Supervised Approaches and Dependency Parsing for Chinese Opinion Analysis at NTCIR-8

Bin LU, Benjamin K. TSOU and Tao JIANG

Language Information Sciences Research Centre
City University of Hong Kong

香港城市大學
City University of Hong Kong

Outline

- Introduction
- Linguistic analysis of opinions
- Supervised approaches for subjectivity/ polarity classification
- Opinion holder/target identification with dependency parsing
- Official Results
- Conclusion

Introduction

- The NTCIR-8 Multilingual Opinion Analysis Task (MOAT)

- the CityU (HK)'s system
  - the traditional Chinese task
  - four of the five subtasks:
    - opinionated sentence recognition
    - opinion polarity classification
    - opinion holder identification
    - opinion target identification
  - Three runs were submitted

香港城市大學
City University of Hong Kong

Subjectivity & Polarity Classification

Linguistic analysis of opinions

- Features for opinionated sentence recognition

Reporting verbs: verbs indicating speech events

- *a) 報導中引述KGB在德國的上司卡魯金的話 **說**，普亭的間諜工作並 **不特別成功**。(The report **said** the spying work of Putin was not **quite successful**...)*

Polar items: sentiment-bearing items (words or phrases)

- *b) 金融市場 **可能** **面臨** 不能提供所需資金 **風險**。(The financial market was **perhaps** facing **the danger** of not being able to provide ...essary funds.)*

Adverb clues: adverbs frequently co-occurring with opinions.

Linguistic analysis of opinions (cont'd)

- Features for polarity classification
  - *a) 報導中引述KGB在德國的上司卡魯金的話說，普亭的間諜工作並 不 特別成功。(The report said the spying work of Putin was **not quite successful**...)*

above mentioned **polar items**

Negation markers: words used to reverse the polarity of a polar item.

  - *c) 普亭 **雖然 支持**這項法案，**但** 俄國民意對這項法案的**反對聲**浪高漲。((Although) Putin **supports** this bill, **but** the majority of Russian people is **highly against** it.)*

Discourse markers: those may reverse the polarity of previous clause.

