 75 and 100

Features	Class	Precision	Recall	F1	AUC ROC	Accuracy
25	alt.atheism	0.50	0.56	0.53	0.95	66%
	comp.graphics	0.65	0.54	0.59	0.95	
	comp.os.ms-windows.misc	0.48	0.49	0.48	0.93	
	comp.sys.ibm.pc.hardware	0.48	0.51	0.50	0.95	
	comp.sys.mac.hardware	0.48	0.52	0.50	0.95	
	comp.windows.x	0.65	0.67	0.66	0.97	
	misc.forsale	0.70	0.65	0.67	0.97	
	rec.autos	0.72	0.69	0.71	0.97	
	rec.motorcycles	0.72	0.75	0.74	0.98	
	rec.sport.baseball	0.79	0.78	0.78	0.99	
	rec.sport.hockey	0.85	0.85	0.85	0.99	
	sci.crypt	0.82	0.84	0.83	0.98	
	sci.electronics	0.54	0.59	0.57	0.95	
	sci.med	0.77	0.80	0.78	0.97	
	sci.space	0.83	0.76	0.79	0.98	
	soc.religion.christian	0.84	0.58	0.69	0.98	
	talk.politics.guns	0.83	0.60	0.70	0.97	
	talk.politics.mideast	0.84	0.86	0.85	0.99	
	talk.politics.misc	0.43	0.65	0.51	0.91	
	talk.religion.misc	0.12	0.52	0.20	0.93	
50	alt.atheism	0.49	0.57	0.52	0.94	66.83%
	comp.graphics	0.57	0.59	0.58	0.95	
	comp.os.ms-windows.misc	0.61	0.53	0.57	0.94	
	comp.sys.ibm.pc.hardware	0.51	0.53	0.52	0.95	
	comp.sys.mac.hardware	0.47	0.53	0.50	0.94	
	comp.windows.x	0.71	0.68	0.69	0.97	
	misc.forsale	0.74	0.69	0.72	0.98	
	rec.autos	0.80	0.71	0.75	0.98	
	rec.motorcycles	0.83	0.78	0.80	0.98	
	rec.sport.baseball	0.86	0.86	0.86	0.99	
	rec.sport.hockey	0.80	0.83	0.82	0.98	
	sci.crypt	0.80	0.83	0.82	0.98	
	sci.electronics	0.55	0.60	0.57	0.93	
	sci.med	0.77	0.75	0.76	0.96	
	sci.space	0.81	0.79	0.80	0.98	
	soc.religion.christian	0.85	0.60	0.71	0.97	
	talk.politics.guns	0.84	0.58	0.69	0.97	
	talk.politics.mideast	0.82	0.87	0.85	0.98	
	talk.politics.misc	0.39	0.62	0.48	0.92	
	talk.religion.misc	0.06	0.52	0.10	0.91	

Table 18: Results for 20 classes of 20News data set for doc2vec document representation models with feature vectors of sizes 25 and 50

Features	Class	Precision	Recall	F1	AUC ROC	Accuracy
75	alt.atheism	0.48	0.54	0.50	0.92	65.03%
	comp.graphics	0.54	0.54	0.54	0.93	
	comp.os.ms-windows.misc	0.57	0.56	0.56	0.94	
	comp.sys.ibm.pc.hardware	0.50	0.53	0.51	0.94	
	comp.sys.mac.hardware	0.46	0.58	0.51	0.94	
	comp.windows.x	0.68	0.64	0.66	0.96	
	misc.forsale	0.77	0.62	0.68	0.97	
	rec.autos	0.68	0.65	0.67	0.96	
	rec.motorcycles	0.76	0.70	0.73	0.98	
	rec.sport.baseball	0.75	0.66	0.70	0.97	
	rec.sport.hockey	0.81	0.75	0.78	0.99	
	sci.crypt	0.79	0.78	0.79	0.97	
	sci.electronics	0.49	0.63	0.55	0.92	
	sci.med	0.73	0.73	0.73	0.95	
	sci.space	0.78	0.72	0.75	0.97	
	soc.religion.christian	0.79	0.59	0.67	0.97	
	talk.politics.guns	0.80	0.56	0.66	0.96	
	talk.politics.mideast	0.78	0.87	0.82	0.98	
	talk.politics.misc	0.37	0.59	0.46	0.91	
	talk.religion.misc	0.06	0.80	0.12	0.90	
100	alt.atheism	0.50	0.56	0.52	0.93	62%
	comp.graphics	0.44	0.57	0.50	0.92	
	comp.os.ms-windows.misc	0.54	0.49	0.52	0.93	
	comp.sys.ibm.pc.hardware	0.48	0.55	0.51	0.93	
	comp.sys.mac.hardware	0.46	0.55	0.50	0.92	
	comp.windows.x	0.68	0.62	0.65	0.96	
	misc.forsale	0.72	0.61	0.66	0.96	
	rec.autos	0.68	0.69	0.69	0.96	
	rec.motorcycles	0.73	0.61	0.66	0.96	
	rec.sport.baseball	0.74	0.69	0.71	0.97	
	rec.sport.hockey	0.84	0.76	0.80	0.99	
	sci.crypt	0.81	0.75	0.78	0.97	
	sci.electronics	0.43	0.52	0.47	0.90	
	sci.med	0.69	0.63	0.66	0.94	
	sci.space	0.74	0.72	0.73	0.97	
	soc.religion.christian	0.82	0.57	0.67	0.96	
	talk.politics.guns	0.77	0.53	0.63	0.95	
	talk.politics.mideast	0.79	0.84	0.81	0.98	
	talk.politics.misc	0.34	0.63	0.44	0.91	
	talk.religion.misc	0.03	0.50	0.05	0.90	

Table 19: Results for 20 classes of 20News data set for doc2vec document representation models with feature vectors of sizes 75 and 100

