Use of Discourse Knowledge to Improve Lexicon-based Sentiment Analysis

Pedro Paulo Balage Filho

University of Wolverhampton, Universidade do Algarve

Supervisors:

Dr. Constantin Orăsan (University of Wolverhampton)

Prof. Dr. Mário Silva (Instituto Superior Técnico)

June, 2012

Outline

Sentiment Analysis

Definition

Sentiment analysis deals with the computational treatment of opinion, sentiment and subjectivity in text (Pang el al., 2002).

- Task: text classification
- Sentiment: positive and negative

Sentiment Analysis - Example

It could have been a great movie. It could have been excellent, and to all the people who have forgotten about the older, greater movies before it, will think that as well. It does have beautiful scenery, some of the best since Lord of the Rings. The acting is well done, and I really liked the son of the leader of the Samurai. He was a likeable chap, and I hated to see him die... But, other than all that, this movie is nothing more than hidden rip-offs.

Sentiment Analysis - Approaches

- Machine Learning
  - corpus for training
  - bag-of-words features
  - covers domain dependence
- Lexicon based
  - uses a dictionary of terms and their semantic orientation
  - averages the semantic orientations for the words found in the text
  - good for general domain
  - easy to include linguistic knowledge

SO-CAL (Taboada et al.,2006; Taboada and Grieve, 2004)

- Each word has a semantic orientation (SO) measured by a value

  This is a **good** *(+3) movie.*
  $SO = +3$

- Negation:

  Not *good (+3)*
  $SO = 3 - 4 = -1$

- Intensifier:

  really very *good (+3)*
  $SO = (3 \times [100\% + 25\%]) \times (100\% + 15\%) = 4.3$

- Irrealis:

  This **should** *have been a great (+3) movie.*
  $SO = 0$

SO-CAL (Taboada et al.,2006; Taboada and Grieve, 2004)

- Each word has a semantic orientation (SO) measured by a value

  This is a **good** *(+3) movie.*
  $SO = +3$

- Negation:

  Not *good (+3)*
  $SO = 3 - 4 = -1$

- Intensifier:

  really very *good (+3)*
  $SO = (3 \times [100\% + 25\%]) \times (100\% + 15\%) = 4.3$

- Irrealis:

  This **should** *have been a great (+3) movie.*
  $SO = 0$

SO-CAL (Taboada et al.,2006; Taboada and Grieve, 2004)

- Each word has a semantic orientation (SO) measured by a value

  This is a **good** *(+3) movie.*
  $SO = +3$

- Negation:

  Not *good (+3)*
  $SO = 3 - 4 = -1$

- Intensifier:

  really very *good (+3)*
  $SO = (3 \times [100\% + 25\%]) \times (100\% + 15\%) = 4.3$

- Irrealis:

  This **should** *have been a great (+3) movie.*
  $SO = 0$

SO-CAL (Taboada et al.,2006; Taboada and Grieve, 2004)

- Each word has a semantic orientation (SO) measured by a value

  This is a **good** *(+3) movie.*
  $SO = +3$

- Negation:

  Not *good (+3)*
  $SO = 3 - 4 = -1$

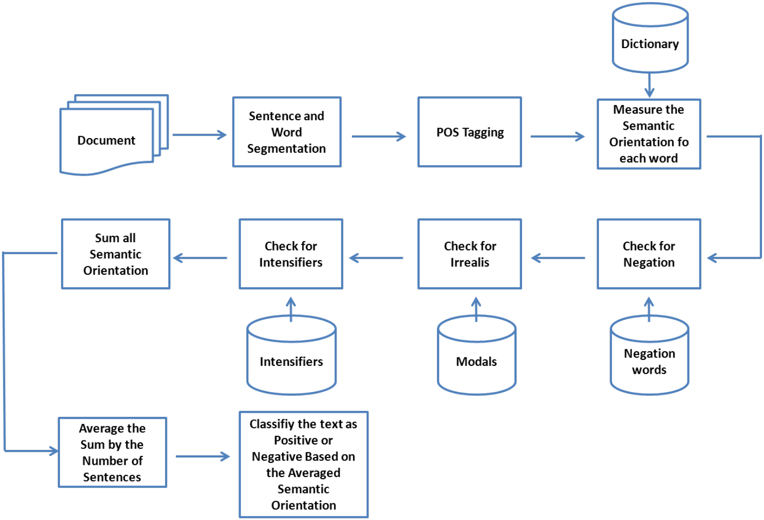
- Intensifier:

  really very *good (+3)*
  $SO = (3 \times [100\% + 25\%]) \times (100\% + 15\%) = 4.3$

