RECOGNIZING THE ELLIPSIS OF OPINION TARGET IN CHINESE TEXT
中文文本中评价对象省略识别方法

Zhu Zhu Min Dai Shoushan Li Guodong Zhou

Natural Language Processing Lab

Soochow University

OUTLINE

- **Background**

- **Related Work**

- **Recognizing the Ellipsis of Opinion Target in Chinese Text**
  - Framework
  - Features
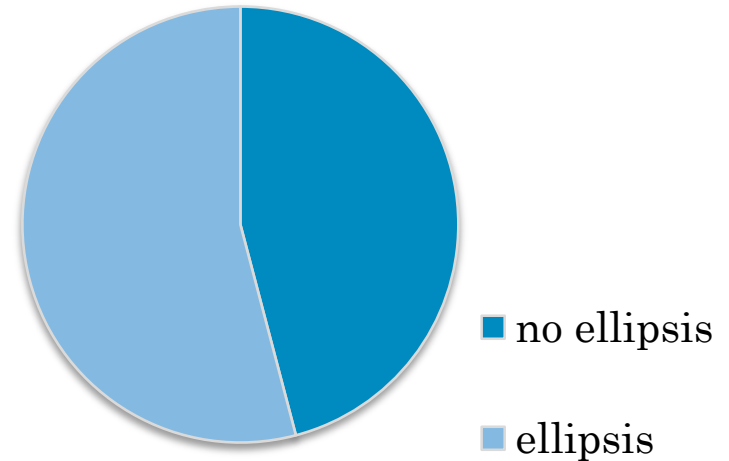 - Experiments

- **Conclusion**

BACKGROUND

- **Opinion target extraction** is a basic task in sentiment analysis.

- **Opinion target extraction** is to extract the object faced by opinion expressions.

✓ In product reviews:

    ✓ the product itself: *I like the computer.*

    ✓ product attributes: *The design of iphone6 is amazing.*

BACKGROUND (CONT.)

- Ellipsis phenomenon is particularly common and acute in Chinese text.

- Opinion target is often omitted as an important element of sentiment information.

■ no ellipsis

■ ellipsis

✓ For example:
- ✓ 这次 我们 住 在 了 火车站 附近 。
- ✓ 很 不错 ， 很 安静 。

酒店 *(the hotel)*

OUTLINE

- **Background**

- **Related Work**

- **Recognizing the Ellipsis of Opinion Target in Chinese Text**
  - Framework
  - Features
  - Experiments

- **Conclusion**

RELATED WORK

- Opinion Target Extraction
  - Unsupervised-learning
    - Hu and Liu, *SIGKDD-2004;*
    - Popescu and Etzioni, *EMNLP-2005;*
    - Scaffidi etc., *EC-2007*.

  - Supervised-learning
    - Zhuang etc., *CIKM-2006;*
    - Kessler and Nicolov, *AAAI-2009;*
    - Jakob etc., *EMNLP-2010;*
    - Li etc., *AAAI-2012*.

RELATED WORK（CONT.）

- Ellipsis in Chinese
  - Characteristics:
    - Flexibility
    - Uncertainty

  - There is no concrete study related to the ellipsis of opinion target in Chinese sentiment texts.

  - Approaches (zero anaphora):
    - Rule-based
      - Yeh etc., *Computational Linguistic-1997;*
      - *Nielsen, COLING-2004.*
    - Machine-learning
      - Zhao etc., *EMNLP-2007;*
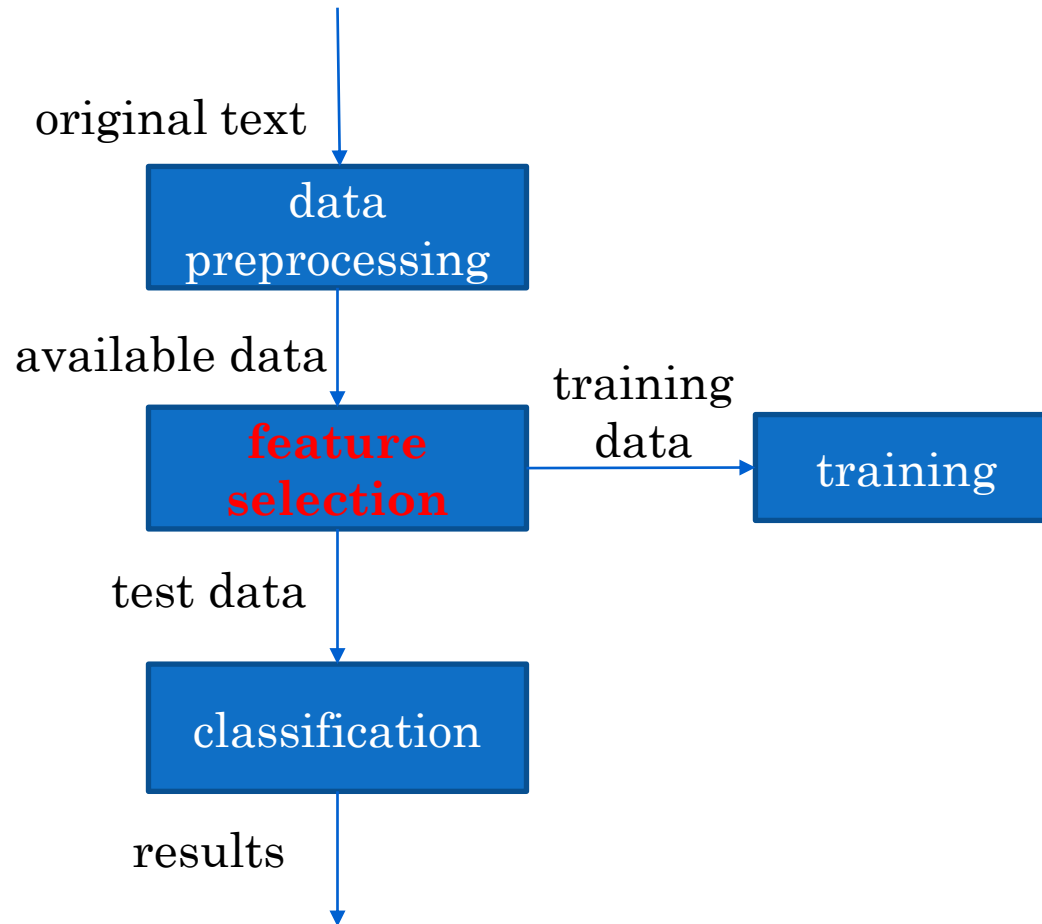   - *Huang etc., Computer Science-2012.*

