

Mining Product Features and Customer Opinions from Reviews

Ana-Maria Popescu
University of Washington
<http://www.cs.washington.edu/research/knowitall/>

Outline

- 1 Mining Customer Reviews: Related Work
- 2 OPINE: Tasks and Results
- 3 Product Feature Extraction
- 4 Customer Opinion Extraction
- 5 Conclusions and Future Work

2

Mining Customer Reviews

Finding and analyzing subjective phrases or sentences
Positive: **The hotel had a great location.**
Takamura'05, Wilson'04, Turney'03, Riloff et al.'03, etc.

Classifying consumer reviews
polarity classification, strength classification
Trump International : Review #4 : positive: ***
Turney'02, Pang et al.'05, Pang et al.'02, Kushal et al.'03, etc.

Extracting product features and opinions from reviews
hotel_location:great[+]
Hu & Liu'04, Kobayashi'04, Yi et al.'05, Gamon et al.'05, OPINE

3

Tasks and Results

Identify product features P = 94% R = 77%

Identify the semantic orientation of potential opinion words (adjectives, nouns, etc.) in the context of product features and review sentences. P = 78% R = 88%

Identify opinion phrases P = 79% R = 76%
Identify opinion phrase polarity P = 86% R = 89%

5

OPINE

KnowItAll is a Web-based information extraction system (Etzioni et al'05)

Given a target class (Country)
The Extractor instantiates extraction rules ("country such as [X]") and uses search engine to find candidate instances
The Assessor eliminates incorrect candidates using high-precision lexical patterns

$$PMI("([X] and other countries", "Garth Brooks") = \frac{\text{hits}("Garth Brooks and other countries")}{\text{hits}("and other countries") * \text{hits}("Garth Brooks")}$$

OPINE is built on top of KnowItAll
It uses and extends KnowItAll's architecture
It extensively uses high-precision lexical patterns
It uses the Web to collect statistics

6

Outline

- 1 Mining Customer Reviews: Related Work
- 2 OPINE: Tasks and Results
- 3 **Product Feature Extraction**
- 4 Customer Opinion Extraction
- 5 Conclusions and Future Work

7

Feature Extraction

Product classes: Hotels
Instances: Trump International

Features	Examples
Properties	Quality Size
Parts	Room
Features of parts	RoomSize
Related concepts	Neighborhood
Features of related concepts	NeighborhoodSafety

8

Feature Extraction

I loved the hot water and the clean bathroom. Extract noun phrases *np* such that *np* contains only nouns and $\text{frequency}(np) > 1$ as potential features.

The fan was broken and our room was hot the entire time.

I like a nice, hot room when the snow piles up outside.

9

Feature Extraction

I loved the hot water and the clean bathroom. Assess potential features using bootstrapped lexical patterns (*discriminators*)

The fan was broken and our room was hot the entire time.

I like a nice, hot room when the snow piles up outside.

Examples

- X of Y
- Y with X
- Y equipped with X
- Y boasts X
- Y has X
- Y comes with X
- Y contains X
- Y offers X

10

Feature Extraction

I loved the hot water and the clean bathroom. Assess potential features using discriminators

The fan was broken and our room was hot the entire time.

I like a nice, hot room when the snow piles up outside.

$$\text{PMI}(\text{hotel's } [Y], \text{room}) = \frac{\text{hits}(\text{"hotel's room"})}{\text{hits}(\text{"hotel's"}) * \text{hits}(\text{"room"})}$$

$$\text{PMI}(\text{hotel's } [Y], \text{room}) = 0.54 * 10^{-13}$$

$$\text{PMI}(\text{hotel's } [Y], \text{snow}) = 0.64 * 10^{-16}$$

$\text{PMI}(\text{hotel's } [Y], \text{room}) \gg \text{PMI}(\text{hotel's } [Y], \text{snow})$

11

Feature Extraction

I loved the hot water and the clean bathroom. Assess potential features using discriminators

The fan was broken and our room was hot the entire time.

