

Apples to Apples: A Systematic Evaluation of Topic Models [APPENDIX]

Anonymous RANLP 2021 submission

1 Coherence scores on 20NG (20 topics)

name	C_v	C_{NPMI}	C_{UMASS}	C_{UCI}
CTM	0.56	-0.04	-5.78	-3.09
D2T	0.57	0.01	-2.94	-1.56
GSDMM	0.50	0.00	-3.86	-2.02
HDP	0.44	-0.09	-5.59	-3.85
LDA	0.64	0.10	-1.98	0.27
LFTM	0.53	-0.01	-2.97	-1.46
LSI	0.53	0.03	-3.25	-1.37
NMF	0.61	0.10	-2.37	-0.03
PVTM	0.54	0.06	-1.63	0.21

2 Coherence scores on AFP (17 topics)

name	C_v	C_{NPMI}	C_{UMASS}	C_{UCI}
CTM	0.54	-0.05	-6.56	-2.75
D2T	0.58	0.06	-2.25	-0.02
GSDMM	0.51	0.09	-1.72	0.70
HDP	0.42	0.02	-2.23	-0.20
LDA	0.65	0.11	-1.40	0.80
LFTM	0.59	0.06	-1.97	0.11
LSI	0.58	0.07	-1.80	0.09
NMF	0.67	0.13	-1.27	0.95
PVTM	0.52	0.07	-1.16	0.49

3 Coherence scores on Yahoo unbal. (26 topics)

name	C_v	C_{NPMI}	C_{UMASS}	C_{UCI}
CTM	0.46	-0.22	-10.84	-6.17
D2T	0.37	-0.02	-3.22	-0.81
GSDMM	0.47	0.06	-2.57	0.26
HDP	0.48	-0.23	-14.15	-6.76
LDA	0.61	0.10	-2.88	0.47
LFTM	0.52	0.01	-4.35	-1.27
LSI	0.38	-0.01	-2.96	-0.48
NMF	0.51	0.04	-2.19	0.23
PVTM	0.52	0.04	-1.78	0.28

4 Coherence scores on Yahoo balanced (26 topics)

name	C_v	C_{NPMI}	C_{UMASS}	C_{UCI}
CTM	0.44	-0.18	-9.34	-5.26
D2T	0.33	-0.04	-3.32	-1.00
GSDMM	0.51	0.05	-2.41	0.15
HDP	0.46	-0.23	-13.41	-6.56
LDA	0.62	0.10	-2.75	0.64
LFTM	0.54	0.04	-3.83	-0.71
LSI	0.40	-0.00	-2.81	-0.49
NMF	0.51	0.03	-2.68	-0.18
PVTM	0.52	0.05	-1.85	0.31

5 Ground truth scores on 20NG (20 topics)

In all following tables, V-measure scores are not reported because equivalent to NMI.

	Purity	Homog.	Comple.	NMI
CTM	0.25	0.21	0.22	0.21
D2T	0.26	0.17	0.17	0.17
GSDMM	0.18	0.20	0.39	0.27
HDP	0.12	0.15	0.54	0.24
LDA	0.49	0.54	0.58	0.56
LFTM	0.33	0.32	0.36	0.34
LSI	0.13	0.08	0.34	0.13
NMF	0.15	0.08	0.11	0.09
PVTM	0.61	0.58	0.60	0.59

6 Ground truth scores on AFP (17 topics)

	Purity	Homog.	Comple.	NMI
CTM	0.20	0.22	0.54	0.31
D2T	0.17	0.19	0.46	0.26
GSDMM	0.14	0.07	0.49	0.12
HDP	0.25	0.19	0.84	0.31
LDA	0.23	0.31	0.78	0.44
LFTM	0.23	0.26	0.73	0.39
LSI	0.17	0.13	0.56	0.22
NMF	0.14	0.12	0.37	0.19
PVTM	0.25	0.27	0.72	0.40

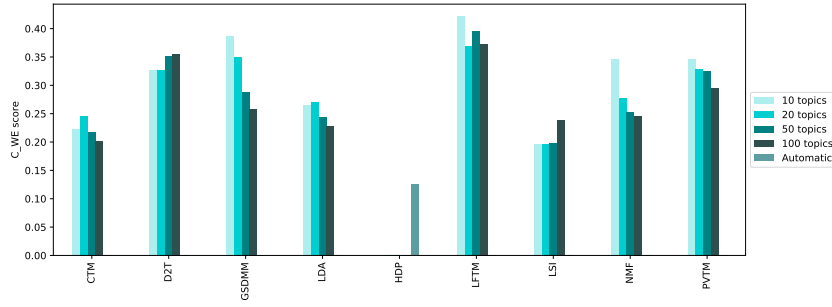


Figure 1: Word embedding coherence on 20NG when varying the number of topics

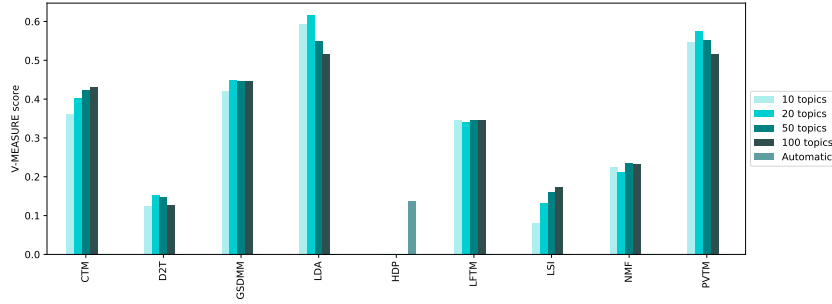


Figure 2: V-measure on 20NG when varying the number of topics

7 Ground truth scores on Yahoo balanced (26 topics)

	Purity	Homog.	Comple.	NMI
CTM	0.06	0.01	0.01	0.01
D2T	0.08	0.02	0.02	0.02
GSDMM	0.30	0.27	0.33	0.29
LDA	0.39	0.32	0.33	0.33
LFTM	0.34	0.27	0.27	0.27
LSI	0.15	0.09	0.15	0.11
HDP	0.06	0.01	0.09	0.02
NMF	0.11	0.06	0.06	0.06
PVTM	0.18	0.14	0.14	0.14

9 Embedding-based coherence scores

topics	20NG		AFP		Yahoo	
	20	50	17	50	bal.	unb.
CTM	0.17	0.19	0.15	0.10	0.22	0.18
D2T	0.26	0.24	0.21	0.23	0.23	0.27
GSDMM	0.18	0.11	0.29	0.18	0.27	0.26
HDP	0.14	0.14	0.19	0.20	0.12	0.11
LDA	0.27	0.24	0.31	0.31	0.34	0.33
LFTM	0.37	0.38	0.39	0.39	0.42	0.43
LSI	0.18	0.18	0.21	0.20	0.28	0.28
NMF	0.23	0.23	0.31	0.31	0.38	0.42
PVTM	0.30	0.29	0.35	0.33	0.41	0.43

8 Ground truth scores on Yahoo unbalanced (26 topics)

	Purity	Homog.	Comple.	NMI
CTM	0.15	0.02	0.01	0.02
D2T	0.16	0.03	0.03	0.03
GSDMM	0.33	0.23	0.29	0.26
HDP	0.16	0.02	0.07	0.03
LDA	0.47	0.34	0.32	0.33
LFTM	0.38	0.24	0.22	0.23
LSI	0.23	0.09	0.13	0.11
NMF	0.20	0.07	0.06	0.07
PVTM	0.32	0.20	0.19	0.19