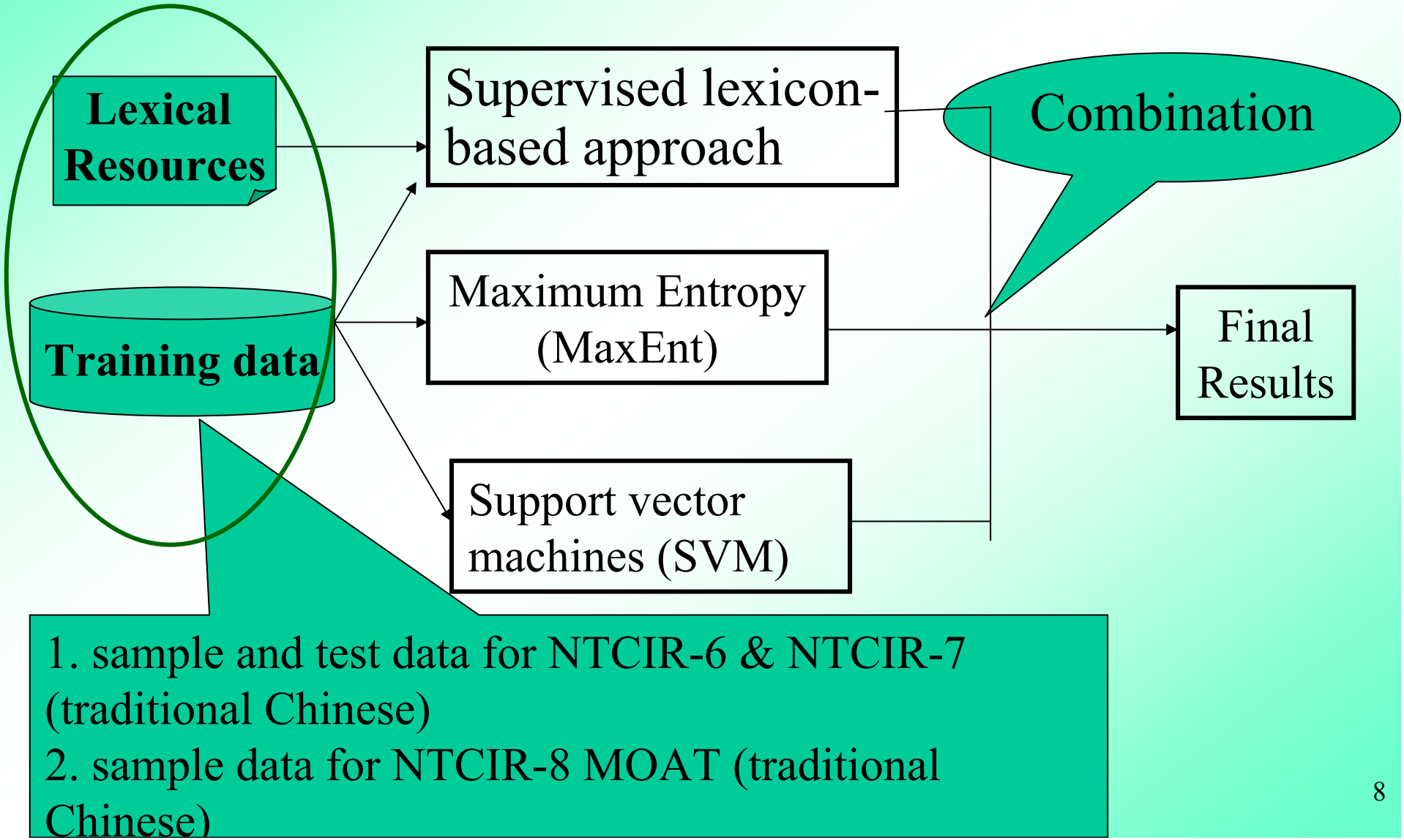
香港城市大學
City University of Hong Kong

Supervised approaches and ensemble techniques

- Motivation
  - make full use
    - the manual labeled lexicons
    - annotated corpora
  - the training corpus
    - the sample and test data for NTCIR-6 OAPT (traditional Chinese)
    - the sample data for NTCIR-7 MOAT (traditional Chinese)

香港城市大學
City University of Hong Kong

System Architecture



Lexical resources

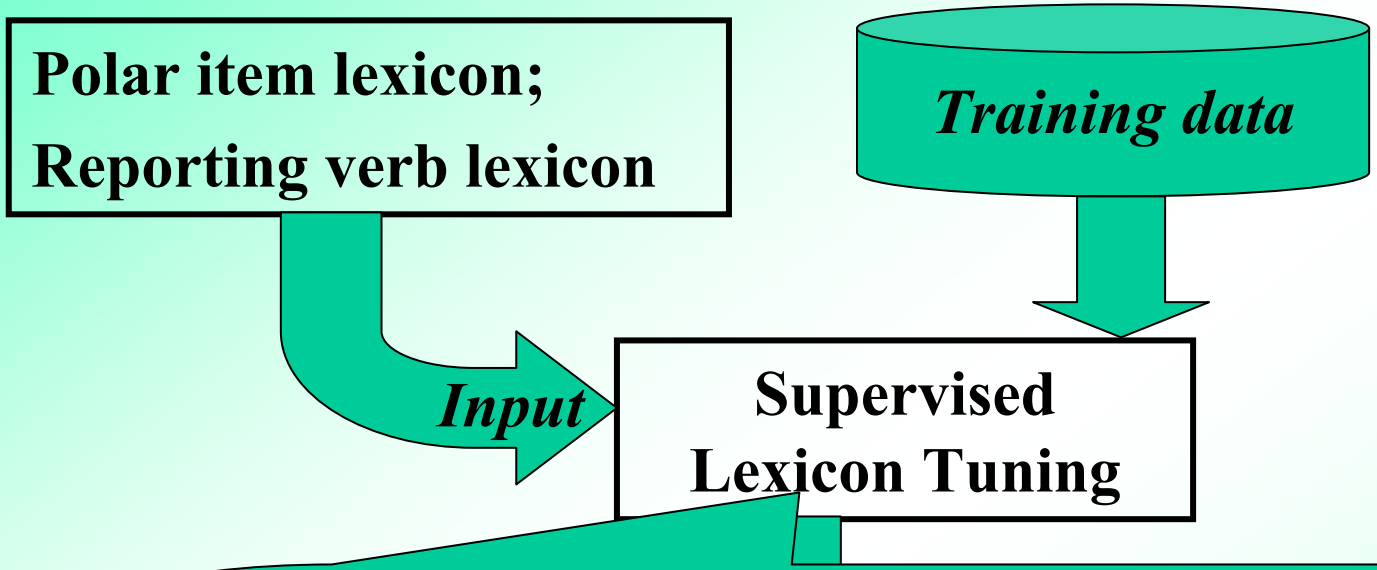
- Lists of Polar item
  - NTU Sentiment Dictionary (NTUSD)
  - *The Lexicon of Chinese Positive Words (*LCPW*)*
  - *The Lexicon of Chinese Negative Words* (LCNW)
  - CityU's polar word and phrase list (CPWP)
  - Polar items from sample data of NTCIR-6 OAPT (SIST)
    - marked with the *SENTIMENT_KW* tag