I like a nice, hot room when the snow piles up outside.

$$\text{PMI}(\text{hotel's } [Y], \text{room}) = \frac{\text{hits}(\text{"hotel's room"})}{\text{hits}(\text{"hotel's"}) * \text{hits}(\text{"room"})}$$

$$\text{PMI}(\text{hotel's } [Y], \text{room}) = 0.54 * 10^{-13}$$

$$\text{PMI}(\text{hotel's } [Y], \text{snow}) = 0.64 * 10^{-16}$$

$\text{PMI}(\text{hotel's } [Y], \text{room}) \gg \text{PMI}(\text{hotel's } [Y], \text{snow})$

12

Results

5 consumer electronics product classes (Hu&Liu'04) 314 reviews

	Hu&Liu	OPINE-Web	OPINE
P	0.72	0.79	0.94
R	0.80	0.66	0.77

1/3 of OPINE precision increase is due to OPINE assessment
2/3 of OPINE precision increase is due to Web PMI statistics

2 product classes (Hotels, Scanners) 1307 reviews
P = 89% R = 73%

Outline

- 1 Mining Customer Reviews: Related Work
- 2 OPINE: Tasks and Results
- 3 Product Feature Extraction
- 4 **Customer Opinion Extraction**
- 5 Conclusions and Future Work

14

Potential Opinions

I loved the hot water and the clean bathroom.

Use syntax-based rules to extract potential opinions po for each feature f

The fan was broken and our room was hot the entire time.

If [subj= f , pred = be , arg]
 $po := arg$

I like a nice, hot room when the snow piles up outside.

If [subj, pred, obj= f]
 $po := pred$
...
(similar intuition to Kim&Hovy'04, Hu&Liu'04)

15

Semantic Orientation

The room was hot(-) and stuffy(-).
After freezing for hours, the room was nice(+) and hot(+).

cold basic loud visible casual modern central quiet

Task

Compute the SO label for a (word, feature, sentence) tuple

Solution

- I Compute the SO label for each word
- II Compute the SO label for each (word, feature) pair
- III Compute the SO label for each (word, feature, sentence) tuple

Each solution step = labeling problem \rightarrow relaxation labeling

16

Relaxation Labeling

Input

objects, labels
an initial $object \leftrightarrow label$ mapping
an object's neighborhood
a support function q for an object label

Output

final $object \leftrightarrow label$ mapping

RL Update Equation

$$P(w = L)_{m+1} = \frac{P(w = L)_m (1 + q(w, L)_m)}{\sum_{L'} P(w = L')_m (1 + q(w, L')_m)}$$

$w =$ word, $L =$ SO label

17

Word Semantic Orientation

Building word neighborhoods:

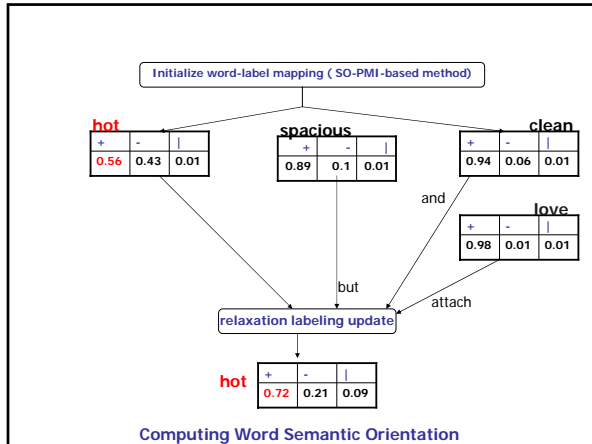
conjunctions, disjunctions,
syntactic attachment rules
WordNet synonymy/antonymy
morphology information

I loved the hot water and the clean bathroom.

neighbor(hot, love, synt_dep_path) neighbor(hot, clean, and)

The room was spacious but hot.

neighbor(hot, spacious, but)



