Phrase-Level Contextual Sentiment Analysis

Wilson et al. (2005), "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis"

Choi and Cardie (2008), "Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis"

C.J. Hsu, Ryan Bielby

Introduction

- Tasks such as summarization, opinion-oriented information extraction and mining product reviews require sentence-level or even phrase-level sentiment analysis.
- General methods like bag-of-words are not enough, words or constituents can interact with each other to yield a particular overall polarity.

Frameworks for Phrase-Level Sentiment Analysis

- The first framework: Try to predict the polarity of the subjective words of the given phrase. Instead of predicting the polarity of overall phrase.
- The second framework: Try to predict the polarity of the overall phrase, however, the boundary is already defined by human annotators.
- None of these frameworks try to find the polar boundary of the given sentence.

Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis *(Wilson et al., 2005)*

- Although the name of paper refers to phrase-level sentiment analysis, they try to solve this problem by looking at "words"
- Goal: identify polarity of the subjective keywords in an expression
- E.g.
  - Thousands of coup supporters <u>celebrated</u> (**positive**) overnight, waving flags, …
  - … the three countries in question are <u>repressive</u> (**negative**) and <u>grave human rights violators</u> (**negative**)

Dataset: Corpus

- Multi-perspective Question Answering (MPQA) Opinion corpus
  - Marks subjective expressions in sentences
  - Does not indicate polarity of the expressions
  - Need human annotation

Dataset: Annotation

- Manually annotate subjective expressions from MPQA for polarity
- *positive*, *negative*, *both*, *neutral*
- Inter-annotator agreement study
  - 447 expressions
  - 2 annotators
  - 82% agreement, $\kappa = 0.72$
  - *uncertain* tag
    - 18% of expressions
    - 90% agreement, $\kappa = 0.84$

Dataset: Breakdown

- 15,991 subjective expressions from 425 docs
  - devset: 66 docs, 2808 subjective expressions
  - 10-fold cross-validation: 359 docs, 7611 expressions
- Prior-polarity subjectivity lexicon (8,000 words)
  - Riloff and Wiebe, 2003
  - Hatzivassiloglou and McKeown, 1997
  - General Inquirer, 2000
  - Reliability tags: *strongsubj* and *weaksubj*
  - 33.1% *positive*, 59.7% *negative,* 0.3% *both*, 6.9% *neutral*

Setup

- Gold Standard
  - *neutral*: not in subjective expression
  - *both*: in at least one positive and one negative expression
  - *negative*: only in negative or neutral expressions
  - *positive*: only in positive or neutral expressions
- BoosTexter AdaBoost.HM machine learning algorithm

Two Part Classification

- Using only prior-polarity classifier:
  - 48% accuracy

		Prior-Polarity Classifier				
		Neut	Pos	Neg	Both	Total
Gold	Neut	798	784	698	4	2284
	Pos	81	371	40	0	492
	Neg	149	181	622	0	952
	Both	4	11	13	5	33
	Total	1032	1347	1373	9	3761

  - 76% errors from non-neutral prior polarity words appearing in contextually neutral polarity phrases

Neutral-Polar Classification

- Features:
  - Word (5 features)
    - Word part-of-speech
  - Modification (8 features)
    - Preceded by adjective
  - Sentence (11 features)
    - Adjectives in sentence
  - Structure (3 features)
    - In subject
  - Document (1 feature)
    - Document topic
- Results:

	Acc	Polar Rec	Polar Prec	Polar F	Neut Rec	Neut Prec	Neut F
word token	73.6	45.3	72.2	55.7	89.9	74.0	81.2
word+priorpol	74.2	54.3	68.6	60.6	85.7	76.4	80.7
28 features	75.9	56.8	71.6	63.4	87.0	77.7	82.1

Polarity Classification

- Features:
  - Word (2 features)
    - Word prior polarity
  - Polarity (8 features)
    - Negated
- Results:

		Positive			Negative			Both			Neutral		
	Acc	Rec	Prec	F	Rec	Prec	F	Rec	Prec	F	Rec	Prec	F
word token	61.7	59.3	63.4	61.2	83.9	64.7	73.1	9.2	35.2	14.6	30.2	50.1	37.7
word+priorpol	63.0	69.4	55.3	61.6	80.4	71.2	75.5	9.2	35.2	14.6	33.5	51.8	40.7
10 features	65.7	67.1	63.3	65.1	82.1	72.9	77.2	11.2	28.4	16.1	41.4	52.4	46.2

