

NTCIR-7 MOAT Overview

Yohei Seki, Lun-Wei Ku, David Kirk Evans, Le Sun

Opinion Analysis

- Given a sentence:
 - Does it contain any opinions?
 - What are the opinions
 - Polarity? (Positive, Negative, Neutral)
 - Who expresses the opinion?
 - What is the opinion target?
 - Is the sentence relevant to the topic?

Opinion Analysis

Japan should carefully consider visiting the Yasukuni Shrine and abide by the solemn statement it has made so far, Chinese Foreign Ministry spokesman Sun Yuxi said here Friday.

Opinion Analysis

Japan should carefully consider visiting the Yasukuni Shrine and abide by the solemn statement it has made so far, Chinese Foreign Ministry spokesman Sun Yuxi said here Friday.

Opinion Analysis

Japan should carefully consider visiting the Yasukuni Shrine and abide by the solemn statement it has made so far, Chinese Foreign Ministry spokesman Sun Yuxi said here Friday.

Opinion holder: Chinese Foreign Ministry
spokesman Sun Yuxi

Opinion Analysis

Japan should carefully consider visiting the Yasukuni Shrine and abide by the solemn statement it has made so far, Chinese Foreign Ministry spokesman Sun Yuxi said here Friday.

Opinion holder: Chinese Foreign Ministry
spokesman Sun Yuxi

Negative, Neutral, **Negative**, Neutral

NTCIR 6→7

- Added **Simplified Chinese**
- Added **opinion target** annotation
- **Sub-sentence** opinion annotation
- Relevance judged only for opinionated sentences
- Web-based annotation system for JA

Corpus Annotation

Feature	Value	Level	Req'd?
Opinionated	YES, NO	Sentence	Yes
Relevant	YES, NO	Sentence	No
Polarity	Positive, Neutral, Negative	Clause	No
Opinion Holder	String	Clause	No
Target	String	Clause	No

Corpus Sources

- Japanese: 1998-2001 Mainichi newspaper
- English: 1998-2001 Mainichi Daily News, Korea Times, Xinghua, Hong Kong Standard, Straights times

Corpus Sources

- Trad. Chinese: 1998-2001 | China Times (+ Express), Commercial Times, (United | Central | China | Economic) Daily News, Min Sheng Daily, Star News, United Evening News
- Simp. Chinese: 1998-2001 | Xinhua News and Lianhe Zaobao

Corpus Information

Language	Topics			Documents		
	Sum	Sample	Test	Sum	Sample	Test
Trad. Chinese	17	3	14	246	58	188
Japanese	22	4	18	287	38	249
English	17	3	14	167	25	142
Simp. Chinese	16	2	14	271	19	252

Corpus Information

Language	Sentences			Opinion Clauses		
	Sum	Sample	Test	Sum	Sample	Test
Trad. Chinese	6174	1509	4655	N/A	N/A	4657
Japanese	7163	1278	5885	7569	1348	6221
English	4711	399	4312	4733	404	4329
Simp. Chinese	5301	242	4877	7523	570	6953

Opinion Percentage

Language	Opinionated		Relevant (of Opinionated)		Polarity (POS/NEG/NEU)	
	Lenient	Strict	Lenient	Strict	Lenient	Strict
T-Chinese	46.8	44.3	82.72	90.16	34.1 / 40.3 / 25.6	33.2 / 41.2 / 25.6
Japanese	28.9	21.1	43.2	22.6	5.5 / 15.3 / 79.2	4.3 / 10.2 / 85.5
English	25.2	7.5	99.4	95.7	25.0 / 48.0 / 6.0	18.0 / 46.4 / 0.9
S-Chinese	38.3	18.4	95.1	88.7	30.7 / 25.8 / 43.5	30.9 / 6.5 / 62.6