- Irrealis:

  This **should** *have been a great (+3) movie.*
  $SO = 0$

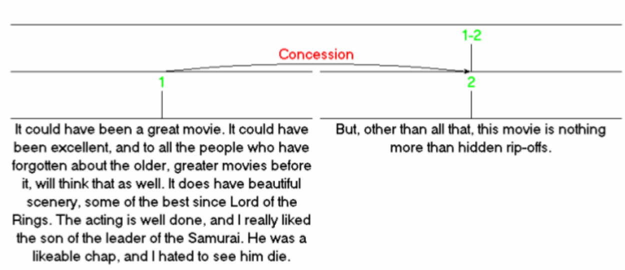
SO-CAL


Dictionary
Document
Sentence and Word Segmentation
POS Tagging
Measure the Semantic Orientation fo each word
Sum all Semantic Orientation
Check for Intensifiers
Check for Irrealis
Check for Negation
Intensifiers
Modals
Negation words
Average the Sum by the Number of Sentences
Classifiy the text as Positive or Negative Based on the Averaged Semantic Orientation

Discourse and RST

- Discourse is a linguistic level of analysis where the author represents his intentions
- Rhetorical Structure Theory is a descriptive theory proposed by Mann (1987) that explain the use of rhetorical relations in the text in order to keep the coherence.
- 26 relations
- Each relation links two spans of text in terms of the intentions desired by the author at the discourse level.
- Nucleus and Satellite

RST


1-2
Concession
1
2
It could have been a great movie. It could have been excellent, and to all the people who have forgotten about the older, greater movies before it, will think that as well. It does have beautiful scenery, some of the best since Lord of the Rings. The acting is well done, and I really liked the son of the leader of the Samurai. He was a likeable chap, and I hated to see him die.
But, other than all that, this movie is nothing more than hidden rip-offs.

Motivation

- The use of discourse structure to represent ideas is evident in text with sentiment.
- Sentiment classifiers can use such structure to better understand the text and emphasizes what is more important.

Objective

Research Questions

1. Can discourse knowledge help lexicon-based sentiment classifiers?
2. Which RST relations are more important for lexicon-based sentiment classification?
3. How to incorporate those important relations into the classifier algorithm?

SO-RST

(1) I like the product appearance.
(2) One day, it broke down.
(3) Hence, I believe it is a bad product.

I like (+4) the product appearance.
$SO = 4 \times w_{none}$

One day it broken (-2) down.
$SO = -2 \times w_{ResultNucleus}$

Hence, I believe it is a bad (-2) product.
$SO = -2 \times w_{ResultSatellite}$

SO-RST

(1) I like the product appearance.
(2) One day, it broke down.
(3) Hence, I believe it is a bad product.

I like (+4) the product appearance.
$SO = 4 \times w_{none}$

One day it broken (-2) down.
$SO = -2 \times w_{ResultNucleus}$

Hence, I believe it is a bad (-2) product.
$SO = -2 \times w_{ResultSatellite}$

SO-RST

(1) I like the product appearance.
(2) One day, it broke down.
(3) Hence, I believe it is a bad product.

I like (+4) the product appearance.
$SO = 4 \times w_{none}$

One day it broken (-2) down.
$SO = -2 \times w_{ResultNucleus}$

Hence, I believe it is a bad (-2) product.
$SO = -2 \times w_{ResultSatellite}$

SO-RST

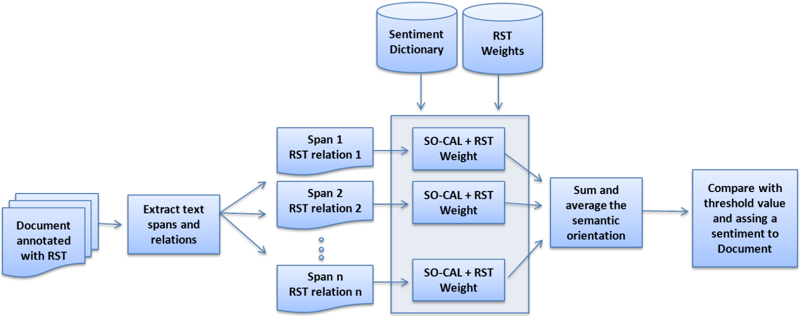
(1) I like the product appearance.
(2) One day, it broke down.
(3) Hence, I believe it is a bad product.

