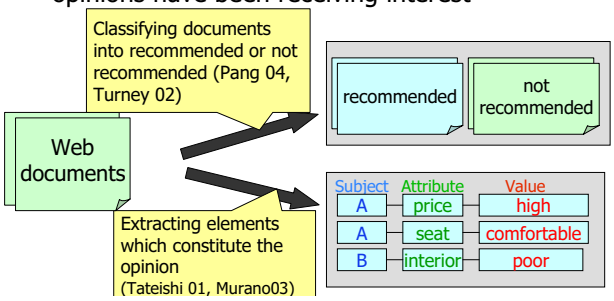
Collecting Evaluative Expressions for Opinion Extraction

Computational Linguistics Lab.
Nozomi Kobayashi

Background

- A huge amount of text data on the Web
  - Including human opinions
- Automatic extraction and classification of human opinions have been receiving interest



Aim

- To develop the automatic method to extract human opinions
  - Extraction of three elements

〈Subject, Attribute, Value〉
Subject: The subject of evaluation
Attribute: A particular aspect of a Subject
Value: A value of the Attribute

The Prius is quiet and the seats are very comfortable.

→ 〈Prius, φ, quiet〉
〈Prius, seats, comfortable〉

Subject/Attribute/Value

- Subject: product name, service name
  - Jaguar, Opel, Prius, BMW, …
- Attribute: constituents, qualitative and quantitative properties

domain	
Automobile	weight, seat, engine, space
Computer	CPU, speed, usability, design

- Value: quantity or quality of the Attribute

domain	
Automobile	heavy, elegant, good, exciting
Computer	attractive, poor, slow, robust

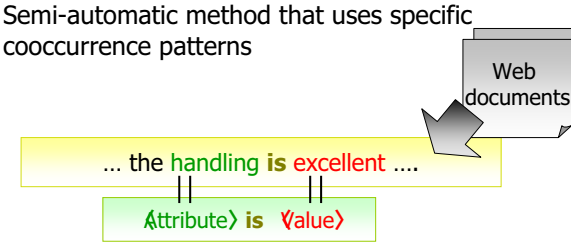
Related work

- Opinion extraction from Web documents
  - Using dictionaries and patterns (Tateishi01, Murano03)
  - Creating dictionaries manually
    - Attribute/Value expression are domain-dependent → costly

How to reduce the cost of creating a list of evaluative expressions?

