SemEval-2016 Task 4: Sentiment Analysis in Twitter

Preslav Nakov♣, Alan Ritter⋄, Sara Rosenthal⋄, Fabrizio Sebastiani♣, Veselin Stoyanov♠

♣Qatar Computing Research Institute, Hamad bin Khalifa University, Qatar

⋄Department of Computer Science and Engineering, The Ohio State University, USA

⋄Department of Computer Science, Columbia University, USA

♠Facebook, USA

Fabrizio Sebastiani is currently on leave from Consiglio Nazionale delle Ricerche, Italy.

		20	13		2014			2016
#	System	Tweet	SMS	Tweet	Tweet sarcasm	Live- Journal	Tweet	Tweet
1	SwissCheese	0.700_{5}	0.6372	0.716 ₅	0.5661	0.6957	0.6711	0.633 ₁
2	SENSEI-LIF	0.706_{4}	0.634_{3}	0.744_2	0.467_{8}	0.741_{1}	0.662_2	0.630_2
3	unimelb	0.6877	0.593_{9}	0.706_{7}	0.449_{11}	0.683_{9}	0.651_4	0.617 ₃
4	INESC-ID	0.723_2	0.609_{6}	0.727_{3}	0.554_{3}	0.702_{4}	0.657_{3}	0.610_{4}
5	aueb.twitter.sentiment	0.6668	0.618_{5}	0.708_{6}	0.410_{17}	0.695_{7}	0.623_{7}	0.605 ₅
6	SentiSys	0.714_{3}	0.633_{4}	0.723_4	0.515_{5}	0.726_{2}	0.644_{5}	0.598 ₆
7	I2RNTU	0.6936	0.597_{7}	0.680_{8}	0.469_{6}	0.696_{6}	0.638_{6}	0.5967
8	INSIGHT-1	0.602_{16}	0.582_{12}	0.644_{16}	0.391_{23}	0.559_{23}	0.595_{16}	0.593 ₈
9	twise	0.610_{15}	0.540_{17}	0.645_{14}	0.450_{10}	0.649_{13}	0.6218	0.586 ₉
10	ECNU	0.643 ₁₀	0.593_{9}	0.662_{9}	0.425_{14}	0.663_{10}	0.606_{11}	0.585 ₁₀
11	NTNUSentEval	0.623_{12}	0.641_{1}	0.651_{11}	0.427_{13}	0.719_{3}	0.599_{13}	0.583 ₁₁
12	MDSENT	0.589_{19}	0.509_{21}	0.587_{20}	0.386_{24}	0.606_{19}	0.593_{18}	0.580_{12}
	CUFE	0.642_{11}	0.596_{8}	0.662_{9}	0.466_{9}	0.697_{5}	0.598_{14}	0.580 ₁₂
14	THUIR	0.616_{13}	0.575_{14}	0.648_{12}	0.399_{20}	0.640_{15}	0.617_{10}	0.576 ₁₄
	PUT	0.565_{21}	0.511_{20}	0.614_{19}	0.360_{27}	0.648_{14}	0.597_{15}	0.576 ₁₄
16	LYS	0.650_{9}	0.579_{13}	0.647_{13}	0.407_{18}	0.655_{11}	0.603_{12}	0.575 ₁₆
17	IIP	0.598_{17}	0.465_{23}	0.645_{14}	0.405_{19}	0.640_{15}	0.619_9	0.574 ₁₇
18	UniPI	0.592_{18}	0.585_{11}	0.627_{18}	0.381_{25}	0.654_{12}	0.586_{19}	0.571 ₁₈
	DIEGOLab16	0.813_{1}	0.543_{16}	0.759_1	0.562_{2}	0.615_{18}	0.595_{16}	0.571 ₁₈
20	GTI	0.612_{14}	0.524_{18}	0.639_{17}	0.468_{7}	0.623_{17}	0.584_{20}	0.539 ₂₀
21	OPAL	0.567_{20}	0.562_{15}	0.556_{23}	0.395_{21}	0.593_{21}	0.531_{21}	0.505_{21}
22	DSIC-ELIRF	0.494_{25}	0.404_{26}	0.546_{26}	0.342_{29}	0.517_{24}	0.531_{21}	0.502_{22}
23	UofL	0.490_{26}	0.443_{24}	0.547_{25}	0.372_{26}	0.574_{22}	0.502_{25}	0.499_{23}
	ELiRF	0.462_{28}	0.408_{25}	0.514_{28}	0.310_{33}	0.493_{25}	0.493_{26}	0.499_{23}
25	ISTI-CNR	0.538_{22}	0.492_{22}	0.572_{21}	0.327_{30}	0.598_{20}	0.508_{24}	0.494 ₂₅
26	SteM	0.518_{23}	0.315_{29}	0.571_{22}	0.320_{32}	0.405_{28}	0.517_{23}	0.478 ₂₆
27	Tweester	0.506_{24}	0.340_{28}	0.529_{27}	0.540_{4}	0.379_{29}	0.479_{28}	0.455 ₂₇
28	Minions	0.489_{27}	0.521_{19}	0.554_{24}	0.420_{16}	0.475_{26}	0.481_{27}	0.415 ₂₈
29	aicyber	0.418_{29}	0.361_{27}	0.457_{29}	0.326_{31}	0.440_{27}	0.432_{29}	0.402 ₂₉
30	mib	0.394_{30}	0.310_{30}	0.415_{31}	0.352_{28}	0.359_{31}	0.413_{31}	0.401 ₃₀
31	VCU-TSA	0.383_{31}	0.307_{31}	0.444_{30}	0.425_{14}	0.336_{32}	0.416_{30}	0.372 ₃₁
32	SentimentalITists	0.339_{33}	0.238_{33}	0.393_{32}	0.288_{34}	0.323_{34}	0.343_{33}	0.339 ₃₂
33	Wisers_Research	0.355_{32}	0.284_{32}	0.393_{32}	0.430_{12}	0.366_{30}	0.377_{32}	0.330 ₃₃
34	CICBUAPnlp	0.193_{34}	0.193_{34}	0.335_{34}	0.393_{22}	0.326_{33}	0.303_{34}	0.303 ₃₄
	baseline (positive)	0.292	0.190	0.346	0.277	0.272	0.303	0.255

