Opinion Mining Reviews

- A popular topic in opinion analysis is extracting sentiments related to products, entertainment, and service industries.
  - cameras, laptops, cars
  - movies, concerts
  - hotels, restaurants
- Common scenario: acquire reviews about an entity from the Web and extract opinion information about that entity.
- A single review often contains opinions that relate to multiple “aspects” of the entity, so each aspect and the opinion (evaluation) of that aspect must be identified.
  - laptop: fast processor, bulky charger
  - hotel: great location, tiny rooms

Opinion Extraction Task

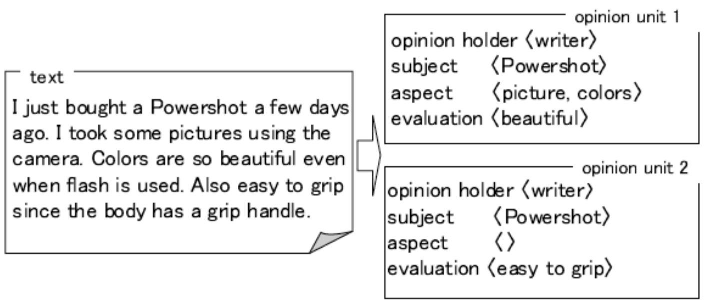
[Kobayashi et al., 2007] take the approach that most evaluative opinions can be structured as a frame consisting of:

- **Opinion Holder**: the person making the evaluation
- **Subject (Target)**: a named entity belonging to a class of interest (e.g., *iPhone*)
- **Aspect**: a part, member or related object, or attribute of the Subject (Target) (e.g., *size, cost*)
- **Evaluation**: a phrase expressing an evaluation or the opinion holder’s mental/emotional attitude (e.g., *too bulky*)

Opinion Extraction Task = filling these slots for each evaluation expressed in text.

Opinion Extraction Example

A review often contains multiple opinions, which are captured in separate frames. Each frame is referred to as an Opinion Unit.



Data Set

- 116 Japanese weblog posts about restaurants were randomly sampled from the *gourmet* category of a blog site.
- Two human annotators independently identified evaluative phrases and judged whether they related to a particular subject (restaurant).
- For these cases, the annotators were required to fill the opinion holder and subject slots. The aspect slot was filled only when a hierarchical relation between aspects was identified (e.g., *noodle* and its *volume*).
- An opinion unit was created for each evaluation in a sentence.

Inter-Annotator Agreement

Inter-annotator agreement (IAA) was measured as:

$$agr(A_1 \,||\, A_2) = \frac{\text{\# tags agreed by } A_1 \text{ and } A_2}{\text{\# tags annotated by } A_1}$$

For identifying evaluations:
$agr(A_1 \,||\, A_2) = .73$ & $agr(A_2 \,||\, A_1) = .83$ ⟹ F score = .79

For aspect-evaluation and subject-evaluation:
$agr(A_1 \,||\, A_2) = .86$ & $agr(A_2 \,||\, A_1) = .90$ ⟹ F score = .88

For subject-aspect and aspect-aspect relations:
$agr(A_1 \,||\, A_2) = .80$ & $agr(A_2 \,||\, A_1) = .79$ ⟹ F score = .79

Relation Subtasks

They evaluated the ability to identify specific relations within an opinion unit.

- **Aspect-Evaluation Relation**: evaluation of an aspect

  <*curry with chicken*, *was good*>

- **Aspect-Of Relation**: aspect of the entity being reviewed

  <*Bombay House*, *curry with chicken*>

- **Aspect-Aspect Relation**: hierarchical aspects

  <*picture*, *colors*> (e.g., colors in the picture … are beautiful!)