OUTLINE

- **Background**

- **Related Work**

- **Recognizing the Ellipsis of Opinion Target in Chinese Text**
  - Framework
  - Features
  - Experiments

- **Conclusion**

FRAMEWORK

original text

data
preprocessing

available data

feature
selection

training
data

training

test data

classification

results

FEATURES

- 这台笔记本虽然好看，但是不实用。

  - *Clause1*: 这台笔记本虽然好看，
  - *Clause2*: 但是(这台笔记本)不实用。

  - Opinion target-----这台笔记本:
    - at the beginning of the sentence
    - noun
    - appeared in clause 1, while omitted in clause 2

FEATURES (CONT.)

- sentence's position-dependent features

Features	Description
f_iw	the first i words at the beginning of the sentence
f_ip	the POS of the first i words at the beginning of the sentence
f_iw_p	the first i words at the beginning of the sentence with the POS
l_iw	the last i words at the end of the sentence
l_ip	the POS of the last i words at the end of the sentence
l_iw_p	the last i words at the end of the sentence with the POS

FEATURES (CONT.)

○ sentence's position-independent features

Features	Description
word	all words in the sentence (bagword)
pos	the POS of all words in the sentence
w_p	all words in the sentence with the POS

FEATURES (CONT.)

- contextual features

Features	Description
PreW	all words in the previous clause
PreP	the POS of all words in the previous clause
PreW_P	all words in the previous sentence with the POS

EXPERIMENTS

- Corpus—from Amazon

Domain	Docs.	Sen.	Sen. (with the ellipsis of opinion target)
Notebook	2000	4649	1082 (23%)
Hotel	1000	4368	614 (14%)
Beauty	2000	3291	1060 (32%)

EXPERIMENTS (CONT.)

- Single feature
- Greedy feature selection

Greedy Feature Selection

○ *Notebook*

Features	Performance (F1%)
w_p	71.9
+$word$	75.8
+f_2p	76.6
+f_1w	77.1
+pos	77.5
+f_3p	77.7
+l_1p	**78.3**

GREEDY FEATURE SELECTION (CONT.)

- *Hotel*

Features	Performance (F1%)
word	72.4
+*w_p*	75.4
+*pos*	76.9
+f_2p	77.7
+f_1p	78.2
+f_1w_p	78.7
+f_1w	79.0
+*PreW_P*	**80.1**

GREEDY FEATURE SELECTION (CONT.)

- *Beauty*

Features	Performance (F1%)
word	76.3
+*w_p*	76.4
+f_1w	76.6
+f_2p	77.2
+*pos*	77.5
+f_1p	77.7
+l_3p	77.9
+l_1w_p	78.3
+*PreW*	**78.7**

FEATURE CONCLUSION

- sentence's position-dependent features
  - $f_i p$

- sentence's position-independent features
  - *word*, *pos*, *w_p*

- contextual features
  - combination > single

OUTLINE

- **Background**

- **Related Work**

- **Recognizing the Ellipsis of Opinion Target in Chinese Text**
  - Framework
  - Features
  - Experiments

- **Conclusion**

Conclusion and Future Work

- More features

- Applied to the task of opinion target extraction

- Combined with opinion expressions

- ......

THANK YOU