Table 1: Historical results for Subtask A "Message Polarity Classification". The systems are ordered by their score on the Twitter2016 test dataset; the rankings on the individual datasets are indicated with a subscript.

		Tweet 2016			
#	System	AvgF1	AvgR	Acc	
1	SwissCheese	0.6331	0.6672	0.6461	
2	SENSEI-LIF	0.630_2	0.670_1	0.6177	
3	unimelb	0.617 ₃	0.6415	0.6168	
4	INESC-ID	0.610_4	0.662_{3}	0.600_{10}	
5	aueb.twitter.sentiment	0.605 ₅	0.6444	0.6296	
6	SentiSys	0.5986	0.6415	0.609_9	
7	I2RNTU	0.5967	0.6377	0.593 ₁₂	
8	INSIGHT-1	0.593 ₈	0.616_{12}	0.635_{5}	
9	twise	0.586 ₉	0.598 ₁₇	0.528 ₂₄	
10	ECNU	0.585 ₁₀	0.617 ₁₁	0.571 ₁₇	
11	NTNUSentEval	0.583 ₁₁	0.619_{9}	0.6432	
12	MDSENT	0.580 ₁₂	0.592 ₁₈	0.545_{20}	
	CUFE	0.580 ₁₂	0.619_{9}	0.6374	
14	THUIR	0.576 ₁₄	0.605_{16}	0.596 ₁₁	
	PUT	0.576 ₁₄	0.607 ₁₄	0.584 ₁₄	
16	LYS	0.575 ₁₆	0.615 ₁₃	0.585 ₁₃	
17	IIP	0.574 ₁₇	0.579_{19}	0.537 ₂₃	
18	UniPI	0.571 ₁₈	0.607_{14}	0.639_3	
	DIEGOLab16	0.571 ₁₈	0.6248	0.582_{15}	
20	GTI	0.539 ₂₀	0.557_{21}	0.518_{26}	
21	OPAL	0.505 ₂₁	0.560_{20}	0.541_{22}	
22	DSIC-ELIRF	0.502_{22}	0.511_{25}	0.513 ₂₇	
23	UofL	0.499 ₂₃	0.537 ₂₂	0.572 ₁₆	
	ELiRF	0.499 ₂₃	0.516_{24}	0.543_{21}	
25	ISTI-CNR	0.494 ₂₅	0.529_{23}	0.567 ₁₈	
26	SteM	0.478 ₂₆	0.496_{27}	0.452_{31}	
27	Tweester	0.455 ₂₇	0.503_{26}	0.523_{25}	
28	Minions	0.415 ₂₈	0.485_{28}	0.556_{19}	
29	aicyber	0.402 ₂₉	0.45729	0.506_{28}	
30	mib	0.401 ₃₀	0.438 ₃₀	0.480_{29}	
31	VCU-TSA	0.372 ₃₁	0.390_{32}	0.382_{32}	
32	SentimentalITists	0.339 ₃₂	0.424 ₃₁	0.480_{29}	
33	Wisers_Research	0.330 ₃₃	0.333_{34}	0.298_{34}	
34	CICBUAPnlp	0.303 ₃₄	0.377 ₃₃	0.374 ₃₃	
	baseline (positive)	0.255	0.333	0.342	