Features	Class	Precision	Recall	F1	AUC	ROC	Accuracy
200	alt.atheism	0.37	0.52	0.43	0.90		
	comp.graphics	0.40	0.49	0.44	0.90		
	comp.os.ms-windows.misc	0.45	0.51	0.48	0.90		
	comp.sys.ibm.pc.hardware	0.51	0.44	0.47	0.92		
	comp.sys.mac.hardware	0.42	0.53	0.47	0.90		
	comp.windows.x	0.67	0.53	0.59	0.94		
	misc.forsale	0.71	0.60	0.65	0.95		
	rec.autos	0.68	0.56	0.61	0.94		59%
	rec.motorcycles	0.63	0.69	0.66	0.96		
	rec.sport.baseball	0.69	0.70	0.70	0.95		
	rec.sport.hockey	0.85	0.70	0.77	0.98		
	sci.crypt	0.74	0.63	0.68	0.95		
	sci.electronics	0.34	0.55	0.42	0.84		
	sci.med	0.63	0.62	0.63	0.90		
	sci.space	0.67	0.66	0.67	0.94		
	soc.religion.christian	0.82	0.50	0.62	0.95		
	talk.politics.guns	0.69	0.47	0.56	0.94		
	talk.politics.mideast	0.73	0.74	0.74	0.96		
	talk.politics.misc	0.23	0.61	0.33	0.87		
	talk.religion.misc	0.02	0.75	0.05	0.87		
500	alt.atheism	0.25	0.63	0.36	0.89		
	comp.graphics	0.48	0.47	0.48	0.88		
	comp.os.ms-windows.misc	0.51	0.45	0.48	0.90		
	comp.sys.ibm.pc.hardware	0.42	0.42	0.42	0.90		
	comp.sys.mac.hardware	0.45	0.49	0.47	0.90		
	comp.windows.x	0.65	0.58	0.61	0.94		
	misc.forsale	0.77	0.53	0.63	0.95		
	rec.autos	0.58	0.63	0.61	0.93		54.3%
	rec.motorcycles	0.67	0.58	0.62	0.95		
	rec.sport.baseball	0.63	0.66	0.65	0.94		
	rec.sport.hockey	0.85	0.68	0.76	0.98		
	sci.crypt	0.67	0.63	0.65	0.93		
	sci.electronics	0.30	0.46	0.36	0.83		
	sci.med	0.63	0.59	0.61	0.91		
	sci.space	0.64	0.69	0.66	0.93		
	soc.religion.christian	0.80	0.45	0.57	0.95		
	talk.politics.guns	0.61	0.52	0.56	0.92		
	talk.politics.mideast	0.74	0.68	0.71	0.95		
	talk.politics.misc	0.11	0.56	0.18	0.84		
	talk.religion.misc	0.00	0.17	0.01	0.83		
1000	alt.atheism	0.22	0.53	0.31	0.89		
	comp.graphics	0.46	0.47	0.46	0.88		
	comp.os.ms-windows.misc	0.48	0.46	0.47	0.89		
	comp.sys.ibm.pc.hardware	0.42	0.50	0.45	0.92		
	comp.sys.mac.hardware	0.47	0.53	0.50	0.91		
	comp.windows.x	0.75	0.56	0.64	0.95		
	misc.forsale	0.78	0.56	0.65	0.96		
	rec.autos	0.55	0.57	0.56	0.93		56.14%
	rec.motorcycles	0.74	0.58	0.65	0.96		
	rec.sport.baseball	0.76	0.69	0.72	0.96		
	rec.sport.hockey	0.85	0.75	0.80	0.98		
	sci.crypt	0.69	0.66	0.67	0.94		
	sci.electronics	0.29	0.47	0.36	0.82		
	sci.med	0.63	0.61	0.62	0.92		
	sci.space	0.65	0.69	0.67	0.94		
	soc.religion.christian	0.81	0.46	0.59	0.95		
	talk.politics.guns	0.63	0.52	0.57	0.92		
	talk.politics.mideast	0.80	0.66	0.72	0.96		
	talk.politics.misc	0.10	0.62	0.18	0.84		
	talk.religion.misc	0.01	0.33	0.02	0.84		

Table 20: Results for 20 classes of 20News data set for doc2vec document representation models with feature vectors of sizes 200, 500 and 1000

Features	Class	Precision	Recall	F1	AUC	ROC	Accuracy
Average	alt.atheism	0.05	0.07	0.06	0.58		
	comp.graphics	0.11	0.11	0.11	0.62		
	comp.os.ms-windows.misc	0.10	0.08	0.09	0.61		
	comp.sys.ibm.pc.hardware	0.09	0.09	0.09	0.60		
	comp.sys.mac.hardware	0.09	0.10	0.09	0.62		
	comp.windows.x	0.15	0.12	0.13	0.65		
	misc.forsale	0.26	0.19	0.22	0.72		
	rec.autos	0.05	0.06	0.06	0.54		10.5%
	rec.motorcycles	0.16	0.11	0.13	0.66		
	rec.sport.baseball	0.08	0.11	0.09	0.55		
	rec.sport.hockey	0.10	0.13	0.11	0.60		
	sci.crypt	0.10	0.08	0.09	0.60		
	sci.electronics	0.07	0.08	0.08	0.56		
	sci.med	0.05	0.05	0.05	0.52		
	sci.space	0.07	0.08	0.08	0.55		
	soc.religion.christian	0.16	0.12	0.14	0.64		
	talk.politics.guns	0.18	0.17	0.17	0.71		
	talk.politics.mideast	0.17	0.16	0.17	0.67		
	talk.politics.misc	0.09	0.13	0.11	0.62		
	talk.religion.misc	0.02	0.06	0.03	0.55		
Histogram	alt.atheism	0.04	0.09	0.05	0.60		
	comp.graphics	0.13	0.12	0.12	0.66		
	comp.os.ms-windows.misc	0.10	0.10	0.10	0.61		
	comp.sys.ibm.pc.hardware	0.11	0.10	0.11	0.63		
	comp.sys.mac.hardware	0.08	0.09	0.09	0.63		
	comp.windows.x	0.12	0.10	0.11	0.66		
	misc.forsale	0.29	0.19	0.23	0.75		
	rec.autos	0.06	0.08	0.07	0.55		11.12%
	rec.motorcycles	0.22	0.12	0.15	0.67		
	rec.sport.baseball	0.07	0.10	0.08	0.53		
	rec.sport.hockey	0.11	0.14	0.12	0.58		
	sci.crypt	0.12	0.10	0.11	0.61		
	sci.electronics	0.05	0.05	0.05	0.55		
	sci.med	0.07	0.08	0.07	0.51		
	sci.space	0.04	0.06	0.05	0.56		
	soc.religion.christian	0.19	0.13	0.15	0.63		
	talk.politics.guns	0.19	0.19	0.19	0.73		
	talk.politics.mideast	0.25	0.21	0.23	0.71		
	talk.politics.misc	0.04	0.11	0.06	0.65		
	talk.religion.misc	0.01	0.08	0.02	0.56		
Quantiles	alt.atheism	0.01	0.14	0.01	0.60		
	comp.graphics	0.14	0.11	0.12	0.66		
	comp.os.ms-windows.misc	0.08	0.08	0.08	0.63		
	comp.sys.ibm.pc.hardware	0.12	0.08	0.09	0.62		
	comp.sys.mac.hardware	0.06	0.11	0.07	0.64		
	comp.windows.x	0.10	0.12	0.11	0.66		
	misc.forsale	0.32	0.17	0.23	0.73		
	rec.autos	0.02	0.08	0.03	0.55		11.61%
	rec.motorcycles	0.46	0.10	0.17	0.68		
	rec.sport.baseball	0.06	0.09	0.07	0.52		
	rec.sport.hockey	0.06	0.26	0.09	0.59		
	sci.crypt	0.12	0.09	0.10	0.62		
	sci.electronics	0.05	0.09	0.06	0.58		
	sci.med	0.04	0.06	0.05	0.51		
	sci.space	0.01	0.02	0.01	0.58		
	soc.religion.christian	0.20	0.11	0.14	0.66		
	talk.politics.guns	0.14	0.22	0.17	0.71		
	talk.politics.mideast	0.26	0.21	0.23	0.69		
	talk.politics.misc	0.04	0.16	0.06	0.63		
	talk.religion.misc	0.00	0	0	0.57		

