Extracting aspect-aspect pairs for opinion extraction

Nozomi Kobayashi
Computational Linguistics Lab.
Nara Institute of Science and Technology

1

Background

- There are an increasing number of Web documents that include human opinions
  - This information is scattered all over the web
  - Acquiring users' opinions on products or services is helpful

This situation is generating increasing interest in automatic text analysis of opinions

2

Aim

- Extracting opinions that describe how customers feel about certain aspects of certain products, shops, services, etc.
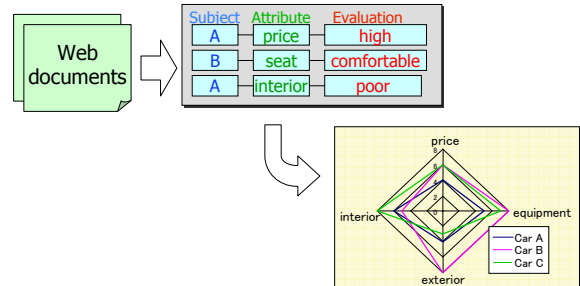 - Extracting structured "opinions"

positive opinion for Mazda XXX

I bought a new Mazda XXX this summer. I liked it because the design was cool.

Subject	Attribute	Evaluation
Mazda XXX	design	cool
Mazda XXX		liked

3

Example application

- Generation of radar charts

Web documents

Subject	Attribute	Evaluation
A	price	high
B	seat	comfortable
A	interior	poor

4

Issues

- How the task of opinion extraction and structurization should be designed

- How to extract the opinions which we defined

5

Outline

- Background and aim
- Structuring opinions
- Approach for opinion extraction
- Experiment and result
- Conclusion and future work

6

Opinion structurization

- Opinion：describes the opinion holders' positive or negative feelings for certain product (or a shop, a service, etc.)

text
> I bought powershot last week through amazon.
> I took hundreds of pictures and they were great.
> great colors and white balance.

Opinion
Opinion Holder
Subject
Aspect
Evaluation

7

Opinion structurization

- Opinion：describes the opinion holders' positive or negative feelings for certain product (or a shop, a service, etc.)

text
> I bought powershot last week through amazon.
> I took hundreds of pictures and they were great.
> great colors and white balance.

Opinion
Opinion Holder 〈writer〉
Subject 〈powershot〉
Aspect 〈picture, color〉
Evaluation 〈great〉

- writer's evaluation (e.g. high, good, beautiful)
- writer's emotion or mental attitude (e.g. like, dislike)

8

Opinion structurization

- Opinion：describes the opinion holders' positive or negative feelings for certain product (or a shop, a service, etc.)

text
> I bought powershot last week through amazon.
> I took hundreds of pictures and they were great.
> great colors and white balance.

Opinion
Opinion Holder 〈writer〉
Subject 〈powershot〉
Aspect 〈picture, color〉
Evaluation 〈great〉

particular aspect of the subject that opinion holder have expressed evaluation on

9

Opinion structurization

- Opinion：describes the opinion holders' positive or negative feelings for certain product (or a shop, a service, etc.)

text
> I bought powershot last week through amazon.
> I took hundreds of pictures and they were great.
> great colors and white balance.

Opinion
Opinion Holder 〈writer〉
Subject 〈powershot〉
Aspect 〈picture, color〉
Evaluation 〈great〉

subject being evaluated
(e.g. product name, shop name,…)

10

Opinion-tagged corpus

- Japanese weblogs, 2,800 articles
  - Domain: restaurant, automobile, cell-phone, video game
  - Annotated tags
    - Subject (Aspect)-Evaluation
    - Aspect-Aspect/ Subject-Aspect

Subject-Aspect
> I bought powershot last week .
> I took hundreds of pictures and they were great.
> great colors and white balance.