Table 2: Results for Subtask A "Message Polarity Classification" (on test-2016 only). The systems are ordered by their score on the Twitter2016 test dataset; the rankings on the individual datasets are indicated with a subscript.

#	System	AvgR	AvgF1	Acc
1	Tweester	0.797 ₁	0.799_1	0.8623
2	LYS	0.791_2	0.720_{10}	0.762_{17}
3	thecerealkiller	0.784_{3}	0.762_{5}	0.823_{9}
4	ECNU	0.768 ₄	0.770_{4}	0.843_{5}
5	INSIGHT-1	0.767 ₅	0.786_{3}	0.864_2
6	PUT	0.763 ₆	0.732_{8}	0.794_{14}
7	unimelb	0.758 ₇	0.788_{2}	0.870_{1}
8	twise	0.756 ₈	0.752_{6}	0.826_{8}
9	GTI	0.736 ₉	0.731_9	0.811_{11}
10	finki	0.720 ₁₀	0.748_{7}	0.848_{4}
11	pkudblab	0.689 ₁₁	0.716_{11}	0.832_{7}
12	CUFE	0.679 ₁₂	0.708_{12}	0.834_{6}
13	ISTI-CNR	0.671 ₁₃	0.690_{13}	0.811_{11}
14	SwissCheese	0.648 ₁₄	0.674_{14}	0.820_{10}
15	SentimentalITists	0.624 ₁₅	0.643_{15}	0.802_{13}
16	PotTS	0.618 ₁₆	0.610_{17}	0.712_{18}
17	OPAL	0.616 ₁₇	0.633_{16}	0.792_{15}
18	Wisers_Research	0.522 ₁₈	0.502_{18}	0.577_{19}
19	VCU-TSA	0.502 ₁₉	0.448_{19}	0.775_{16}
	baseline (positive)	0.500	0.438	0.778

Table 3: Results for Subtask B "Tweet classification according to a two-point scale". The systems are ordered by their ρ^{PN} score (higher is better).

#	System	\mathbf{MAE}^{M}	\mathbf{MAE}^{μ}
1	twise	0.719 ₁	0.6325
2	ECNU	0.806_2	0.7268
3	PUT	0.860_{3}	0.7739
4	LYS	0.864_{4}	0.6947
5	finki	0.869 ₅	0.6726
6	INSIGHT-1	1.006_{6}	0.607 ₃
7	ISTI-CNR	1.074_{7}	0.580_1
8	YZU-NLP	1.111 ₈	0.588_2
9	SentimentalITists	1.148 ₉	0.6254
10	PotTS	1.237 ₁₀	0.860_{10}
11	pkudblab	1.697 ₁₁	1.300 ₁₁
	baseline (0)	1.200	0.537

Table 4: Results for Subtask C "Tweet classification according to a five-point scale". The systems are ordered by their MAE^M score (lower is better).