	NTUSD	LCPW	LCNW	CPWP	SWST	**Combined**
# Positive items	2812	5046	0	5838	2426	**13,437**
# Negative items	8276	0	3499	9002	1252	**18,365**
Total	11088	5046	3499	14840	4234	**31,802**

the Lexicon-based method

- Identify opinionated sentences
  - check whether a polar item (including adverbs) or a reporting verb occur in it.
  - If yes, then opinionated,
  - Otherwise, not opinionated.

Lexicon adjustment process



Two reasons for this adjustment on a large sentiment lexicon

a) **may contain errors or typos:** especially those polar items extracted from last year's sample data are not quite clean, such as *随着 (with)*, *可以 (be able to)*, etc.

b) **words could be contextual or not suitable for news domain**: since they are marked with annotators' own subjectivities;

Supervised Lexicon Tuning

compute precision for each reporting verb / polar item (on the training data)

learn the best threshold combination (threshold for reporting verbs + threshold for polar items)

1. Tune **separately** for two subtasks (i.e. opinionated and polarity)
2. For polarity classification, only filter polar item lexicon, and reporting verbs were not used.

Lexicon adjustment

- Two kinds of items filtered out:
  - 1) noisy terms: actually not reporting verbs or polar items according to our judgment, e.g.
    - 觀光 (sightseeing) in LCPW, 定下 (set) and 前往 (head for) in SKPI;
  - 2) reporting verbs or polar items: may present facts and frequently occur in factual sentences,
    - e.g. 暴雨 (downpour) in NTUSD and 襲擊 (attack) in NTUSD and CPWP.

香港城市大學
City University of Hong Kong

Combination method

- Combination method for the *opinionated sentence recognition* task:
  - majority voting
  - if two of the three component classifiers label a sentence as opinionated, the sentence would be marked as opinionated;

Official Results

Group ID	Run	Opinionated			Polarity		
		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

Official Results

Combination of SVM, MaxEnt, Supervised Lexicon-based method

Run	Opinionated			Polarity		
	P	R	F	P	R	F
2	56.39	85.71	68.03	44.14	38.5	41.13
1/3	50.92	91.98	65.55	45.17	41.93	43.49