Table 21: Results for 20 classes of 20News data set for graph-of-word representation models using averaging, histograms and quantiles for aggregations of local measures

4 The results for Brown data set

4.1 The results for Brown2: 2 classes

Accuracy	Per class Precision	Per class Recall	Per class F1	Per class AUC ROC	Per class all
tf	tf	tf	tf	tf	tf
word2vec_50	word2vec_50	word2vec_50	word2vec_50	word2vec_50	word2vec_50
doc2vec_25	doc2vec_25	doc2vec_25	doc2vec_25	doc2vec_25	doc2vec_25
tf-idf	doc2vec_75	doc2vec_75	doc2vec_75	doc2vec_75	doc2vec_75
word2vec_100	tf-idf	tf-idf	word2vec_100	tf-idf	word2vec_50
word2vec_75	doc2vec_50	doc2vec_50	word2vec_75	word2vec_75	word2vec_75
doc2vec_75	word2vec_100	word2vec_100	word2vec_75	word2vec_50	tf-idf
gow_avg	word2vec_75	word2vec_75	gow_avg	tf	doc2vec_200
gow_histogram	doc2vec_1000	doc2vec_1000	gow_histogram	doc2vec_100	word2vec_100
word2vec_25	doc2vec_200	doc2vec_200	word2vec_25	doc2vec_1000	doc2vec_1000
doc2vec_50	gow_histogram	gow_histogram	doc2vec_50	word2vec_100	word2vec_25
gow_quantiles	word2vec_25	word2vec_25	gow_quantiles	word2vec_25	doc2vec_100
doc2vec_100	doc2vec_100	doc2vec_100	doc2vec_100	gow_avg	doc2vec_500
doc2vec_1000	doc2vec_500	doc2vec_500	doc2vec_1000	doc2vec_500	gow_avg
doc2vec_200	gow_avg	gow_avg	doc2vec_200	gow_quantiles	gow_histogram
doc2vec_500	gow_quantiles	gow_quantiles	doc2vec_500	gow_histogram	gow_quantiles

Table 22: Ranks of document representation models for Brown2 dataset in terms of accuracy, per class precision, per class recall, per class F1 and per class AUC ROC

Features	Class	Precision	Recall	F1	AUC ROC	Accuracy
tf	Fiction	0.92	0.96	0.94	0.99	96.97%
	Informative	0.99	0.97	0.98	0.99	
tf-idf	Fiction	0.80	0.95	0.87	0.99	93.94%
	Informative	0.99	0.94	0.96	0.99	

Table 23: Results for 2 classes of Brown2 data set for bag-of-words document representation models

Features	Classes	Precision	Recall	F1	AUC ROC	Accuracy
25	Fiction	0.80	0.91	0.85	0.99	92.93%
	Informative	0.97	0.94	0.95	0.99	
50	Fiction	0.84	1.00	0.91	0.99	95.96%
	Informative	1.00	0.95	0.97	0.99	
75	Fiction	0.84	0.91	0.88	0.99	93.94%
	Informative	0.97	0.95	0.96	0.99	
100	Fiction	0.84	0.91	0.88	0.99	93.94%
	Informative	0.97	0.95	0.96	0.99	

Table 24: Results for 2 classes of Brown2 data set for word2vec document representation models with feature vectors of size 25, 50, 75 and 100

Features	Classes	Precision	Recall	F1	AUC ROC	Accuracy
200	Fiction	0.52	1	0.68	0.99	87.88%
	Informative	1	0.86	0.93	0.99	
500	Fiction	0.44	1	0.61	0.98	85.86%
	Informative	1	0.84	0.91	0.98	
1000	Fiction	0.52	1	0.68	0.99	87.88%
	Informative	1	0.86	0.925	0.99	

Table 26: Results for 2 classes of Brown2 data set for doc2vec document representation models with feature vectors of size 200, 500 and 1000

Features	Classes	Precision	Recall	F1	AUC ROC	Accuracy
25	Fiction	0.84	0.95	0.89	0.99	94.95%
	Informative	0.99	0.95	0.97	0.99	
50	Fiction	0.68	1.00	0.81	1	91.92%
	Informative	1.00	0.90	0.95	1	
75	Fiction	0.72	1	0.84	1	92.93%
	Informative	1	0.91	0.95	1	
100	Fiction	0.68	0.94	0.79	0.99	90.91%
	Informative	0.99	0.9	0.94	0.99	

Table 25: Results for 2 classes of Brown2 data set for doc2vec document representation models with feature vectors of size 25, 50, 75 and 100

Features	Classes	Precision	Recall	F1	AUC ROC	Accuracy
Average	Fiction	0.84	0.88	0.86	0.98	92.93%
	Informative	0.96	0.95	0.95	0.98	
Quantiles	Fiction	0.76	0.9	0.83	0.96	91.92%
	Informative	0.97	0.92	0.95	0.96	
Histogram	Fiction	0.8	0.91	0.85	0.95	92.93%
	Informative	0.97	0.94	0.95	0.95	

Table 27: Results for 2 classes of Brown2 data set for graph-of-word representation models using averaging, histograms and quantiles for aggregations of local measures

4.2 The results for Brown4: 4 classes

Accuracy	Per class Precision	Per class Recall	Per class F1	Per class AUC ROC	Per class all
doc2vec_75	1	tf	1	doc2vec_50	1.5
tf	2	doc2vec_50	2.5	doc2vec_75	1.5
doc2vec_50	3	doc2vec_75	2	tf	5
doc2vec_100	4	doc2vec_25	6.5	doc2vec_100	5
doc2vec_200	5	doc2vec_100	6.5	doc2vec_200	5
tf-idf	6.5	doc2vec_200	6.5	doc2vec_25	5
doc2vec_25	6.5	word2vec_100	6.5	gow_histogram	5
word2vec_50	8	word2vec_50	6.5	word2vec_100	7
doc2vec_500	10	tf-idf	6.5	word2vec_25	7
word2vec_100	10	doc2vec_1000	13	word2vec_50	7
word2vec_25	10	doc2vec_500	13	tf-idf	13.5
gow_avg	12.5	gow_avg	13	doc2vec_1000	13.5
gow_histogram	12.5	gow_histogram	13	doc2vec_500	13.5
gow_quantiles	14	gow_quantiles	13	gow_avg	13.5
word2vec_75	15	word2vec_25	13	word2vec_25	12
doc2vec_1000	16	word2vec_75	12.5	word2vec_50	12
				word2vec_75	12