Critique

- No reasoning given for feature choices
- Low polarity recall (56.8%) from first classifier whose results feed into second classifier
- No scores given for the polarity
- Try to classify phrase-level sentiment based only on polarity of one word in the phrase which must appear in their lexicon

Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis

Yejin Choi and Claire Cardie 2008

The Weakness of Previous Work

- Previous work mostly focuses on the model of "a bag of words", structure information is barely employed.
- The previous work employing structure information uses rule-based system to depict this structure information.
- Previous work focuses on function word negators, other negators like content word negators are also important. Eg: "eliminate"

The Contributions of this Work

- Heuristic-Based and Learning-Based methods which both incorporate structure information motivated by compositional semantics.
- The results of both methods get improved when employing the features of compositional semantics.
- Incorporating the features about content word negators.
- Explorations about incorporating the features outside the phrase

Dataset : Two Lexicons

- The polarity lexicon is initialized with the lexicon of Wilson et al (2005) and expanded using the General Inquirer Dictionary.
- The negator lexicon is initialized with a handful of seed words and then expanded by the corresponding synonyms of WordNet.

Dataset : Corpus

- Multi-Perspective Question Answering (MPQA) corpus provides the documents annotated with phrase-level subjectivity information.
- Performance is reported using 10-fold cross-validation on 400 documents; a separate 135 documents were used as a development set.

Heuristic-Based Methods

- Six types of Heuristic-Based methods are proposed. Except the one employing compositional semantics, the other five methods are all based on counting and voting.

	VOTE	NEG(1)	NEG(N)	NEGEX(1)	NEGEX(N)	COMPO
type of negators	none	function-word		function-word & content-word		
maximum # of negations applied	0	1	n	1	n	n
scope of negators	N/A	over the entire expression				compositional

Heuristic-Based Methods

- VOTE : Assign the majority polarity to given phrase.
- NEG(1) : If function word negators found in this phrase, flip the majority polarity.
- NEG(N) : Same up NEG(1), except the number of flips is the number of function word negators in this phrase.

Heuristic-Based Methods

- NEGEX(1) : Same as NEG(1), but also takes content word negators into consideration.
- NEGEX(N) : Same as NEG(N), but also takes content word negators into consideration.
- COMPO : Apply a series of phrase structure rules to decide the polarity of the given phrase.

The Rules about Compositional Semantics

	Rules		Examples
1	Polarity(not_[arg1]) =	$\neg$ Polarity(arg1)	not [bad]$_{arg1}$.
2	Polarity([VP]_[NP]) =	Compose([VP], [NP])	[destroyed]$_{VP}$ [the terrorism]$_{NP}$.
3	Polarity([VP1]_to_[VP2]) =	Compose([VP1], [VP2])	[refused]$_{VP_1}$ to [deceive]$_{VP_2}$ the man.
4	Polarity([adj]_to_[VP]) =	Compose([adj], [VP])	[unlikely]$_{adj}$ to [destroy]$_{VP}$ the planet.
5	Polarity([NP1]_[IN]_[NP2]) =	Compose([NP1], [NP2])	[lack]$_{NP_1}$ [of]$_{IN}$ [crime]$_{NP_2}$ in rural areas.
6	Polarity([NP]_[VP]) =	Compose([VP], [NP])	[pollution]$_{NP}$ [has decreased]$_{VP}$.
7	Polarity([NP]_be_[adj]) =	Compose([adj], [NP])	[harm]$_{NP}$ is [minimal]$_{adj}$.

Definition of Compose(arg1, arg2)	
	Compose(arg1, arg2) =
For COMPOMC: (**COMPO**sition with **M**ajority **C**lass)	if (arg1 is a negator) then $\neg$ Polarity(arg2) else if (Polarity(arg1) == Polarity(arg2)) then Polarity(arg1) else the majority polarity of data
	Compose(arg1, arg2) =
For COMPOPR: (**COMPO**sition with **PR**iority)	if (arg1 is a negator) then $\neg$ Polarity(arg2) else Polarity(arg1)

Learning-based Methods

- Class Labels: Positive and Negative
- Algorithms: Online SVM MIRA. The general idea is like SVM. However, this algorithm integrates one training instance one by one instead of taking all training instances into consideration at one time.

