

1. Introduction

In this paper, we describe our participating system, which is based on supervised approaches and dependency parsing, for opinion analysis on traditional Chinese texts at NTCIR-8:

- 1) For opinionated sentence recognition, the supervised lexicon-based approach, SVM and Maximum Entropy are combined together.
- 2) For polarity classification, we use only the supervised lexicon-based approach.
- 3) on the basis of dependency parsing,
 - a) identify opinion holders by means of reporting verbs and
 - b) identify opinion targets by considering both opinion holders and opinion-bearing words.

The results show that among all the teams participating in the traditional Chinese task, our system achieve:

- 1)the highest F-measure on the opinionated sentence recognition task,
- 2)the second highest F-measure on the identification of both opinion holders and targets,
- 3)the middle ranking for opinion polarity classification.

2. Linguistic Analysis of Opinions

2.1 Subjectivity and Polarity

- Reporting verbs
- Sentiment-bearing items
- Adverb clues
- Negation marker
- Discourse marker

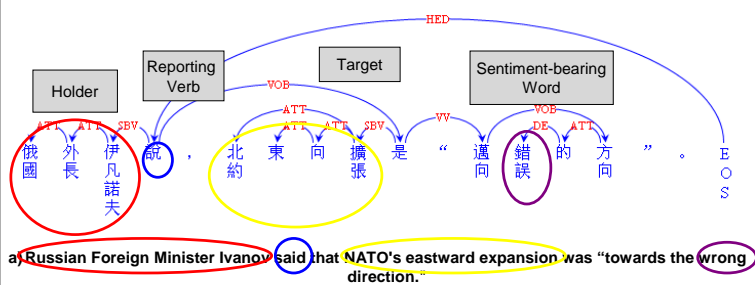
2.2 Opinion Holder & Target

Opinion holders/targets are more diverse in *news texts* than in product reviews:

- Holders could be any named entities and noun phrases;
- Targets are more abstract, could be noun phrases, verb phrases or even clauses.

2.3 Dependency Parsing and Opinion Holders / Targets

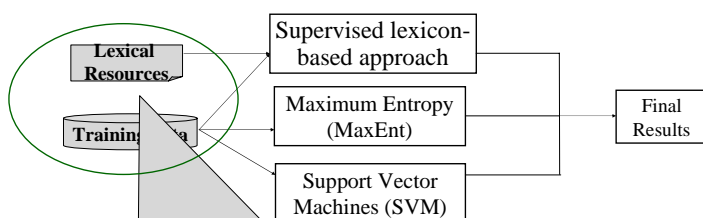
- a) 俄國外長 伊凡諾夫 說，北約東向擴張是“邁向 錯誤的 方向”。



3. Subjectivity and Polarity Classification

Motivation: make full use of

- 1) the manual labeled lexicons
- 2) annotated corpora



1. sample and test data for NTCIR-6 & NTCIR-7 (traditional Chinese)
2. sample data for NTCIR-8 MOAT (traditional Chinese)

4. Identifying Opinion Holders/Targets with Dependency Parsing

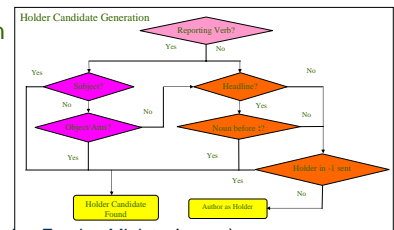
4.1 Sentence Preprocessing for better parsing (SP)

- Named entities were first recognized with a large dictionary
- Parentheses enclosing only English words/numbers are removed from sentences

4.2 Opinion Holder Identification

a) Holder Candidate Generation

- Subject of Reporting Verb
- Heuristic Rules (HR)



b) Candidate Expansion (EP)

- Attributive modifier:
 - e.g. 俄國外長 伊凡諾夫 (Russian Foreign Minister Ivanov)
- Quantifier modifier and 和/及 (and/or)
 - e.g. 蘇哈托和另外兩名軍方將領 (Suharto **and two** other army generals)

4.3 Opinion Target Identification with Opinion Holder and Opinion-bearing Words

a) Target Candidate Generation (Heuristic Rules, HR)

- Subject in the embedded clause if holder is identified by a reporting verb
 - the subject of the object (verb) of the reporting verb or find (after the reporting verb) the subject whose parent is an opinion-bearing word
- Subject/object of the whole sentence if no holder is found
- Remove a target candidate if it is in the holder candidates (called holder conflict, **HC**)

b) Target Candidate Expansion (EP)

5. Results

Group ID	Run	Supervised Lexicon-based method			Combination of SVM, MaxEnt, Supervised Lexicon-based method		
		P	R	F	P	R	F
CTL	1	65.14	68.79	66.92	76.5	53.06	62.66
CityUHK	2	56.39	85.71	68.03	44.14	38.5	41.13
CityUHK	1	50.92	91.98	65.55	45.17	41.93	43.49
CityUHK	3	50.92	91.98	65.55	45.17	41.93	43.49
WIA	1	53.41	83.68	65.2	50.68	41.14	45.41
WIA	2	53.41	83.68	65.2	50.66	40.45	44.98
KLELAB	3	44.51	87.92	59.1			
KLELAB	1	41.98	94.94	58.22			
KLELAB	2	41.98	94.94	58.22			
NTU	2	41.85	92.22	57.57	44.35	41.19	42.71
NTU	1	41.41	93.82	57.46	45.57	42.83	44.16
cyut	1	42.71	87.74	57.45	40.49	35.6	37.89
cyut	2	41.13	82.41	54.87	31.26	25.95	28.36
UNINE	1	52.37	48.47	50.34	47.01	23.27	31.13
cyut	3	47.55	43.99	45.7	36.68	16.19	22.46

Group ID	Run	Holder	Target
CTL	1	84.9	54.4
CityUHK	2	72.1	48.5
CityUHK	1	70	25.9
CityUHK	3	68.1	23.3
WIA	1	62.1	28.3
WIA	2	60.5	24.6
KLELAB	1	29.6	
KLELAB	2	26.2	

6. Conclusion

The system ranked on the traditional Chinese task

- No. 1 for opinionated sentence recognition,
- No. 2 for identification of both opinion holders and targets,
- the middle position for polarity classification.

The result show that

- 1)the combination of supervised lexicon-based approach and machine learning techniques (namely, SVM and Maximum Entropy) is effective for opinionated sentence recognition;
- 2)the dependency parsing-based approach on opinion holder and target identification is effective.