Table 28: Ranks of document representation models for Brown4 dataset in terms of accuracy, per class precision, per class recall, per class F1 and per class AUC ROC

Features	Class	Precision	Recall	F1	AUC ROC	Accuracy
tf	Fiction	0.96	0.89	0.92	0.99	82.83%
	Non Fiction	0.77	0.85	0.81	0.91	
	Other	0.80	0.78	0.79	0.89	
	Tisak	0.76	0.81	0.79	0.96	
tf-idf	Fiction	0.80	0.91	0.85	0.99	75.76%
	Non Fiction	0.64	0.88	0.74	0.89	
	Other	0.86	0.64	0.73	0.87	
	Tisak	0.65	0.79	0.71	0.93	

Table 29: Results for 4 classes of Brown4 data set for bag-of-words document representation models

Features	Classes	Precision	Recall	F1	AUC ROC	Accuracy
25	Fiction	0.88	0.88	0.88	0.99	71.72%
	Non Fiction	0.64	0.93	0.76	0.91	
	Other	0.69	0.62	0.65	0.76	
	Tisak	0.65	0.55	0.59	0.84	
50	Fiction	0.84	0.88	0.86	0.99	72.73%
	Non Fiction	0.64	0.93	0.76	0.91	
	Other	0.77	0.60	0.68	0.78	
	Tisak	0.59	0.67	0.63	0.86	
75	Fiction	0.84	0.88	0.86	0.99	68.69%
	Non Fiction	0.59	0.87	0.70	0.90	
	Other	0.74	0.57	0.64	0.76	
	Tisak	0.47	0.57	0.52	0.86	
100	Fiction	0.84	0.88	0.86	0.99	71.72%
	Non Fiction	0.64	0.88	0.74	0.92	
	Other	0.80	0.60	0.68	0.79	
	Tisak	0.47	0.67	0.55	0.86	

Table 30: Results for 4 classes of Brown4 data set for word2vec document representation models with feature vectors of size 25, 50, 75 and 100

Features	Classes	Precision	Recall	F1	AUC ROC	Accuracy
200	Fiction	0.84	0.95	0.89	0.99	76.77%
	Non Fiction	0.73	1.00	0.84	0.92	
	Other	1.00	0.61	0.76	0.85	
	Tisak	0.24	1.00	0.38	0.94	
500	Fiction	0.8	0.91	0.85	0.98	71.72%
	Non Fiction	0.68	1.00	0.81	0.93	
	Other	0.97	0.57	0.72	0.87	
	Tisak	0.12	1.00	0.21	0.88	
1000	Fiction	0.88	0.85	0.86	0.99	65.66%
	Non Fiction	0.45	1.00	0.63	0.91	
	Other	0.89	0.51	0.65	0.81	
	Tisak	0.12	1.00	0.21	0.88	

Table 32: Results for 4 classes of Brown4 data set for doc2vec document representation models with feature vectors of size 200, 500 and 1000

Features	Classes	Precision	Recall	F1	AUC ROC	Accuracy
25	Fiction	0.84	0.95	0.89	1	75.76%
	Non Fiction	0.68	0.75	0.71	0.91	
	Other	0.80	0.68	0.74	0.90	
	Tisak	0.65	0.69	0.67	0.95	
50	Fiction	0.92	0.92	0.92	0.99%	81.82%
	Non Fiction	0.77	0.85	0.81	0.95	
	Other	0.83	0.76	0.79	0.92	
	Tisak	0.71	0.75	0.73	0.94	
75	Fiction	0.92	1	0.96	1	83.84%
	Non Fiction	0.73	0.84	0.78	0.93	
	Other	0.91	0.73	0.81	0.93	
	Tisak	0.71	0.92	0.80	0.96	
100	Fiction	0.88	0.92	0.9	0.99	78.79%
	Non Fiction	0.59	0.87	0.70	0.90	
	Other	0.89	0.70	0.78	0.90	
	Tisak	0.71	0.75	0.73	0.93	

Table 31: Results for 4 classes of Brown4 data set for doc2vec document representation models with feature vectors of size 25, 50, 75 and 100

Features	Classes	Precision	Recall	F1	AUC ROC	Accuracy
Average	Fiction	0.76	0.73	0.75	0.95	70.7%
	Non Fiction	0.77	0.77	0.77	0.92	
	Other	0.60	0.62	0.61	0.79	
	Tisak	0.76	0.76	0.76	0.93	
Quantiles	Fiction	0.8	0.74	0.77	0.93	69.7%
	Non Fiction	0.73	0.73	0.73	0.92	
	Other	0.51	0.62	0.56	0.72	
	Tisak	0.88	0.71	0.79	0.93	
Histogram	Fiction	0.84	0.72	0.78	0.93	70.7%
	Non Fiction	0.73	0.76	0.74	0.90	
	Other	0.54	0.61	0.58	0.73	
	Tisak	0.82	0.78	0.80	0.94	

Table 33: Results for 4 classes of Brown4 data set for graph-of-word representation models using averaging, histograms and quantiles for aggregations of local measures

4.3 The results for Brown10: 10 classes

Accuracy		Per class Precision	Per class Recall	Per class F1	Per class AUC ROC	Per class all	
tf	1	tf	3.5	tf-idf	3	tf	2
doc2vec_25	2.5	tf-idf	3.5	doc2vec_100	3	doc2vec_25	2
doc2vec_50	2.5	doc2vec_25	3.5	doc2vec_25	3	doc2vec_50	2
doc2vec_75	4	doc2vec_500	3.5	doc2vec_50	3	doc2vec_75	2
tf-idf	5	doc2vec_75	3.5	doc2vec_75	3	tf-idf	10
doc2vec_100	6	word2vec_75	3.5	tf	7	doc2vec_100	10
doc2vec_200	7	doc2vec_I00	11.5	doc2vec_500	7	doc2vec_200	10
doc2vec_500	8.5	doc2vec_1000	11.5	gow_avg	7	tf-idf	10
word2vec_50	8.5	doc2vec_200	11.5	doc2vec_1000	12.5	doc2vec_1000	10
word2vec_75	10	doc2vec_50	11.5	doc2vec_200	12.5	doc2vec_500	10
word2vec_25	11	gow_avg	11.5	gow_histogram	12.5	gow_avg	10
gow_avg	12.5	gow_histogram	11.5	gow_quantiles	12.5	gow_histogram	10
word2vec_100	12.5	gow_quantiles	11.5	word2vec_100	12.5	gow_quantiles	10
doc2vec_I000	14	word2vec_100	11.5	word2vec_25	12.5	word2vec_100	10
gow_quantiles	15	word2vec_25	11.5	word2vec_50	12.5	word2vec_25	10
gow_histogram	16	word2vec_50	11.5	word2vec_75	12.5	word2vec_50	10
						word2vec_75	10

Table 34: Ranks of document representation models for Brown10 dataset in terms of accuracy, per class precision, per class recall, per class F1 and per class AUC ROC