Data Set Statistics

Ultimately, they collected weblog posts for 4 domains:

(Restaurant, Automobile, cellular phone and video game)

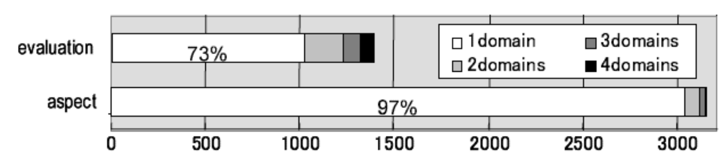
		Rest	Auto	Phone	Game
	articles	1,356	564	481	361
	sentences	21,666	14,005	11,638	6,448
	# of opinion units	4,267	1,519	1,518	775
I	Asp-Eval	3,692	943	965	521
	Asp-Asp	1,426	280	296	221
	Subj-Asp	2,632	877	850	451
II	Subj-Eval	575	576	553	243
	Subj-Asp-Eval	2,314	736	768	351
	Subj-Asp-Asp-Eval	1,065	175	172	127
	other	313	32	25	54
	Non-writer op. holder	95	17	22	2

The opinion holder was nearly always the writer, so they abandoned this subtask.

Domain Specificity

The aspect phrases are highly domain-specific:
only 3% occurred in > 1 domain!

The evaluation phrases also can vary across domains, but 27% occurred in multiple domains.



To further investigate, they created a ***dictionary of 5,550 evaluative expressions*** from 230,000 sentences in car reviews plus resources such as thesauri. The coverage was:

84% restaurants, 88% phones, 91% cars, 93% video games

Overall Approach

They adopt a 3-step procedure for opinion extraction:

1.Aspect-evaluation relation extraction: using dictionary look-up, find candidate evaluation expressions and identify the target (subject or aspect).

2.Opinion-hood determination: for each <target, evaluation> pair, determine whether it is an opinion based on its context.

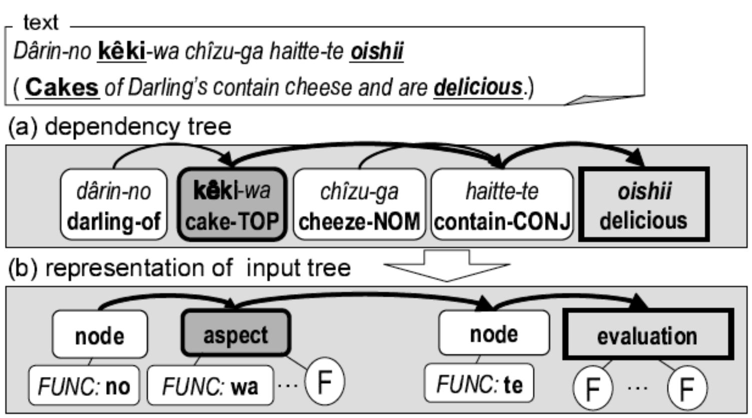
3.Aspect-of relation extraction: for each <aspect, evaluation> pair judged to be an opinion, search for the aspect's antecedent (either a higher aspect or its subject).

Interesting observation: Aspect-of relations are a type of bridging reference!

Aspect-Evaluation and Aspect-Of Relation Detection

- Given an evaluation phrase and candidate aspect, a "contextual" classifier is trained to determine whether the pair have an aspect-evaluation relation.
- If the classifier finds > 1 aspect that is related to the evaluation, then the one with the highest score is chosen.
- To encode training examples, each sentence with an evaluation is parsed. The path linking the evaluation and candidate is extracted, along with the children of each node.
- A classifier is trained with a Boosting learning algorithm using a variety of features.
- A similar classifier is also trained for the AspectOf relation.

Example of Instance Representation



Feature Sets

Features for contextual clues
- Position of c / t in the sentence (beginning, end, other)
- Base phrase distance between c and t (1, 2, 3, 4, other)
- Whether c and t has a immediate dependency relation
- Whether c precedes t
- Whether c appears in a quoted sentence
- Part-of-speech of c / t
- Suffix of c (*-sei*, *-sa* (-ty), etc.)
- Character type of c (*English*, *Chinese*, *Katakana*, etc.)
- Semantic class of c derived from *Nihongo Goi Taikei* (Ikehara et al., 1997).

Features for statistical clues
- Co-occurrence score rank of c (1st, 2nd, 3rd, 4th, other)
- Aspect-hood score rank of c (1st, 2nd, 3rd, 4th, other)

Context-Independent Statistical Clues

- **Co-occurrence Clues**: aspect-aspect and aspect-evaluation co-occurrences were extracted from 1.7 million weblog posts using 2 simple patterns.

