

# Detecting Opinions and their Opinion Targets in NTCIR-8

Yoonjung Choi, Seongchan Kim, Sung-Hyon Myaeng

Korea Advanced Institute of Science and Technology (KAIST), South Korea

## Task Scope

### Subjectivity Judgment

- Deciding whether a sentence conveys an opinion

ex) Bush **said** we need to think about Saddam Hussein using al-Qaida to do his dirty work to not leave fingerprints behind. → **Opinionated sentence**

### Opinion Target Identification

- Resolving opinion targets within opinionated sentence

ex) A U.S. official said he was pleased that **Gen.I Made Mangku Pastika** was doing **painstaking** work.

## Subjectivity Judgment

### Hypothesis

- Opinionated sentences contain opinion clue words (e.g., happy, criticize)

### Lexical Patterns

Main verb	"insist", "claim", "criticize", "think", "believe", "advise"
Auxiliary verb	"would", "could", "should", "might", "will", "may"
Special phrase	"in fact", "unfortunately", "consequently"

### Subjectivity Scoring

- Subjective Score (SentiWordNet)

$$SubjScr(t) = |PosScr(t) - NegScr(t)|$$

- Appraisal Score (Appraisal Verbs)

$$AppScr(t) = 0.3 \text{ (empirically set)}$$

- Opinion Score

$$OpiScr(t) = SubjScr(t) + AppScr(t)$$

$$OpiScr(Sent) = \sum_{t \in Sent} OpiScr(t)$$

## Experiments

- 20 topics (138 docs / 6,165 sentences)

Task	Precision	Recall	F-measure
Opinion Extraction	0.1888	0.6526	0.2943
Opinion Target Identification	0.231	0.346	0.277

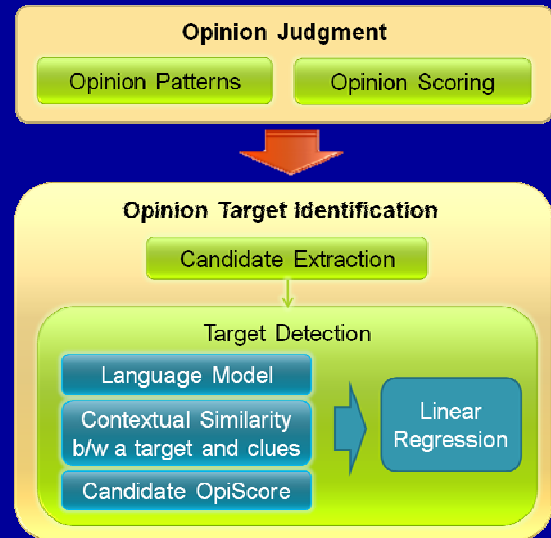
### Subjectivity Judgment Task

- Many sentences include opinion words in general

### Opinion Target Identification Task

- Strict Matching was used → boundary problem
- Anaphoric target problem
- Mostly depends on TF and the presence of opinion clues

## System Flow



## Opinion Target Identification

### Hypothesis

- Document-level theme can be an opinion target
- Opinion target will frequently co-occur with opinion clue words

### Candidates: Noun Phrases

### Statistical Classifier: Linear Regression

- Title Section: Whether given candidate's keywords appears on the *Title* section

- Document-level Language Model

$$p(NP | D) = \prod_{i=1}^n p(t_i | D)$$

$$p(t_i | d_j) = \frac{\text{count}(t_i, d_j)}{\sum_{t \in d_j} \text{count}(t, d_j)}$$

- Candidate's Opinion Score

- Some opinion targets include strong opinion clue words such as "threat" and "terrorism"

- Collocation Information between a candidate and opinion clues

$$CI(NP, Clues) = \frac{|Sent_{NP} \cap Sent_{Clues}|}{\sqrt{|Sent_{NP}|} \times \sqrt{|Sent_{Clues}|}}$$

## Future Works

- Filtering weak opinion clue words
- Opinion target boundary detection
- Anaphor and synonym problem