Features	Classes	Precision	Recall	F1	AUC ROC	Accuracy
tf	Belle Letters	0.73	0.42	0.54	0.83	63.92%
	Editorial	0.60	0.75	0.67	0.98	
	Fiction	0.96	0.80	0.87	0.98	
	Government	0.50	0.75	0.60	0.92	
	Hobbies	0.43	0.60	0.50	0.97	
	Learned	0.63	0.59	0.61	0.81	
	Love	0.11	0.50	0.18	0.74	
	News	0.63	0.71	0.67	0.86	
	Religion	0.00	0.00	0.00	0.98	
	Reviews	0.67	1.00	0.80	1	
tf-idf	Belle Letters	0.67	0.37	0.48	0.84	58.76%
	Editorial	0.40	0.67	0.50	0.94	
	Fiction	0.88	0.81	0.85	0.96	
	Government	0.50	1.00	0.67	0.92	
	Hobbies	0.00	0.00	0.00	0.94	
	Learned	0.63	0.43	0.51	0.78	
	Love	0.11	1.00	0.20	0.71	
	News	0.63	0.83	0.71	0.87	
	Religion	0.33	1.00	0.50	0.99	
	Reviews	1.00	0.75	0.86	1	

Table 35: Results for 10 classes of Brown10 data set for bag-of-words document representation models

Features	Classes	Precision	Recall	F1	AUC ROC	Accuracy
25	Belle Letters	0.40	0.29	0.33	0.74	46.4%
	Editorial	0.60	0.33	0.43	0.94	
	Fiction	0.80	0.71	0.75	0.94	
	Government	0.50	0.60	0.55	0.88	
	Hobbies	0.14	0.50	0.22	0.73	
	Learned	0.44	0.44	0.44	0.83	
	Love	0.00	0.00	0.00	0.59	
	News	0.50	0.67	0.57	0.91	
	Religion	0.33	1.00	0.50	0.78	
	Reviews	0.00	0.00	0.00	0.89	
50	Belle Letters	0.53	0.31	0.39	0.76	48.45%
	Editorial	0.40	0.33	0.36	0.92	
	Fiction	0.80	0.77	0.78	0.96	
	Government	0.50	0.60	0.55	0.88	
	Hobbies	0.14	0.50	0.22	0.76	
	Learned	0.50	0.50	0.50	0.82	
	Love	0.00	0.00	0.00	0.64	
	News	0.50	0.57	0.53	0.92	
	Religion	0.33	1.00	0.50	0.81	
	Reviews	0.00	0.00	0.00	0.82	
75	Belle Letters	0.53	0.33	0.41	0.74	47.42%
	Editorial	0.60	0.43	0.50	0.92	
	Fiction	0.80	0.77	0.78	0.95	
	Government	0.50	0.75	0.60	0.90	
	Hobbies	0.14	0.33	0.20	0.76	
	Learned	0.50	0.50	0.50	0.83	
	Love	0.00	0.00	0.00	0.60	
	News	0.25	0.40	0.31	0.90	
	Religion	0.33	1.00	0.50	0.79	
	Reviews	0.00	0.00	0.00	0.82	
100	Belle Letters	0.47	0.28	0.35	0.75	45.36%
	Editorial	0.60	0.38	0.46	0.93	
	Fiction	0.80	0.80	0.80	0.96	
	Government	0.50	0.50	0.50	0.90	
	Hobbies	0.14	0.50	0.22	0.76	
	Learned	0.44	0.47	0.45	0.82	
	Love	0.00	0.00	0.00	0.60	
	News	0.25	0.50	0.33	0.90	
	Religion	0.33	1.00	0.50	0.80	
	Reviews	0.00	0.00	0.00	0.82	

Table 36: Results for 10 classes of Brown10 data set for word2vec document representation models with feature vectors of size 25, 50, 75 and 100

Features	Classes	Precision	Recall	F1	AUC ROC	Accuracy
25	Belle Letters	0.60	0.39	0.47	0.87	61.85%
	Editorial	0.80	0.57	0.67	0.98	
	Fiction	0.92	0.70	0.79	0.98	
	Government	0.50	1.00	0.67	0.92	
	Hobbies	0.29	1.00	0.44	0.94	
	Learned	0.69	0.58	0.63	0.86	
	Love	0.22	0.67	0.33	0.71	
	News	0.50	1.00	0.67	0.89	
	Religion	0.00	0.00	0.00	0.99	
	Reviews	0.67	0.67	0.67	1	
50	Belle Letters	0.6	0.41	0.49	0.86	61.86%
	Editorial	0.60	0.38	0.46	0.94	
	Fiction	0.92	0.77	0.84	0.98	
	Government	0.50	1.00	0.67	0.95	
	Hobbies	0.43	0.60	0.50	0.96	
	Learned	0.69	0.61	0.65	0.86	
	Love	0.33	0.50	0.40	0.74	
	News	0.25	1.00	0.40	0.93	
	Religion	0.00	0.00	0.00	0.96	
	Reviews	1.00	1.00	1.00	1.00	
75	Belle Letters	0.8	0.44	0.57	0.89	60.82%
	Editorial	0.40	0.40	0.40	0.96	
	Fiction	0.96	0.67	0.79	0.97	
	Government	0.33	1.00	0.50	0.90	
	Hobbies	0.29	0.67	0.40	0.93	
	Learned	0.63	0.59	0.61	0.83	
	Love	0.22	1.00	0.36	0.76	
	News	0.38	1.00	0.55	0.88	
	Religion	0.00	0.00	0.00	0.95	
	Reviews	0.67	1.00	0.80	1.00	
100	Belle Letters	0.8	0.43	0.56	0.88	55.67%
	Editorial	0.20	0.50	0.29	0.97	
	Fiction	0.96	0.59	0.73	0.97	
	Government	0.33	0.67	0.44	0.95	
	Hobbies	0.14	1.00	0.25	0.92	
	Learned	0.63	0.56	0.59	0.86	
	Love	0.00	0.00	0.00	0.67	
	News	0.50	1.00	0.67	0.90	
	Religion	0.00	0.00	0.00	0.95	
	Reviews	0.00	0.00	0.00	1.00	

Table 37: Results for 10 classes of Brown10 data set for doc2vec document representation models with feature vectors of size 25, 50, 75 and 100