  Probabilistic latent semantic indexing (PLSI) was used to estimate the conditional probabilities:

  P(*Aspect* | *Evaluation*) P(*Aspect_A* | *Aspect_B*)

- **Aspect-hood of Candidate Aspects**: the plausibility of a term being an aspect is estimated based on how often it directly co-occurs with a subject in the domain.

  PMI is used to measure the strength of association between candidates X and Y extracted from specific patterns.

Inter-sentential Relation Extraction

- If no aspect is identified for an evaluation expression within the same sentence, then the preceding sentences are searched.
- This task is viewed as **zero-anaphora resolution**, so a specialized zero-anaphora resolution supervised learning model is used.
- Zero anaphora occur when a reference to something is understood but there is no lexical realization of it. (This is very common in Japanese and many other languages, but less common in English.) Example:

  "John fell and broke his leg."

Opinion-hood Determination

- Evaluative phrases may not refer to the target (or any aspect of it). For example:

  "The weather was good so I took some pictures with my new camera."

- So an SVM classifier was trained to determine whether an <aspect, evaluation> pair truly represents an opinion.
- Positive training examples came from the annotated corpus. Negative training examples are artificially generated:
  - for each evaluation phrase in the dictionary, extract the most plausible candidate aspect using the prior method
  - if the candidate is not correct, it's a negative example

Experimental Results

Experiments were performed on 395 weblog posts in the restaurant domain using 5-fold cross validation. A previous pattern-based method (*Patterns*) was used as a baseline.

Table 3: The results of aspect-evaluation relation

		intra-sent.	inter-sent.
Patterns	P	0.56 (432/774)	-
	R	0.53 (432/809)	-
Contextual	P	0.70 (504/723)	0.13 (46/360)
	R	0.62 (504/809)	0.17 (46/274)
Contextual +statistics	P	0.72 (502/694)	0.14 (53/389)
	R	0.62 (502/809)	0.19 (53/274)

Inter-sentential performed poorly because the syntactic features could not be used, only the statistical clues.

Aspect-Of Relation Results

Since the Aspect-Of relation is similar to bridging references, a statistical co-occurrence model (*Co-occurrence*) used for bridging reference resolution was used as a baseline.

GIven an aspect, "the nearest candidate that has the highest positive score of the PMI" is selected.

Table 4: The results of aspect-of relation

	precision	recall
Co-occurrence	0.27 (175/ 682)	0.17 (175/1048)
Contextual	0.44 (458/1047)	0.44 (458/1048)
Contextual+statistics	0.45 (474/1047)	0.45 (474/1048)

Opinion-hood Evaluation

- The opinion-hood classifier achieved only 50% precision with 45% recall.
- They note that this task encompasses two subproblems:
  - is the evaluation expression truly an opinion?
  - does the evaluation expression apply to the domain (target/aspect)?
- To illustrate how challenging the aspect-evaluation task can be, note that similar sentences can have different labels:

"I like shrimps." (general personal preference)

"I like shrimps of the restaurant." (opinion about restaurant)

Cross-Domain Portability

Table 5: Comparing intra-sentential models among three domains (upper: aspect-eval, lower: aspect-of)

test		restaurant	cellular phone	automobile
same dom.	P	0.72 (502/694)	0.75 (522/693)	0.76 (562/738)
	R	0.62 (502/809)	0.63 (522/833)	0.65 (562/870)
other dom	P	0.73 (468/638)	0.72 (517/710)	0.74 (565/768)
	R	0.58 (468/809)	0.62 (517/833)	0.65 (565/870)
same dom.	P	0.43 (139/321)	0.62 (139/224)	0.66 (185/280)
	R	0.59 (139/234)	0.60 (139/230)	0.66 (185/279)
other dom	P	0.42 (124/293)	0.53 (138/260)	0.59 (195/329)
	R	0.52 (124/234)	0.60 (138/230)	0.70 (195/279)

Conclusions

- There are a ton of applications for opinion extraction! Most people think only of the opinion expression, but for real applications:
  - many additional things need to be extracted: *holder, target, aspects*
  - and each linked to an opinion expression!
- This area has been very active, and a lot of progress has been made.
- But this is a challenging task because of the diversity of opinion expressions and the underlying information extraction subtasks. Much future work to be done!