#	System	KLD	AE	RAE
1	finki	0.0341	0.0741	0.7073
2	LYS	0.053_2	0.099_{4}	0.8445
	twise	0.053_2	0.101_{5}	0.8646
4	INSIGHT-1	0.054_{4}	0.085_2	0.4231
5	GTI	0.055_{5}	0.104_{6}	1.200_{10}
	QCRI	0.055_{5}	0.095_{3}	0.8646
7	NRU-HSE	0.084_{7}	0.120_{8}	0.7674
8	PotTS	0.094_{8}	0.150_{12}	1.838 ₁₂
9	pkudblab	0.099_{9}	0.109_{7}	0.9478
10	ECNU	0.121 ₁₀	0.148_{11}	1.1719
11	ISTI-CNR	0.127 ₁₁	0.147_{9}	1.371 ₁₁
12	SwissCheese	0.191 ₁₂	0.147_{9}	0.638_2
13	UDLAP	0.261 ₁₃	0.274_{13}	2.973 ₁₃
14	HSENN	0.399 ₁₄	0.336_{14}	3.930_{14}
	baseline (1 0)	0.887	0.242	1.155
	baseline (avg on train+dev+devtest)	0.175	0.184	2.110

Table 5: Results for Subtask D "Tweet quantification according to a two-point scale". The systems are ordered by their KLD score (lower is better).

#	System	Score
1	QCRI	0.243
2	finki	0.316
3	pkudblab	0.331
4	NRU-HSE	0.334
5	ECNU	0.341
6	ISTI-CNR	0.358
7	LYS	0.360
8	INSIGHT-1	0.366
9	HSENN	0.545
10	PotTS	0.818
	baseline (0 0 0 1 0)	0.734
	baseline (avg on train+dev+devtest)	0.474

Table 6: Results for Subtask E "Tweet quantification according to a five-point scale". The systems are ordered by their EMD score (lower is better).

Subtasks	Team ID	Affiliation	Nation	Paper
A	aicyber	aicyber.com	China	i -
A		3	Greece	
		Instituto Politenico Nacional		
A	CICBUAPnlp	Benemrita Universidad Autonoma de Puebla	Mexico	
A B	CUFE		Egypt	
A	DIEGOLab16	•	USA	
A	DSIC-ELIRF	, , , , , , , , , , , , , , , , , , ,	Spain	
ABCDE			China	
A	ELiRF	·	Spain	
BCDE			Macedonia	
A B D	GTI	7 13	Spain	
	HSENN		Russia	
DE		Institute for Infocomm Research, A*STAR	Kussia	
A	I2RNTU	School of Computer Engineering, Nanyang Technological University	Singapore	
A	IIP		India	
А		INESC-ID, Lisboa	muia	
A	INESC-ID	Instituto Superior Tenico, Universidade de Lisboa	Portugal	
		INSIGHT Research Centre, National University of Ireland, Galway		
A B C D E	INSIGHT-1	AYLIEN Inc.	Ireland	
		Universidade da Corua		
A B C D E	LYS	Universidade da Corda Universidade de Vigo	Spain	
A	MDSENT	e	USA	
A	mib		Italy	
A	Minions	. 9	Romania	
A B C D E		Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche		
-	NRU-HSE		•	
		, ,	Russia Norway	
A	NTNUSentEval			
A B	OPAL		Italy	
BCDE	ркиавіав		China	
BCDE	PotTS	University of Potsdam	Germany	
4 D G	DUTE	Retresco GmbH	-	
АВС	PUT	7 27	Poland	(0)
	QCRI (**)		Qatar	(?)
A	SENSEI-LIF	,	France	
АВС	SentimentallTists	, , , , , , , , , , , , , , , , , , ,	Romania	
A	SentiSys	·	France	
			Turkey	
A	SteM		Turkey	
		, , ,	Germany	
A B D	SwissCheese		Switzerland	
В	thecerealkiller		USA	
A	THUIR		China	
		School of ECE, Technical University of Crete		
АВ	Tweester	Department of Informatics, University of Athens	Greece	
	1005101	Signal Analysis and Interpretation Laboratory (SAIL)	2.1000	
		Institute for Language & Speech Processing - ILSP		
A B C D	twise	The state of the s	France	
D	UDLAP		Mexico	
A B	unimelb	•	Australia	
A	UniPI		Italy	
A	UofL	University of Louisville	USA	
A B	VCU-TSA	Virginia Commonwealth University	USA	
A B	Wisers_Research	Wisers Information Limited	Hong Kong SAR, China	ı
C	VZU NI D	Yuan Ze University, Taoyuan	Taiwan	
С	YZU-NLP		China	
		T =		

Table 7: Participating teams (Column 2), their affiliation (Column 3) and nationality (Column 4), the subtasks they have participated in (Column 1), and the paper they have contributed (Column 5). Teams marked with a (**) include some of the SemEval 2016 Task 4 organizers.