Lenient / Strict

Topic Information

ID	Title	Language ID			
		Japanese	English	Chinese	
				Traditional	Simplified
M00	Microsoft Anti-monopoly	N00			
M01	Regenerative medicine	N01	N01		N01
M02	American stance on depleted uranium bullets	N02	N02	N02	N02
M03	The impact of 911 terrorist attacks on America's economy	N03	N03	N03	N03
M04	HIV-tainted blood scandal	N04	N04		
M05	Cosovo civil war	N05	N05	N05	N05
M06	Incident with Nepal's ruling family (royalty)	N06	N06		N06
M07	Attacks toward Chinese Indonesian people	N07	N07	N07	N07
M08	Lawsuit American Government against Microsoft	N08	N08		N08
M09	Nuclear weapons tests	N09	N09	N09	N09
M10	Suriyah in the Middle East Peace Process.	N10	N10	N10	N10
M11	The relationship between AOL and Netscape	N11	N11	N11	N11
M12	El Nino	N12	N12	N12	N12
M13	The relationship between China and Russia	N13	N13		N13
M14	Greenhouse gasses	N14	N14	N14	N14
M15	The relationship between NATO and Poland	N15	N15	N15	N15
M16	Thailand in the Asian economic crisis	N16	N16	N16	N16
M17	Yasukuni Shrine	N17	T01		
M18	Chechin (Chechnia) civil war	N18		T96	
M19	Indonesian President Suharto	N19		N04	
M20	Nuclear missile abandonment of North Korea	N20		N13	
M21	Airplane crashes in Asia	N21		N08	
M22	The floods in the Mainland China			N01	
M23	The births of the cloned animals known to the world			N06	
M24	The responses of other countries to Lockerbie Air Disaster				N04

Annotator Agreement

	Japanese	Trad. Chinese	Simp. Chinese	English
Opinionated	0.7135	0.4581	0.4362	0.2369
Polarity	0.6341	0.7709	0.3634	0.1954
Relevance	0.5905	0.3329	0.6185	0.3459

Macro Average

Participation

Language		Japanese	English	Chinese	
				Trad.	Simp.
Total		8	9	7	9
Multi-lingual participants	J-E-TC-SC	1			
	J-E-TC	1			
	E-SC		1		1
	E-J	2			
	TC-SC			4	

Annotator Training

- JA: 5 annotator pool, 1 training topic, 6 hours of meetings
- EN: 6 annotator pool, 1 training topic, 6 hours of meetings
- ZH-TW: 10 annotator pool, 2 hours training, 1 training topic, 4th or 5th pass based on kappa
- ZH-CN: 12 annotator pool, 1 training topic, 8/4 hours of meetings (4 group leaders / annotator)

Some Guidelines

- General beliefs, “common sense knowledge” are not opinions
- Expressions of future plans are not opinions
- Generally used NTCIR-6 data for examples

Evaluation Metrics

- Precision, Recall, F-Measure over opinionated, relevant, polarity
- Semi-automatic evaluation of opinion holders and targets (precision, recall, f-measure)
- Multiple approaches used

Evaluation Metrics

- Lenient: 2/3 annotators must agree
- Strict: 3/3 annotators must agree
- Polarity ZH: Set precision + recall biased
- ZH: Rules to select polarity if annotators do not agree
- EN/JA: polarity needs majority agreement

$$\frac{\#system_correct(polar = POS, NEU, NEG)}{\#system_correct(opn = Y)} \cdot \frac{\#system_correct(polar = POS, NEU, NEG)}{\#system_proposed(opn = Y)}$$

Holder/Target evaluation

- Semi-automatic evaluation
- Match system extracted holders/targets to annotator holder list, automate the process in some way
- Time consuming, only first priority run evaluated
- No JA evaluation (only 1 run by organizer)