Aspect-Evaluation   Aspect-Aspect

11

Statistics of opinion-tagged corpus

	restaurant	automobile	cell-phone	video game
# of articles	1,445	564	494	361
# of sentences	25,500	14,593	12,326	6,823
Asp—Eval	4,504	1,017	1,144	551
Asp—Asp	2,054	280	304	221
Subj—Eval	622	577	584	242
Subj—Asp	3,253	876	881	451

Asp: Aspect, Subj: Subject, Eval: Evaluation

12

Outline

- Background and aim
- Structuring opinions
- Approach for opinion extraction
- Experiment and result
- Conclusion and future work

13

Approach for opinion extraction

1. Extraction of Aspect（Subject）－Evaluation pairs
2. Extraction of aspect-aspect/subject-aspect

Mazda axela, I like the design.
But, the capacity of the trunk is insufficient.

| like → design |
| insufficient → capacity |

14

Approach for opinion extraction

1. Extraction of Aspect（Subject）－Evaluation pairs
2. Extraction of aspect-aspect/subject-aspect

Mazda axela, I like the design.
But, the capacity of the trunk is insufficient.

| like → design → Mazda axela |
| insufficient → capacity → trunk → Mazda axela |

15

Extraction of aspect-aspect/subject-aspect

- For given aspect
  - Identifying its most likelihood parent from the candidates

But, the capacity of the trunk is insufficient.

parent

aspect   aspect-of

aspect-of relation
Role-of：shop—salesperson, school—teacher, …
Attribute-of：engine—sound, interior—design, …
Part-of：car — engine, shop—interior, …

Our approach (1/2)

- Classification problem

I took hundreds of pictures and they were great.
great colors and white balance.

| colors → pictures | true |
| colors → white balance | false |

17

Our approach (2/2)

- The relation may be expressed in many patterns
  - エンジンのサウンド,エンジンは重厚なサウンド
    sound of the engine   the engine with full sound

→ learning efficient sub-trees to classify the examples

A車のエンジンは重厚なサウンドで良い
(Car A has the engine with full sound, and it is good)

candidate は   aspect で   evaluation

18

Efficient information

- Syntactic information (part-of-speech, etc.)
- Co-occurrence of 「A no B (B of A)」 collecting from large data
  - 「A no B」 express many types of relations
    - 画像の解像度（attribute），車のエンジン（part）
      (the image resolution)   (the engine of the car)
    - グレーの制服（property），私の車（ownership），…
      (the gray uniform)   (my car)

Using 「A no B」s where A and B have an "aspect-of" relation

19

Relation estimation

- Estimating whether a 「A no B」 is aspect-of or not
  - Finding R when P(R|A,B) is maximum
    - R: part-of, role-of, attribute-of, other relation

  - Estimating the probability model from labeled data
    - using maximum entropy method

20

Degree of co-occurrence

- Using conditional probability: P(A|B)
  - A and B has aspect-of relation

$$P(A \mid B, R = asp) = \frac{P(R = asp \mid A, B)}{P(R = asp \mid B)} P(A \mid B)$$

21

Outline

- Background and aim
- Structuring opinions
- Approach for opinion extraction
- Experiment and result
- Conclusion and future work

22

Settings

- Data (opinion tagged corpus)
  - restaurant domain 460 articles
  - 5-fold cross validation

- Learner: bact [Kudo 04]
  - bact: a boosting algorithm using decision stumps that use subtree as weak learners

- Features
  - functional word, part-of-speech,
  - rank of tf·idf score
  - rank of R(A|B,R=asp)

23

Result (parent identification)

baseline: the aspect-aspect pairs are assumed to appear in the 「AのB」 form

24
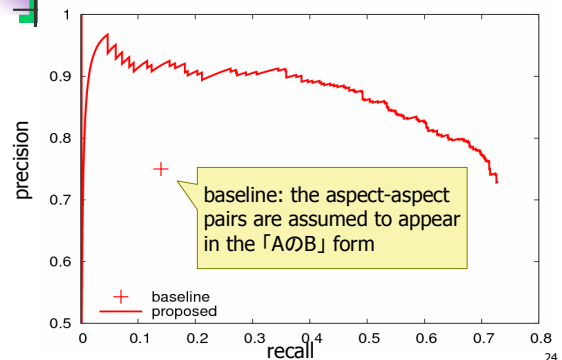
Conclusion

- For extracting opinions
  - Explained how the task of opinion extraction and structurization should be designed
  - Subject-aspect-evaluation chain
- Focus on the aspect (subject)-aspect pair extraction task
  - Applied the machine learning-based method to this task

Future direction

- Evaluation of our approach to opinion extraction
  - Combining
    - Extraction of Subject (Aspect) – Evaluation pair
    - Extraction of aspect-aspect/subject-aspect
- Refinement of opinion-tagged corpus