I like (+4) the product appearance.
$SO = 4 \times w_{none}$

One day it broken (-2) down.
$SO = -2 \times w_{ResultNucleus}$

Hence, I believe it is a bad (-2) product.
$SO = -2 \times w_{ResultSatellite}$

SO-RST Architecture



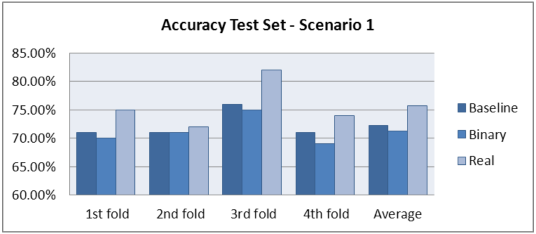
Experiments

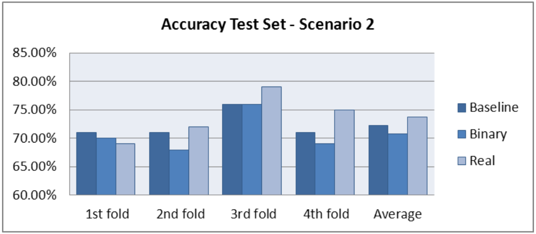
- Experiments:
  - Discover the best weights
  - Shallow RST Parser
- Corpus
  - SFU Review corpus (Taboada and Grieve, 2004)
  - 400 reviews in 8 categories
  - Website Epinions.com
  - RST annotation at sentence level
- Relations
  - Only representative relations (more than 30 instances)
  - 15 relations

Identifying the Best Weights

- Cross-fold-validation with 4 folds
- Training with genetic algorithm
  - 40 individuals in each generation
  - 100 generations
- Two scenarios:
  - Scenario 1: No nucleus and satellite distinction
  - Scenario 2: Different weights for nucleus and satellite
- Two weighting system:
  - binary
  - real values from 0 to 5

Results


Accuracy Test Set - Scenario 1
85.00%
80.00%
75.00%
70.00%
65.00%
60.00%
1st fold
2nd fold
3rd fold
4th fold
Average
Baseline
Binary
Real


Accuracy Test Set - Scenario 2
85.00%
80.00%
75.00%
70.00%
65.00%
60.00%
1st fold
2nd fold
3rd fold
4th fold
Average
Baseline
Binary
Real

Weights Learned for Scenario 1

Relation	1st Fold	2nd Fold	3rd Fold	4th Fold	Average
antithesis	1.35	0.34	0.15	1.81	**0.9125**
background	1.66	2.22	1.86	0.54	**1.57**
cause	1.77	0.69	0.93	0.11	**0.875**
circumstance	1.79	4.15	4.13	3.39	**3.365**
concession	0.2	0.34	0.16	0.09	**0.1975**
condition	2.61	2.89	3.58	3.83	**3.2275**
elaboration	4.02	4.49	4.53	4.53	**4.3925**
evaluation	2.61	3.48	2.25	1.79	**2.5325**
evidence	2.61	2.23	1.2	3.42	**2.365**
interpretation	3.57	4.32	2.25	4.19	**3.5825**
means	4.02	3.48	4.13	1.26	**3.2225**
preparation	1.35	0.69	0.93	0.54	**0.8775**
purpose	3.8	2.63	2.25	1.81	**2.6225**
result	1.35	0.96	0.93	0.54	**0.945**
unless	2.61	3.42	0.93	2.11	**2.2675**

Weights Learned for Scenario 1

Relation	1st Fold	2nd Fold	3rd Fold	4th Fold	Average
antithesis	1.35	0.34	0.15	1.81	**0.9125**
background	1.66	2.22	1.86	0.54	**1.57**
cause	1.77	0.69	0.93	0.11	**0.875**
circumstance	1.79	4.15	4.13	3.39	**3.365**
concession	0.2	0.34	0.16	0.09	**0.1975**
condition	2.61	2.89	3.58	3.83	**3.2275**
elaboration	4.02	4.49	4.53	4.53	**4.3925**
evaluation	2.61	3.48	2.25	1.79	**2.5325**
evidence	2.61	2.23	1.2	3.42	**2.365**
interpretation	3.57	4.32	2.25	4.19	**3.5825**
means	4.02	3.48	4.13	1.26	**3.2225**
preparation	1.35	0.69	0.93	0.54	**0.8775**
purpose	3.8	2.63	2.25	1.81	**2.6225**
result	1.35	0.96	0.93	0.54	**0.945**
unless	2.61	3.42	0.93	2.11	**2.2675**