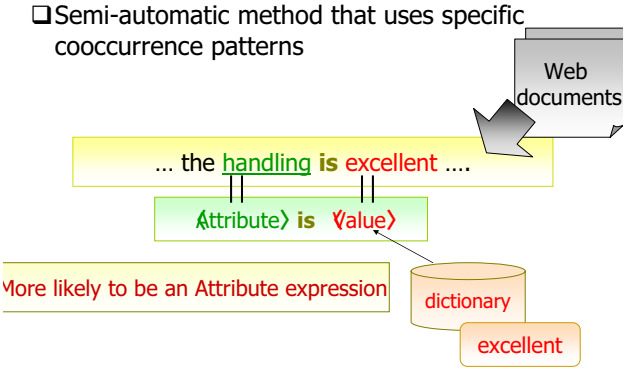
Approach

- Semi-automatic method that uses specific cooccurrence patterns

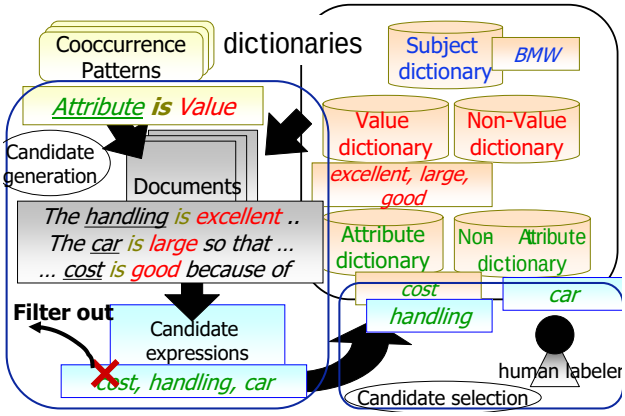


Approach

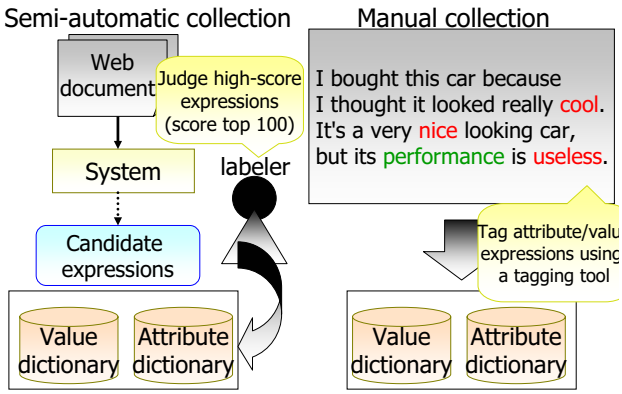
- Semi-automatic method that uses specific cooccurrence patterns



Process of collecting Attribute/Value expressions



Experiment (automobile and videogame)



Resources

- Document collections (from review sites)

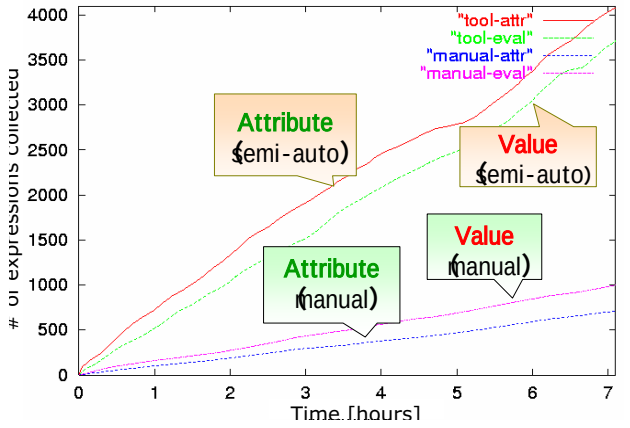
	Automobile [sentence]	Videogame [sentence]
manual	5,000	2,000
semi-auto	230,000	90,000

- Initial Subject dictionary
  - Automobile：389 expressions (*Jaguar, Opel, Chrysler,...*)
  - Videogame：660 expressions (*Dark Chronicle, Seaman,...*)
- Initial Attribute dictionary： 7 expressions
  - *cost, price, service, performance, function, support, design*
- Initial Value dictionary： 247 expressions
  - *good, beautiful, bright, favorite, high, . .*

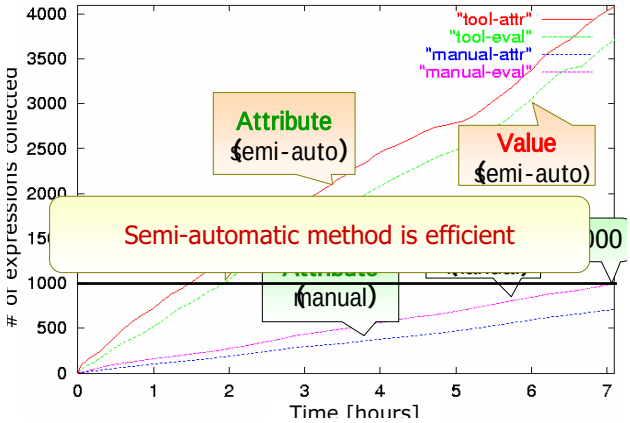
Cooccurrence patterns

- Attribute extraction patterns
  - Subject no Attribute ga/ha/mo/ni/wo Value
  - Subject no Attribute
  - Attribute ga/ha/mo/no/wo Value
  - Value na Attribute
- Value extraction patterns
  - Subject no Attribute ga/ha/mo/ni/wo Value
  - Attribute ga/ha/mo/ni/wo Value
  - Value na Attribute
  - Value Attribute (If Value is adjective)