Supervised Lexicon-based method

Supervised Lexicon-based method

香港城市大學
City University of Hong Kong

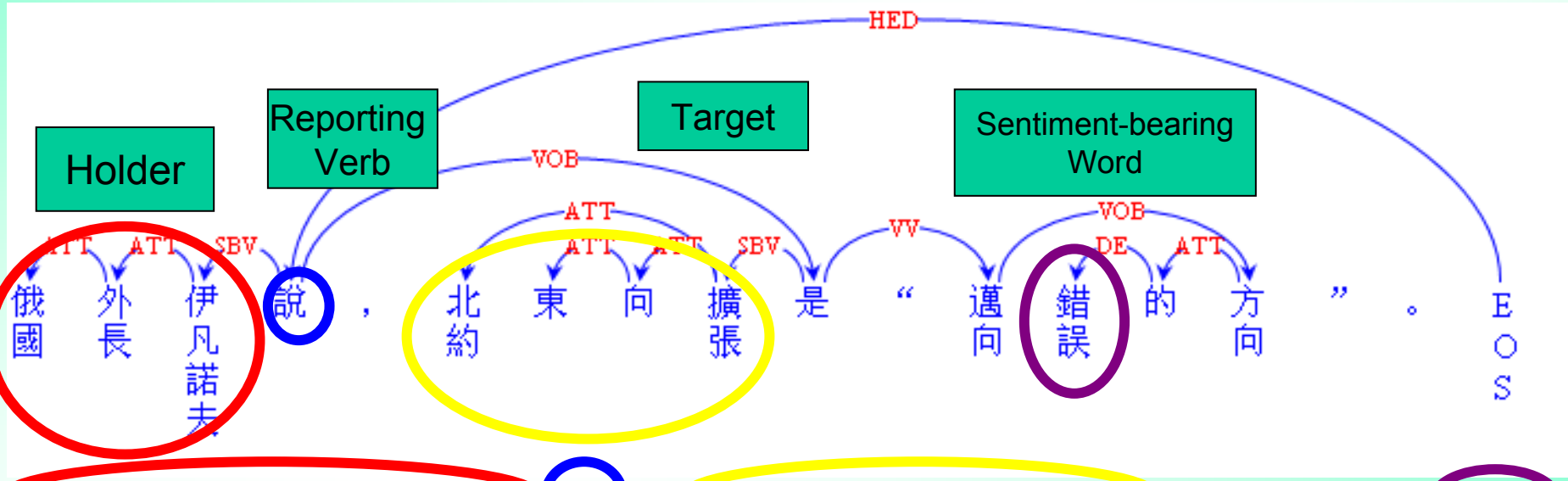
Holder/Target Identification

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**

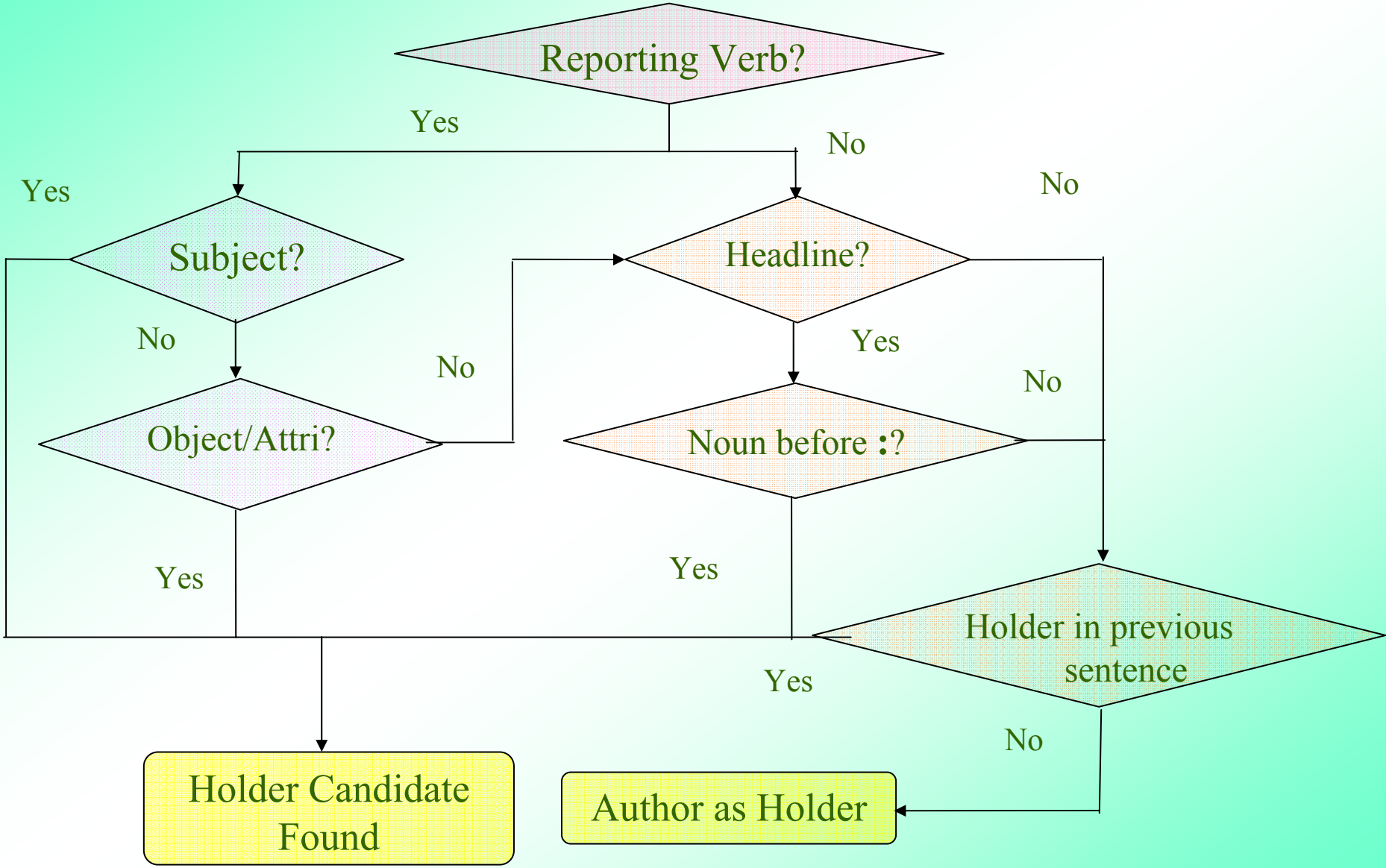
Dependency Parsing and Opinion Holders / Targets