Features	Classes	Precision	Recall	F1	AUC ROC	Accuracy
200	Belle Letters	0.6	0.33	0.43	0.84	49.48%
	Editorial	0.20	1.00	0.33	0.95	
	Fiction	0.96	0.56	0.71	0.97	
	Government	0.00	0.00	0.00	0.94	
	Hobbies	0.00	0.00	0.00	0.93	
	Learned	0.69	0.52	0.59	0.83	
	Love	0.00	0.00	0.00	0.70	
	News	0.38	0.75	0.50	0.89	
	Religion	0.00	0.00	0.00	0.96	
	Reviews	0.00	0.00	0.00	1.00	
500	Belle Letters	0.6	0.38	0.46	0.78	48.45%
	Editorial	0.00	0.00	0.00	0.94	
	Fiction	1.00	0.50	0.67	0.96	
	Government	0.00	0.00	0.00	0.87	
	Hobbies	0.00	0.00	0.00	0.90	
	Learned	0.69	0.52	0.59	0.83	
	Love	0.00	0.00	0.00	0.42	
	News	0.25	1.00	0.40	0.90	
	Religion	0.00	0.00	0.00	0.88	
	Reviews	0.00	0.00	0.00	0.95	
1000	Belle Letters	0.40	0.29	0.33	0.77	44.33%
	Editorial	0.00	0.00	0.00	0.91	
	Fiction	1.00	0.49	0.66	0.95	
	Government	0.00	0.00	0.00	0.86	
	Hobbies	0.00	0.00	0.00	0.94	
	Learned	0.69	0.46	0.55	0.83	
	Love	0.00	0.00	0.00	0.65	
	News	0.13	1.00	0.22	0.87	
	Religion	0.00	0.00	0.00	0.91	
	Reviews	0.00	0.00	0.00	0.78	

Table 38: Results for 10 classes of Brown10 data set for doc2vec document representation models with feature vectors of size 200, 500 and 1000

Features	Classes	Precision	Recall	F1	AUC ROC	Accuracy
Average	Belle Letters	0.47	0.29	0.36	0.76	45.36%
	Editorial	0.00	0.00	0.00	0.84	
	Fiction	0.68	0.71	0.69	0.90	
	Government	0.33	1.00	0.50	0.80	
	Hobbies	0.00	0.00	0.00	0.82	
	Learned	0.56	0.39	0.46	0.78	
	Love	0.33	0.38	0.35	0.59	
	News	0.50	0.50	0.50	0.86	
	Religion	0.00	0.00	0.00	0.31	
	Reviews	0.67	0.33	0.44	0.95	
Quantiles	Belle Letters	0.47	0.26	0.33	0.69	40.21%
	Editorial	0.00	0.00	0.00	0.88	
	Fiction	0.60	0.58	0.59	0.90	
	Government	0.00	0.00	0.00	0.89	
	Hobbies	0.00	0.00	0.00	0.75	
	Learned	0.69	0.39	0.50	0.80	
	Love	0.00	0.00	0.00	0.53	
	News	0.50	0.50	0.50	0.81	
	Religion	0.00	0.00	0.00	0.57	
	Reviews	0.67	0.67	0.67	0.95	
Histogram	Belle Letters	0.53	0.27	0.36	0.69	39.2%
	Editorial	0.00	0.00	0.00	0.86	
	Fiction	0.64	0.64	0.64	0.91	
	Government	0.00	0.00	0.00	0.85	
	Hobbies	0.00	0.00	0.00	0.83	
	Learned	0.63	0.38	0.48	0.78	
	Love	0.00	0.00	0.00	0.55	
	News	0.38	0.50	0.43	0.83	
	Religion	0.00	0.00	0.00	0.34	
	Reviews	0.33	0.20	0.25	0.96	

Table 39: Results for 10 classes of Brown10 data set for graph-of-word representation models using averaging, histograms and quantiles for aggregations of local measures

4.4 The results for Brown15: 15 classes

Accuracy	Per class Precision	Per class Recall	Per class F1	Per class AUC ROC	Per class all
doc2vec_75 1	doc2vec_50 2	doc2vec_75 1	doc2vec_75 1	doc2vec_50 1	tf 8.5
tf 3	doc2vec_75 2	tf 3.5	tf 4	tf 9	tf-idf 8.5
doc2vec_100 3	word2vec_50 2	tf-idf 3.5	doc2vec_100 4	tf-idf 9	doc2vec_100 8.5
doc2vec_50 3	tf 10	doc2vec_100 3.5	doc2vec_25 4	doc2vec_100 9	doc2vec_1000 8.5
doc2vec_25 5	tf-idf 10	doc2vec_25 3.5	doc2vec_50 4	doc2vec_1000 9	doc2vec_200 8.5
tf-idf 6	doc2vec_100 10	doc2vec_1000 11	word2vec_50 4	doc2vec_200 9	doc2vec_25 8.5
doc2vec_200 7.5	doc2vec_1000 10	doc2vec_200 11	tf-idf 11.5	doc2vec_25 9	doc2vec_50 8.5
word2vec_50 7.5	doc2vec_200 10	doc2vec_50 11	doc2vec_1000 11.5	doc2vec_500 9	doc2vec_500 8.5
word2vec_100 9.5	doc2vec_25 10	doc2vec_500 11	doc2vec_200 11.5	doc2vec_75 9	doc2vec_75 8.5
word2vec_25 9.5	doc2vec_500 10	gow_avg 11	doc2vec_500 11.5	gow_avg 9	gow_avg 8.5
gow_avg 11	gow_avg 10	gow_histogram 11	gow_avg 11.5	gow_histogram 9	gow_histogram 8.5
word2vec_75 12	gow_histogram 10	gow_quantiles 11	gow_histogram 11.5	gow_quantiles 9	gow_quantiles 8.5
doc2vec_500 13	gow_quantiles 10	word2vec_100 11	gow_quantiles 11.5	word2vec_100 9	word2vec_100 8.5
doc2vec_1000 14	word2vec_100 10	word2vec_25 11	word2vec_100 11.5	word2vec_25 9	word2vec_25 8.5
gow_histogram 15.5	word2vec_25 10	word2vec_50 11	word2vec_25 11.5	word2vec_50 9	word2vec_50 8.5
gow_quantiles 15.5	word2vec_75 10	word2vec_75 11	word2vec_75 11.5	word2vec_75 9	word2vec_75 8.5

Table 40: Ranks of document representation models for Brown15 dataset in terms of accuracy, per class precision, per class recall, per class F1 and per class AUC ROC

Features	Classes	Precision	Recall	F1	AUC ROC	Accuracy
tf	Adventure	0.60	0.43	0.50	0.96	46.24%
	Belle Letters	0.80	0.46	0.59	0.87	
	Editorial	0.80	0.67	0.73	0.93	
	Fiction	0.20	0.17	0.18	0.90	
	Government	0.33	0.50	0.40	0.83	
	Hobbies	0.43	0.60	0.50	0.92	
	Humor	0.00	0.00	0.00	0.97	
	Learned	0.50	0.40	0.44	0.80	
	Love	0.00	0.00	0.00	0.80	
	Mistic	0.25	0.50	0.33	0.96	
	News	0.63	0.63	0.63	0.92	
	Religion	0.00	0.00	0.00	0.79	
	Reviews	0.67	0.67	0.67	0.99	
	Romance	0.40	0.50	0.44	0.98	
Sci-Fi	0.00	0.00	0.00	0.57		
tf-idf	Adventure	0.4	0.25	0.31	0.91	41.94%
	Belle Letters	0.53	0.40	0.46	0.84	
	Editorial	0.40	0.67	0.50	0.88	
	Fiction	0.40	0.33	0.36	0.90	
	Government	0.50	0.50	0.50	0.78	
	Hobbies	0.29	0.33	0.31	0.90	
	Humor	0.00	0.00	0.00	0.99	
	Learned	0.38	0.43	0.40	0.78	
	Love	0.22	0.67	0.33	0.83	
	Mistic	0.50	0.40	0.44	0.96	
	News	0.75	0.46	0.57	0.90	
	Religion	0.33	0.33	0.33	0.85	
	Reviews	0.67	0.67	0.67	0.98	
	Romance	0.20	0.33	0.25	0.95	
Sci-Fi	0.00	0.00	0.00	0.97		