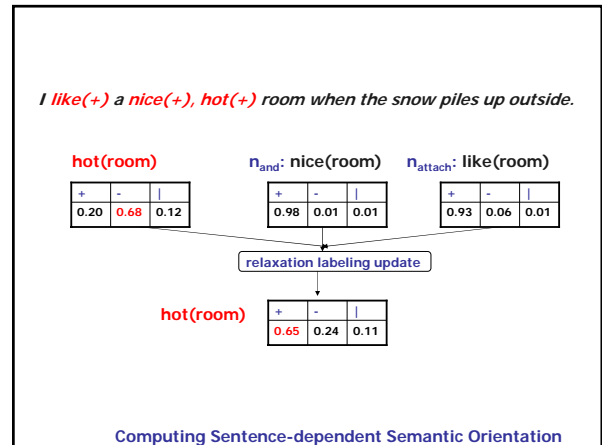
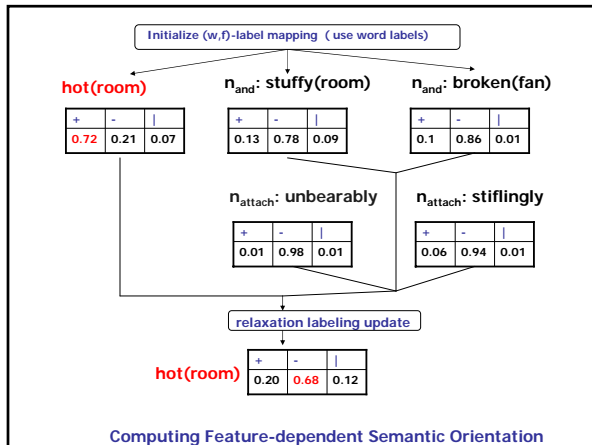
Semantic Orientation

Potential opinion words can change orientation based on features

I loved the hot water and the clean bathroom.
The fan was broken and our room was hot the entire time.
Our room was really hot.

Compute the SO label for each (word, feature) pair

20



Results

PMI ++: Version of PMI-based method for finding SO labels of words or (word, feature) pairs

Hu ++: Version of Hu's WordNet-based method for finding word SO labels

OP-1: OPINE version which only computes the dominant SO label of a word

	PMI ++	Hu ++	OP-1	OPINE
P	0.72	0.74	0.69	0.78
R	0.91	0.78	0.88	0.88

OPINE's improvements are mostly due to contextual information use

23

Issues

- Parsing errors (especially in long-range dependency cases)
 - missed candidate opinions
 - incorrect polarity assignment
- Sparse data problems for infrequent opinion words
 - incorrect polarity assignment
- Complicated opinion expressions
 - opinion nesting, conditionals, subjunctive expressions, etc.

24

Opinion Ranking

Cluster opinion phrases and label clusters

Room Cleanliness: clean, dirty, spotless, incredibly clean
 Room Size: spacious, cramped, tiny, huge

Use lexical patterns to compute relative strength constraints

clean, even spotless
 clean, almost spotless
 clean, but not spotless

→ strength(spotless) > strength(clean)
 (following a suggestion from Hatsivassiloglou'93)

Compute attribute-specific opinion ordering

Room Cleanliness: spotless, incredibly clean, very clean, clean

25

Conclusions

1. OPINE successfully extends KnowItAll's generate-and-test architecture for the task of mining reviews.
2. OPINE benefits from using Web PMI statistics for product feature validation.
3. OPINE benefits from using contextual information when finding the semantic orientation of potential opinion words.

26

Current Work

Identify positive or negative opinion sentences corresponding to a given feature:

The room was small, but clean and overall great for the price.

Identify positive or negative opinion sentences for the product

Identify specific problems with a given product

The laptop froze when he restarted it.
 The laptop froze when a certain battery capacity was trespassed.
 The laptop froze when it was moved.

Extend OPINE to open-domain text (newspaper articles)

27

Opinion Phrases

I loved(+) the hot(+) water in the shower.
 Opinion phrases are phrases with a positive or negative head: love, hot, broken, like, really hot

The fan was broken(-) and our room was hot(-) the entire time.
 Opinion phrase polarity is determined by the context-dependent semantic orientation of the head word

I like(+) a nice(+), hot(+) room when the snow piles up outside.

Our room was really hot(-).

28

Results

Data: 550 sentences containing extracted features
 1036 potential opinion phrases

	PMI ++	Hu ++	OPINE
OP Extraction: Precision	0.71	+0.06	+0.08
OP Extraction: Recall	0.78	-0.08	-0.02
OP Polarity: Precision	0.80	-0.08	+0.06
OP Polarity: Recall	0.82	+0.07	-0.04

29