a) 俄國 外長　伊凡諾夫 說，北約東向擴張是“邁向　錯誤 的 方向　”　。



Russian Foreign Minister Ivanov said that NATO's eastward expansion was "towards the wrong direction."

Holder Candidate Generation



Opinion Holder Identification

- Holder Candidate Generation
  - Subject of Reporting Verb
  - Heuristic Rules (**HR**)

- 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)

Opinion Target Identification with Opinion Holder and Opinion-bearing Words

- 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**)
- Target Candidate Expansion (**EP**)
  - similar to holder candidate expansion

Official Results

the method described above

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	6
KLELAB	1		
KLELAB	2	26.2	

add *more heuristic rules obtained on NTCIR-7 data*

Official Results

Group ID			Target
CTL			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	

a big difference:
annotators have significantly different opinions on opinion analysis of news sentences.

香港城市大學
City University of Hong Kong

Conclusion

- The result show that
  - the combination of supervised lexicon-based approach and machine learning techniques (namely, SVM and Maximum Entropy) is effective for opinionated sentence recognition;
    - No. 1 for opinionated sentence recognition,
  - the dependency parsing-based approach on opinion holder and target identification is effective.
    - No. 2 for identification of both opinion holders and targets,

香港城市大學
City University of Hong Kong

Conclusion on Subjectivity Classification

- Large sentiment lexicons needs some adaptation on the new domain

- the combination lexicon and machine learning can improve the performance on opinionated sentence recognition

Conclusion on Holder/Target Extraction

- **Dependency parsing-based approach on opinion holder and target identification is effective.**

- The existence of reporting verbs is a very important feature for identifying opinion holders in news texts;

- The identification of opinion targets should not be done alone without considering opinion holders in news
  - opinion holders are much easier to be identified in news texts
  - the identified holders are quite useful for the identification of the associated targets.

香港城市大學
City University of Hong Kong

Future Work

- Polarity classification
  - contextual information
  - topic-related features
  - shallow parsing techniques

- Identifying opinion holder/target based only on dependency parsing
  - not robust to the dependency errors
  - to further investigate machine learning approaches by treating dependency structures as features
    - should be more robust to dependency errors

香港城市大學
City University of Hong Kong

Thanks!

Q & A