Table 41: Results for 15 classes of Brown15 data set for bag-of-words document representation models

Features	Classes	Precision	Recall	F1	AUC	ROC	Accuracy
25	Adventure	0.20	0.25	0.22	0.93		37.63%
	Belle Letters	0.53	0.35	0.42	0.77		
	Editorial	0.60	0.43	0.50	0.77		
	Fiction	0.40	0.25	0.31	0.86		
	Government	0.33	0.40	0.36	0.90		
	Hobbies	0.14	0.20	0.17	0.72		
	Humor	0.00	0.00	0.00	0.87		
	Learned	0.38	0.46	0.41	0.80		
	Love	0.11	0.13	0.12	0.59		
	Mistic	0.50	0.50	0.50	0.94		
	News	0.75	0.60	0.67	0.96		
	Religion	0.33	1.00	0.50	0.61		
	Reviews	0.00	0.00	0.00	0.88		
	Romance	0.40	0.67	0.50	0.98		
	Sci-Fi	0.00	0.00	0.00	0.27		
50	Adventure	0.40	0.50	0.44	0.91		40.9%
	Belle Letters	0.47	0.35	0.40	0.80		
	Editorial	0.60	0.43	0.50	0.79		
	Fiction	0.60	0.30	0.40	0.88		
	Government	0.33	0.40	0.36	0.91		
	Hobbies	0.14	0.17	0.15	0.75		
	Humor	0.00	0.00	0.00	0.97		
	Learned	0.50	0.53	0.52	0.78		
	Love	0.11	0.14	0.13	0.64		
	Mistic	0.50	0.67	0.57	0.96		
	News	0.63	0.56	0.59	0.97		
	Religion	0.33	0.50	0.40	0.61		
	Reviews	0.33	0.50	0.40	0.88		
	Romance	0.40	0.67	0.50	0.99		
	Sci-Fi	0.00	0.00	0.00	0.28		
75	Adventure	0.20	0.33	0.25	0.92		34.4%
	Belle Letters	0.60	0.41	0.49	0.77		
	Editorial	0.60	0.43	0.50	0.75		
	Fiction	0.20	0.13	0.15	0.87		
	Government	0.33	0.50	0.40	0.91		
	Hobbies	0.14	0.17	0.15	0.75		
	Humor	0.00	0.00	0.00	0.73		
	Learned	0.44	0.41	0.42	0.79		
	Love	0.11	0.20	0.14	0.54		
	Mistic	0.25	0.25	0.25	0.91		
	News	0.50	0.40	0.44	0.96		
	Religion	0.33	0.50	0.40	0.66		
	Reviews	0.00	0.00	0.00	0.81		
	Romance	0.20	0.25	0.22	0.97		
	Sci-Fi	0.00	0.00	0.00	0.59		
100	Adventure	0.40	0.50	0.44	0.91		37.6%
	Belle Letters	0.47	0.37	0.41	0.79		
	Editorial	0.60	0.38	0.46	0.76		
	Fiction	0.60	0.25	0.35	0.88		
	Government	0.33	0.50	0.40	0.94		
	Hobbies	0.14	0.17	0.15	0.74		
	Humor	0.00	0.00	0.00	0.90		
	Learned	0.44	0.47	0.45	0.79		
	Love	0.11	0.13	0.12	0.59		
	Mistic	0.25	0.50	0.33	0.93		
	News	0.50	0.57	0.53	0.97		
	Religion	0.33	0.50	0.40	0.65		
	Reviews	0.33	0.50	0.40	0.86		
	Romance	0.40	0.50	0.44	0.99		
	Sci-Fi	0.00	0.00	0.00	0.55		

Table 42: Results for 15 classes of Brown15 data set for word2vec document representation models with feature vectors of size 25, 50, 75 and 100

Features	Classes	Precision	Recall	F1	AUC	ROC	Accuracy
25	Adventure	0.40	0.40	0.40	0.94		44.08%
	Belle Letters	0.53	0.47	0.50	0.87		
	Editorial	0.40	0.40	0.40	0.96		
	Fiction	0.40	0.29	0.33	0.94		
	Government	0.17	0.25	0.20	0.91		
	Hobbies	0.43	0.50	0.46	0.88		
	Humor	0.00	0.00	0.00	0.65		
	Learned	0.56	0.45	0.50	0.81		
	Love	0.22	0.33	0.27	0.69		
	Mistic	0.50	0.50	0.50	0.98		
	News	0.88	0.58	0.70	0.94		
	Religion	0.00	0.00	0.00	0.82		
	Reviews	0.67	0.67	0.67	0.98		
	Romance	0.20	0.25	0.22	0.97		
	Sci-Fi	0.00	0.00	0.00	0.96		
50	Adventure	0.4	0.4	0.4	0.93		46.24%
	Belle Letters	0.67	0.48	0.56	0.85		
	Editorial	0.60	0.50	0.55	0.94		
	Fiction	0.00	0.00	0.00	0.90		
	Government	0.33	0.50	0.40	0.92		
	Hobbies	0.43	0.43	0.43	0.89		
	Humor	0.00	0.00	0.00	0.96		
	Learned	0.56	0.50	0.53	0.84		
	Love	0.33	0.43	0.38	0.78		
	Mistic	0.50	0.40	0.44	0.96		
	News	0.75	0.67	0.71	0.92		
	Religion	0.00	0.00	0.00	0.82		
	Reviews	0.67	0.67	0.67	0.99		
	Romance	0.20	0.25	0.22	0.99		
	Sci-Fi	0.00	0.00	0.00	0.93		
75	Adventure	0.4	0.4	0.4	0.90		53.76%
	Belle Letters	0.80	0.48	0.60	0.90		
	Editorial	0.60	0.75	0.67	0.89		
	Fiction	0.40	0.33	0.36	0.91		
	Government	0.33	1.00	0.50	0.93		
	Hobbies	0.43	0.50	0.46	0.87		
	Humor	0.00	0.00	0.00	0.47		
	Learned	0.69	0.50	0.58	0.85		
	Love	0.22	0.40	0.29	0.73		
	Mistic	0.50	0.67	0.57	0.97		
	News	0.88	0.78	0.82	0.91		
	Religion	0.00	0.00	0.00	0.86		
	Reviews	0.67	1.00	0.80	1.00		
	Romance	0.40	0.50	0.44	0.97		
	Sci-Fi	0.00	0.00	0.00	0.95		
100	Adventure	0.4	0.4	0.4	0.93		46.23%
	Belle Letters	0.80	0.41	0.55	0.88		
	Editorial	0.20	0.33	0.25	0.90		
	Fiction	0.20	0.20	0.20	0.91		
	Government	0.33	0.67	0.44	0.94		
	Hobbies	0.29	0.50	0.36	0.87		
	Humor	0.00	0.00	0.00	0.66		
	Learned	0.50	0.40	0.44	0.81		
	Love	0.22	0.67	0.33	0.69		
	Mistic	0.50	0.67	0.57	0.97		
	News	0.88	0.64	0.74	0.94		
	Religion	0.00	0.00	0.00	0.91		
	Reviews	0.67	1.00	0.80	0.99		
	Romance	0.40	0.40	0.40	0.95		
	Sci-Fi	0.00	0.00	0.00	0.90		