Simplified Chinese Lenient Results

Group	Opinionated			Relevance			Polarity S-P	Recall-based Polarity		
	P	R	F	P	R	F		P	R	F
BUPT	0.604	0.3991	0.4807	N/A			N/A			
ICLPKU	0.4803	0.8004	0.6003	0.9775	0.6559	0.785	0.4505	0.2164	0.3606	0.2705
NEUNLP	0.4721	0.7116	0.5676	N/A			N/A			
NLCL	0.4425	0.3991	0.4197	0.963	0.3258	0.4869	N/A			
NLPR	0.5822	0.7753	0.665	N/A			N/A			
NTU	0.5939	0.6089	0.6013	0.9656	0.7693	0.8564	0.4956	0.2944	0.3018	0.298
TTRD	0.412	0.9636	0.5772	0.9507	0.6981	0.8051	0.4348	0.1791	0.4189	0.251
WIA	0.5862	0.8208	0.6839	0.994	0.5032	0.6682	0.7419	0.4348	0.6089	0.5074
ISCAS	0.4649	0.7442	0.5723	0.9703	0.9288	0.9491	N/A			

Simplified Chinese Strict Results

Group	Opinionated			Relevance			Polarity S-P	Recall-based Polarity		
	P	R	F	P	R	F		P	R	F
BUPT	0.6312	0.4421	0.52	N/A			N/A			
ICLPKU	0.4486	0.8207	0.5801	0.9845	0.6743	0.8004	0.2836	0.1272	0.2327	0.1645
NEUNLP	0.4358	0.7339	0.5469	N/A			N/A			
NLCL	0.3857	0.402	0.3937	0.9736	0.3326	0.4959	N/A			
NLPR	0.6096	0.892	0.724	N/A			N/A			
NTU	0.6314	0.7517	0.6863	0.9748	0.7859	0.8702	0.3378	0.2133	0.2539	0.2318
TTRD	0.3481	0.9699	0.5124	0.9631	0.7006	0.8112	0.2882	0.1003	0.2795	0.1476
WIA **	0.6098	0.8964	0.7259	0.9969	0.524	0.687	0.5329	0.3250	0.4777	0.3868

Traditional Chinese Lenient Results

Group	Opinionated			Relevance			Polarity S-P	Recall-based Polarity		
	P	R	F	P	R	F		P	R	F
WIA	0.7298	0.5211	0.6080	0.9949	0.5306	0.6921	0.6931	0.5058	0.3611	0.4214
CityUHK	0.6601	0.8446	0.7411		N/A		0.5361	0.3539	0.4528	0.3973
iclpku	0.7015	0.6279	0.6626	0.9943	0.6768	0.8054	0.4810	0.3374	0.3020	0.3187
NLCL	0.5358	0.2676	0.3570	0.9240	0.1801	0.3015		N/A		
NTU	0.5648	0.8969	0.6931	0.9615	0.7103	0.8170	0.4875	0.2753	0.4372	0.3379
TTRD	0.5110	0.9345	0.6607	0.9673	0.8413	0.8999	0.3747	0.1915	0.3501	0.2476
UniNe	0.5428	0.9267	0.6846	0.9614	0.8456	0.8998	0.4293	0.2330	0.3978	0.2939

Traditional Chinese Strict Results

Group	Opinionated			Relevance			Polarity S-P	Recall-based Polarity		
	P	R	F	P	R	F		P	R	F
WIA	0.8520	0.6003	0.7043	0.9788	0.4061	0.5740	0.7003	0.5966	0.4204	0.4932
CityUHK	0.8364	0.9037	0.8687		N/A		0.5463	0.4569	0.4936	0.4746
iclpku	0.8567	0.6998	0.7704	0.9530	0.5626	0.7075	0.5085	0.4357	0.3559	0.3918
NLCL	0.6259	0.2930	0.3991	0.8487	0.1454	0.2482		N/A		
NTU	0.7076	0.9307	0.8040	0.8849	0.6437	0.7453	0.4979	0.3523	0.4634	0.4003
TTRD	0.6452	0.9395	0.7650	0.8992	0.8044	0.8491	0.3924	0.2531	0.3686	0.3002
UniNe	0.6921	0.9379	0.7965	0.8746	0.8443	0.8592	0.4431	0.3067	0.4156	0.3529

