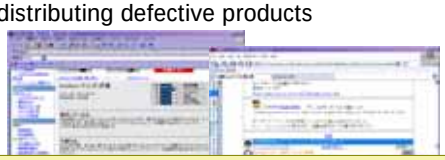
Opinion Mining as Extraction of Attribute-Value Pairs

Nozomi Kobayashi
Computational linguistics lab.
Nara Institute of Science and Technology

Background(1/2)

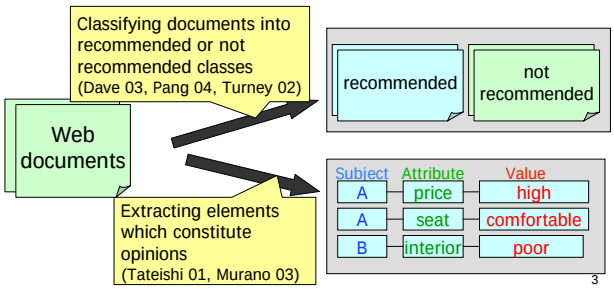
- Acquiring users' opinions on products or services from Web documents
  - It is effective in avoiding the risk of widely distributing defective products



This situation attracts an increasing interest in an automatic text analysis of opinions

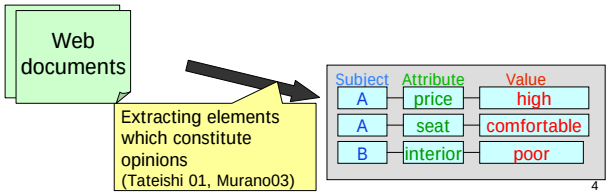
Background(2/2)

- The task of opinion acquisition:



Aim

- To extract opinions that describe evaluation of particular products together with evidence
  - Information extraction viewpoint



Task definition (1/2)

- Opinion:
  - Writer's subjective evaluation of a particular product or a certain aspect of the product

opinion

....the engine is very powerful.

not opinion

I wish they would improve the seats. (request)
If only the engine were more powerful ...(counter-factual)
I have heard that the engine is powerful. (hearsay evidence)

Task definition (2/2)

- Opinions may be linguistically expressed in many ways
  - Typical expressions: attribute-value pair

Attribute: one aspect of a product (subject)
(e.g. engine, design, price,…)
Value: specific expression that quantifies or qualifies the aspect
(e.g. high, good, beautiful, ...)

The task: Extraction of ⟨subject, attribute, value⟩

Extraction of attribute-value pairs

The Prius is quiet and the seats are very comfortable.

⟨Prius, **ϕ**, quiet⟩
⟨Prius, seats, comfortable⟩

Task definition (in this talk)

- To extract Attribute-Value pairs
  - On review sites, products are often clearly specified

BMW 3-series
Reviewed by: Joe Blogs
Summary:
I was so excited when I bought this car!!
Loved it IMMEDIATELY until...about a month
in I started having all sorts of problems,
including replacing the thermostat 3 times in the last 5 years.

Related work

- Pattern-based approach (Tateishi 01,Murano 03)
  - Using pre-defined extraction patterns and a list of evaluative expressions
    - Need to be manually created

- Manual construction of rules is costly
- Difficult to achieve satisfactory coverage

- Semantic parsing-based approach (Kanayama 04)
  - Applying transfer-based MT techniques

- Opinion expressions appear with anaphoric expressions (pronouns or ellipses in many cases)

Our stance

- Focus on the similarity between the anaphora resolution task and the opinion extraction task
  - Attempt to apply the method used for anaphora resolution to opinion extraction

anaphora resolution:
The task to identify the antecedent of a pronoun or an ellipsis

Outline

- Background and aim
- Related work
- Method for opinion extraction
- Experiments and evaluation
- Conclusion

Opinion extraction

- Two subtasks:
  1. Extraction of attribute-value pairs about a product
  2. Determination of its opinion-hood
     - To classify the value expression either as an opinion or not

Extraction of attribute-value pairs

- Problems:
  - Attribute-value pair has a direct dependency relation within a text

デザインは変です.
(The design is weird.)

  - Attribute-value pair does not have a direct dependency relation

Important problem

ellipsis

デザインは変ですが, 私は(φガ)好きです.
(The design is weird, but I like it.)