Feature Template for Learning-based Methods

- Lexical Feature Template:

  current_word (x_i), lemma(x_i), is_stopword(x_i)

- Dictionary Feature Template:

  category(x_i)

- Voting Feature Template:

  Encode the polarity of overall phrase

Learning-Based Methods
Incorporate the compositional semantics

Simple Classification	Classification with Compositional Inference
$y \leftarrow \text{argmax}_y \text{ score}(y)$ $l \leftarrow \text{loss\_flat}(y^*, y)$ $\mathbf{w} \leftarrow \text{update}(\mathbf{w}, l, y^*, y)$	Find K best $\mathbf{z}$ and denote them as $\mathcal{Z} = \{\mathbf{z}^{(1)}, ..., \mathbf{z}^{(K)}\}$ $\quad s.t. \forall\, i < j,\ \text{score}(\mathbf{z}^{(i)}) > \text{score}(\mathbf{z}^{(j)})$ $\mathbf{z}^{bad} \leftarrow \min_k \mathbf{z}^{(k)}\ s.t.\ \text{loss\_compo}(y^*, \mathbf{z}^{(k)}, x) > 0$ (if such $\mathbf{z}^{bad}$ not found in $\mathcal{Z}$, skip parameter update for this.) If $\text{loss\_compo}(y^*, \mathbf{z}^*, x) > 0$ $\quad \mathbf{z}^{good} \leftarrow \min_k \mathbf{z}^{(k)}\ s.t.\ \text{loss\_compo}(y^*, \mathbf{z}^{(k)}, x) = 0$ $\quad z^* \leftarrow \mathbf{z}^{good}$ (if such $\mathbf{z}^{good}$ not found in $\mathcal{Z}$, stick to the original z^*.) $l \leftarrow \text{loss\_compo}(y^*, \mathbf{z}^{bad}, x) - \text{loss\_compo}(y^*, \mathbf{z}^*, x)$ $\mathbf{w} \leftarrow \text{update}(\mathbf{w}, l, \mathbf{z}^*, \mathbf{z}^{bad})$
Definitions of score functions and loss functions	
$\text{score}(y) := \mathbf{w} \cdot \mathbf{f}(x, y)$ $\text{loss\_flat}(y^*, y) := \text{if } (y^* = y)\ 0 \text{ else } 1$	$\text{score}(\mathbf{z}) := \sum_i \text{score}(z_i) := \sum_i \mathbf{w} \cdot \mathbf{f}(x, z_i, i)$ $\text{loss\_compo}(y^*, \mathbf{z}, x) := \text{if } (y^* = \mathcal{C}(x, \mathbf{z}))\ 0 \text{ else } 1$

Evaluations: Heuristics-Based vs. Learning-Based

Heuristic-Based							Learning-Based			
VOTE	NEG (1)	NEG (N)	NEG EX (1)	NEG EX (N)	COMPO MC	COMPO PR	SC VOTE	SC NEG EX	CCI COMPO MC	CCI COMPO PR
86.5	82.0	82.2	87.7	87.7	89.7	89.4	88.5	89.1	90.6	90.7

Evaluations: Fixed Boundary vs. Extended Boundary

- So far, only words inside the phrases are taken into consideration. What if we take more words into consideration?

Data	Heuristic-Based							Learning-Based			
	VOTE	NEG (1)	NEG (N)	NEG EX (1)	NEG EX (N)	COMPO MC	COMPO PR	SC VOTE	SC NEG EX	CCI COMPO MC	CCI COMPO PR
[-0,+0]	86.5	82.0	82.2	87.7	87.7	89.7	89.4	88.5	89.1	90.6	90.7
[-1,+1]	86.4	81.0	81.2	87.2	87.2	89.3	89.0	88.3	88.4	89.5	89.4
[-5,+5]	85.9	79.0	79.4	85.7	85.6	88.2	88.0	86.4	87.1	88.7	88.7
[-∞,+∞]	85.3	75.8	76.9	83.9	83.9	87.0	86.9	85.8	85.8	87.3	87.5

Critique

- Not too much gain from heuristic-based compared to learning-based methods.
- Could we try to encode the info of compositional semantics in the feature level instead of rewriting the training phase of SVM?

Critique

- The average accuracy for the previous paper is around 65%, but the result of this paper could reach to about 90%. What makes this difference?
- Any more sophisticated ways to extend the boundary?