Table 43: Results for 15 classes of Brown15 data set for doc2vec document representation models with feature vectors of size 25, 50, 75 and 100

Features	Classes	Precision	Recall	F1	AUC ROC	Accuracy
200	Adventure	0.4	0.67	0.5	0.93	40.86%
	Belle Letters	0.80	0.33	0.47	0.86	
	Editorial	0.00	0.00	0.00	0.83	
	Fiction	0.20	0.33	0.25	0.89	
	Government	0.17	0.50	0.25	0.92	
	Hobbies	0.14	0.33	0.20	0.82	
	Humor	0.00	0.00	0.00	0.70	
	Learned	0.63	0.34	0.44	0.80	
	Love	0.11	0.50	0.18	0.66	
	Mistic	0.25	0.50	0.33	0.98	
	News	0.88	0.78	0.82	0.93	
	Religion	0.00	0.00	0.00	0.99	
	Reviews	0.00	0.00	0.00	0.79	
	Romance	0.40	0.50	0.44	0.94	
	Sci-Fi	0.00	0.00	0.00	0.87	
500	Adventure	0.6	1	0.75	0.95	33.33%
	Belle Letters	0.80	0.24	0.38	0.84	
	Editorial	0.00	0.00	0.00	0.81	
	Fiction	0.00	0.00	0.00	0.88	
	Government	0.33	1.00	0.50	0.84	
	Hobbies	0.00	0.00	0.00	0.89	
	Humor	0.00	0.00	0.00	0.82	
	Learned	0.50	0.32	0.39	0.84	
	Love	0.00	0.00	0.00	0.67	
	Mistic	0.00	0.00	0.00	0.91	
	News	0.75	0.75	0.75	0.92	
	Religion	0.00	0.00	0.00	0.71	
	Reviews	0.00	0.00	0.00	0.84	
	Romance	0.00	0.00	0.00	0.90	
	Sci-Fi	0.00	0.00	0.00	0.85	
1000	Adventure	0.20	0.50	0.29	0.93	31.18%
	Belle Letters	0.87	0.26	0.40	0.84	
	Editorial	0.00	0.00	0.00	0.82	
	Fiction	0.00	0.00	0.00	0.83	
	Government	0.00	0.00	0.00	0.87	
	Hobbies	0.00	0.00	0.00	0.76	
	Humor	0.00	0.00	0.00	0.28	
	Learned	0.69	0.35	0.47	0.79	
	Love	0.00	0.00	0.00	0.79	
	Mistic	0.00	0.00	0.00	0.98	
	News	0.38	0.60	0.46	0.91	
	Religion	0.00	0.00	0.00	0.79	
	Reviews	0.00	0.00	0.00	0.98	
	Romance	0.20	0.25	0.22	0.94	
	Sci-Fi	0.00	0.00	0.00	0.91	

Table 44: Results for 15 classes of Brown15 data set for doc2vec document representation models with feature vectors of size 200, 500 and 1000

Features	Classes	Precision	Recall	F1	AUC ROC	Accuracy
Average	Adventure	0.4	1	0.57	0.90	35.48%
	Belle Letters	0.53	0.36	0.43	0.80	
	Editorial	0.60	0.38	0.46	0.91	
	Fiction	0.20	0.17	0.18	0.75	
	Government	0.33	1.00	0.50	0.91	
	Hobbies	0.14	0.50	0.22	0.81	
	Humor	0.00	0.00	0.00	0.74	
	Learned	0.56	0.38	0.45	0.76	
	Love	0.00	0.00	0.00	0.44	
	Mistic	0.25	0.33	0.29	0.92	
	News	0.63	0.42	0.50	0.91	
	Religion	0.00	0.00	0.00	0.64	
	Reviews	0.00	0.00	0.00	0.94	
	Romance	0.20	0.25	0.22	0.87	
	Sci-Fi	0.00	0.00	0.00	0.74	
Quantiles	Adventure	0	0	0	0.90	29.03%
	Belle Letters	0.67	0.40	0.50	0.81	
	Editorial	0.40	0.40	0.40	0.91	
	Fiction	0.20	0.14	0.17	0.84	
	Government	0.17	0.25	0.20	0.78	
	Hobbies	0.00	0.00	0.00	0.83	
	Humor	0.00	0.00	0.00	0.85	
	Learned	0.44	0.30	0.36	0.79	
	Love	0.11	0.20	0.14	0.54	
	Mistic	0.00	0.00	0.00	0.92	
	News	0.63	0.38	0.48	0.90	
	Religion	0.00	0.00	0.00	0.70	
	Reviews	0.00	0.00	0.00	0.96	
	Romance	0.00	0.00	0.00	0.79	
	Sci-Fi	0.00	0.00	0.00	0.66	
Histogram	Adventure	0	0	0	0.93	29.03%
	Belle Letters	0.60	0.33	0.43	0.75	
	Editorial	0.20	0.20	0.20	0.88	
	Fiction	0.00	0.00	0.00	0.77	
	Government	0.00	0.00	0.00	0.82	
	Hobbies	0.14	0.50	0.22	0.77	
	Humor	0.00	0.00	0.00	0.95	
	Learned	0.56	0.35	0.43	0.80	
	Love	0.00	0.00	0.00	0.60	
	Mistic	0.25	0.25	0.25	0.87	
	News	0.63	0.50	0.56	0.88	
	Religion	0.00	0.00	0.00	0.66	
	Reviews	0.33	0.33	0.33	0.97	
	Romance	0.00	0.00	0.00	0.85	
	Sci-Fi	0.00	0.00	0.00	0.63	

Table 45: Results for 15 classes of Brown15 data set for graph-of-word representation models using averaging, histograms and quantiles for aggregations of local measures