Anaphora resolution task

- Identify the antecedent of a pronoun or an ellipsis

anaphor

antecedent

太郎は人数を (φ1ガ) 調査して報告書を作成した.
(Taro noted the attendance and wrote a report.)

  - Current work on anaphora resolution
    - Machine learning-based approaches are having achieved a better performance

We consider that the opinion extraction task may be processed in a quite similar way

Extraction of attribute-value pairs

- A value appears in the text but its attribute is missing
  - It is inferable from the value expression and the context

予想よりもかなり(φガ)速いです！
(It is faster than I expected)

- Decide whether the candidate attribute stands for the real attribute or not

We apply supervised machine-learning

Opinion extraction

- Two subtasks:
  1. Extraction of attribute-value pairs about a product
  2. Determination of its opinion-hood
     - To classify the value expression either as an opinion or not

Determination of its opinion-hood

- To determine whether the value expression expresses an opinion or not
  - Binary classification task

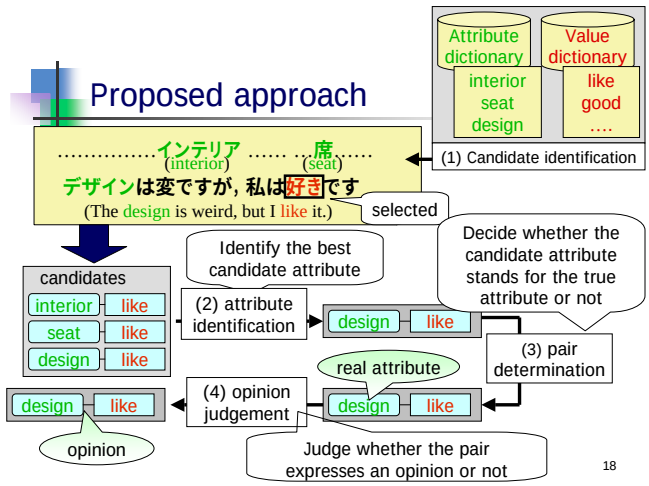
→ we apply the supervised machine-learning technique

Likelihood of the value expression to be an opinion changes according to its attribute

Knowing the candidate attribute is useful to determine the opinion-hood of a value

(Many people customize their car)

Proposed approach



Outline

- Background and aim
- Related work
- Method for opinion extraction
- Experiments and evaluation
- Conclusion

Aim of the experiments

- To evaluate the performance of our model
  - Applying the machine learning-based method designed for anaphora resolution to opinion extraction is useful
  - Information about the candidate attribute is useful for opinion judgement

Settings

- Opinion-tagged corpus
  - 287 review articles in the automobile domain

Attribute-value pairs	2409
Values (no corresponding attribute)	424
Total number	2833

- Dictionaries constructed by the semi-automatic method
  - Attribute dictionary: 3,777 expressions
    - エンジン(engine), 内装(interior), エクステリア(exterior). . .
  - Value dictionary: 3,962 expressions
    - 高い(high), 速い(fast), 良い(good), 好き(like). . .

Result (pair extraction)

デザインは変です.
(The design is weird.)

デザインは変ですが私は好きです.
(The design is weird, but I like it.)

The attribute-value pairs are directly connected with a dependency relation in the sentence

procedure	precision	recall
dependency	69.4% (1443/2803)	59.9% (1443/2409)
attribute identification → pair determination	74.0% (2029/2741)	84.2% (2029/2409)

Applying the machine learning-based method designed for anaphora resolution is useful

Total # of Attribute-Value pairs

Result (opinion judgement)

インテリアに不満が多い
(I have many gripes about the interior)

車を改造する人が多い
(Many people customize their car)

procedure	precision	recall
without attribute information	75.3% (1739/2310)	61.4% (1739/2833)
Use attribute information	79.5% (1933/2430)	68.2% (1933/2833)

Correctly extracted attribute-value opinions

The information about the candidate attribute is useful for opinion judgement

Result (opinion extraction)

- Attribute identification → pair determination → opinion judgement
  - Precision 69.5% (1357/2833)
  - Recall 47.9% (1357/2833)

We will need further investigation for improvement of the performance

Conclusion and future work

- We have proposed a machine learning-based extraction of attribute-value pairs
- Future work
  - To investigate for improvement of the performance