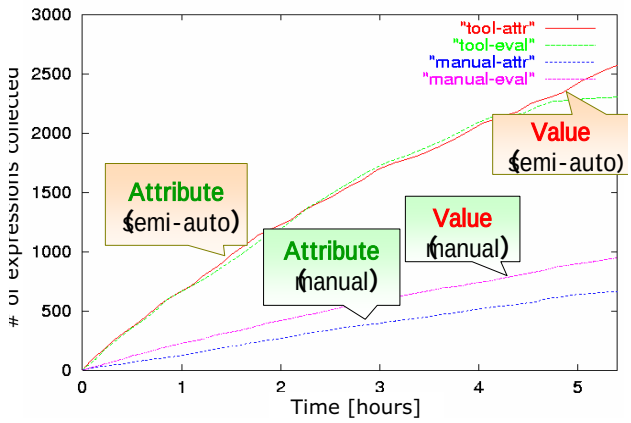
Number of expressions collected (Automobile)



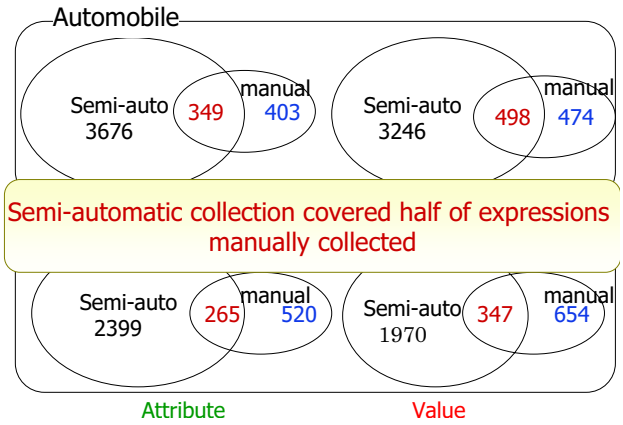
Number of expressions collected (Automobile)



Number of expressions collected (Videogame)

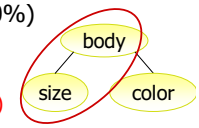


Coverage of expressions collected



Discussion (coverage)

- The main reason why coverage is not high: our semi-automatic method doesn't identify candidates whose span exceeds the boundary of the base phrase (bunsetsu)
  - Value expressions (30% to 40%)
    - Idioms
      - Nitari yottari (all alike)
      - Jigen-ga chigau (be on a different level)
  - Attribute expressions (30% to 40%)
    - Hierarchical relation between attribute (A no B)
      - body-no ookisa (size of body)



Examples of expressions collected

- Automobile

Shaju (weight)	body-no ookisa (size of body)
seijakusei (quietness)	akarui sikichou (light in color tone)

miwakuteki (alluring)	100% (100%)
dame (no good)	tsume-ga amai (easygoing)

- Videogame

CG (CG)	chiteki-na senryaku (intelligent strategy)
free mode	kyara-no graphic (graphic of characters)

soudai (magnificent)	sutoresu-ga tamaru (stressful)
comical	ki-ga nukenai (exciting)

※ left: collected only by semi-auto method
right: collected only by manual method

Conclusion

- Opinion extraction from given documents
  - Extraction of ⟨Subject, Attribute, Value⟩
  - We propose a semi-automatic method to collect Attribute/Value expressions
  - Cooccurrence patterns worked well across different domains
  - Semi-automatic method was more efficient than manual method

Future work

- Extraction of ⟨Subject, Attribute, Value⟩ using dictionaries
  - Classifying whether the collected value expression is used to express an "evaluation" in a given sentence

not evaluation

I bought the 300M brand new off the lot.
The seats are very comfortable.

  - Identifying the Subject/Attribute corresponding to a Value

The seats are very comfortable.

→ ⟨300M, seats, comfortable⟩