Japanese Lenient Results

Group	Opinionated			Relevance			Polarity		
	P	R	F	P	R	F	P	R	F
EHBN	0.4921	0.7313	0.5883	0.4819	0.6354	0.5481	N/A		
HCU	0.619	0.5138	0.5615	N/A			N/A		
MIRAC	0.316	0.0894	0.1394	0.4545	0.0816	0.1384	0.2465	0.0183	0.0341
NAK	0.8115	0.3416	0.4808	N/A			0.4922	0.1801	0.2637
NLCL	0.4255	0.2234	0.293	0.5367	0.1891	0.2797	N/A		
TAK	0.5191	0.2798	0.3636	N/A			0.4638	0.1138	0.1828
TUT	0.6742	0.562	0.613	0.5527	0.2925	0.3825	0.4596	0.214	0.292
UniNe	0.5363	0.1999	0.2912	0.4147	0.1918	0.2623	0.3251	0.0548	0.0938

Japanese Strict Results

Group	Opinionated			Relevance			Polarity		
	P	R	F	P	R	F	P	R	F
EHBN	0.3738	0.7627	0.5017	0.2808	0.7321	0.4059	N/A		
HCU	0.4894	0.5577	0.5213	N/A			N/A		
MIRAC	0.2412	0.0936	0.1349	0.2233	0.0821	0.1201	0.2394	0.0166	0.031
NAK	0.6885	0.3979	0.5043	N/A			0.4933	0.2154	0.2999
NLCL	0.3135	0.226	0.2627	0.301	0.2107	0.2479	N/A		
TAK	0.4166	0.3083	0.3544	N/A			0.5172	0.1316	0.2098
TUT	0.5416	0.6199	0.5781	0.3062	0.3357	0.3203	0.4806	0.2417	0.3216
UniNe	0.4164	0.2131	0.2819	0.1553	0.1464	0.1507	0.2914	0.0497	0.0849

English Lenient Results

Group	Opinionated			Relevance			Polarity		
	P	R	F	P	R	F	P	R	F
ICU	0.2435	0.3687	0.2933	0.2758	0.3648	0.3141	N/A		
kle	0.3529	0.7272	0.4752	N/A			0.2586	0.4301	0.3230
MIRACLE	0.5952	0.0116	0.0227	0.3741	0.3189	0.3444	N/A		
NEUNLP	0.3522	0.7788	0.4851	N/A			N/A		
NLCL	0.3780	0.1014	0.1599	0.1296	0.0685	0.0896	N/A		
sics	0.4192	0.6101	0.4970	N/A			0.1838	0.2413	0.2087
TUT	0.3185	0.4092	0.3582	0.2092	0.1755	0.1909	0.1943	0.1830	0.1885
UKP07	0.3305	0.9060	0.4844	N/A			0.2028	0.4394	0.2775
UniNe	0.3322	0.6995	0.4504	0.4170	0.5992	0.4918	0.2279	0.3671	0.2812

English Strict Results

Group	Opinionated			Relevance			Polarity		
	P	R	F	P	R	F	P	R	F
ICU	0.0743	0.3777	0.1241	0.0981	0.3797	0.1559	N/A		
kle	0.1109	0.7678	0.1938	N/A			0.0687	0.4645	0.1197
MIRACLE	0.2857	0.0116	0.0222	0.0853	0.3040	0.1333	N/A		
NEUNLP	0.1105	0.8204	0.1947	N/A			N/A		
NLCL	0.1168	0.1053	0.1107	0.0526	0.0847	0.0649	N/A		
sics	0.1336	0.6533	0.2219	N/A			0.0524	0.2796	0.0883
TUT	0.0961	0.4149	0.1561	0.0740	0.1853	0.1057	0.0569	0.2180	0.0903
UKP07	0.1026	0.9443	0.1850	N/A			0.0570	0.5024	0.1024
UniNe	0.1050	0.7430	0.1840	0.1614	0.6768	0.2607	0.0637	0.4171	0.1105

Cross-Lingual Results

Type	Opin	Relevance	Polarity
EN			
TC-SC-JA-EN	0.1599	0.0896	--
TC-JA-EN	0.4504	0.4918	0.2812
JA-EN	0.3582	0.1909	0.1885
EN	0.4752	(0.4918)	0.3230
TC			
TC-SC-JA-EN	0.3570	0.3015	--
TC-JA-EN	0.6846	0.8998	0.2939
TC-SC	0.6080	0.6921	0.4214
TC	0.7411	(0.9300)	0.3973
SC			
TC-SC-JA-EN	0.4197	0.4869	--
TC-SC	0.6839	0.6682	0.5074
SC	0.6839	0.6682	0.5074

English Holder & Target

Group	Type	Holder			Target		
		P	R	F	P	R	F
ICU	L	NA			0.1059	0.1761	0.1324
ICU	S	NA			0.0374	0.1793	0.0618
kle	L	0.4000	0.5076	0.4474	NA		
kle	S	0.1333	0.5322	0.2132	NA		
TUT	L	0.3923	0.2833	0.3290	NA		
TUT	S	0.1250	0.2829	0.1735	NA		

Simplified Chinese Holder & Target

Group	L/S	Holder-T			Target-T		
		P	R	F	P	R	F
ICLPKU	L	0.4124	0.4124	0.4124	0.0033	0.0033	0.0033
TTRD	L	0.1129	0.1129	0.1129	0.0151	0.0151	0.0151
NLPR	L	0.4286	0.4286	0.4286		N/A	
NTU	L	0.2909	0.2909	0.2909		N/A	
WIA	L	0.6656	0.6656	0.6656	0.4505	0.4505	0.4505
ICLPKU	S	0.4104	0.4104	0.4104	0	0	0
TTRD	S	0.1719	0.1719	0.1719	0	0	0
NLPR	S	0.4759	0.4759	0.4759		N/A	
NTU	S	0.3839	0.3839	0.3839		N/A	
WIA	S	0.7574	0.7574	0.7574	0.5185	0.5185	0.5185

Traditional Chinese Holder & Target

Group	L/S	Holder-T			Target-T		
		P	R	F	P	R	F
WIA	L	0.8254	0.8254	0.8254	0.6058	0.6058	0.6058
iclpku	L	0.5872	0.5872	0.5872	0.0213	0.0213	0.0213
NTU	L	0.5028	0.5028	0.5028	N/A		
TTRD	L	0.5645	0.5645	0.5645	0.0358	0.0358	0.0358
WIA	S	0.8238	0.8238	0.8238	0.6331	0.6331	0.6331
iclpku	S	0.5797	0.5797	0.5797	0.0228	0.0228	0.0228
NTU	S	0.4825	0.4825	0.4825	N/A		
TTRD	S	0.5496	0.5496	0.5496	0.0372	0.0372	0.0372

Discussion

- Relevance < Opinionated < Polarity < Holder
- CH, EN, JA corpora have different annotator agreement: training issue or data issue?
- How to evaluate when annotators do not agree?
- What do the results MEAN?

Difficulties

- Annotator agreement (address with better experiment design?)
- Shortened schedule (move to News docs)
- Meaningful Cross-language comparison?
- Lack of shared tools for evaluation / processing / etc.
- Relevance this year is suspicious

Future Work

- Increase group participation in multiple languages
- Improve annotator agreement (EN)
- Standardization of tools (evaluation and annotation)
- Confidence / statistical significance for evaluation metrics

NTCIR-8

- Breakout session today from 16:10
- Looking for feedback!
- Plans to incorporate a multi-lingual task:
 - Query in EN, summarize opinion types for JA, EN, ZH-TW, ZH-CN
- New data