Weights Learned for Scenario 1

Relation	1st Fold	2nd Fold	3rd Fold	4th Fold	Average
antithesis	1.35	0.34	0.15	1.81	**0.9125**
background	1.66	2.22	1.86	0.54	**1.57**
cause	1.77	0.69	0.93	0.11	**0.875**
circumstance	1.79	4.15	4.13	3.39	**3.365**
concession	0.2	0.34	0.16	0.09	**0.1975**
condition	2.61	2.89	3.58	3.83	**3.2275**
elaboration	4.02	4.49	4.53	4.53	**4.3925**
evaluation	2.61	3.48	2.25	1.79	**2.5325**
evidence	2.61	2.23	1.2	3.42	**2.365**
interpretation	3.57	4.32	2.25	4.19	**3.5825**
means	4.02	3.48	4.13	1.26	**3.2225**
preparation	1.35	0.69	0.93	0.54	**0.8775**
purpose	3.8	2.63	2.25	1.81	**2.6225**
result	1.35	0.96	0.93	0.54	**0.945**
unless	2.61	3.42	0.93	2.11	**2.2675**

Weights Learned for Scenario 1

Relation	1st Fold	2nd Fold	3rd Fold	4th Fold	Average
antithesis	1.35	0.34	0.15	1.81	**0.9125**
background	1.66	2.22	1.86	0.54	**1.57**
cause	1.77	0.69	0.93	0.11	**0.875**
circumstance	1.79	4.15	4.13	3.39	**3.365**
concession	0.2	0.34	0.16	0.09	**0.1975**
condition	2.61	2.89	3.58	3.83	**3.2275**
elaboration	4.02	4.49	4.53	4.53	**4.3925**
evaluation	2.61	3.48	2.25	1.79	**2.5325**
evidence	2.61	2.23	1.2	3.42	**2.365**
interpretation	3.57	4.32	2.25	4.19	**3.5825**
means	4.02	3.48	4.13	1.26	**3.2225**
preparation	1.35	0.69	0.93	0.54	**0.8775**
purpose	3.8	2.63	2.25	1.81	**2.6225**
result	1.35	0.96	0.93	0.54	**0.945**
unless	2.61	3.42	0.93	2.11	**2.2675**

Important Relations for Real Weights

- Important relations in Scenario 1

circumstance(↑)	condition(↑)
elaboration(↑)	evaluation(↑)
evidence(↑)	interpretation(↑)
means(↑)	result(↑)
concession(↓)	

- Important relations in Scenario 2

circumstance-nucleus(↑)	condition-satellite(↑)
evaluation-satellite(↑)	evidence-satellite(↑)
interpretation-nucleus(↑)	interpretation-satellite(↑)
means-nucleus(↑)	means-satellite(↑)
purpose-nucleus(↑)	purpose-satellite(↑)
evidence-nucleus(↓)	

Shallow RST Parser

- Previous Methodology relies on texts annotated with RST
- Explore how to incorporate the relations from the previous experiment
- Focus on discourse markers and word clues.

Crafting Rules

- Rules according Discourse Tagging Reference Manual (Carlson and Marcu, 2001) and the SFU Reviews Corpus.
- Intra-sentence discourse markers
- Rules provide RST segmentation

Crafting Rules

After its previous mayor committed suicide last year, an investigation disclosed that town officials regularly voted

rule = 40
relation = "CIRCUMSTANCE"
pattern = "(?P<S>after/.+?,/,)(?P<N>.+)$"

Circumstance Nucleus: [an investigation disclosed that town officials regularly voted]
Circumstance Satellite: [After its previous mayor committed suicide last year,]

Crafting Rules

After its previous mayor committed suicide last year, an investigation disclosed that town officials regularly voted

rule = *40*
relation = *"CIRCUMSTANCE"*
pattern = *"(?P<S>after/.+?,/,)(?P<N>.+)$"*

Circumstance Nucleus: [an investigation disclosed that town officials regularly voted]
Circumstance Satellite: [After its previous mayor committed suicide last year,]

Crafting Rules

After its previous mayor committed suicide last year, an investigation disclosed that town officials regularly voted

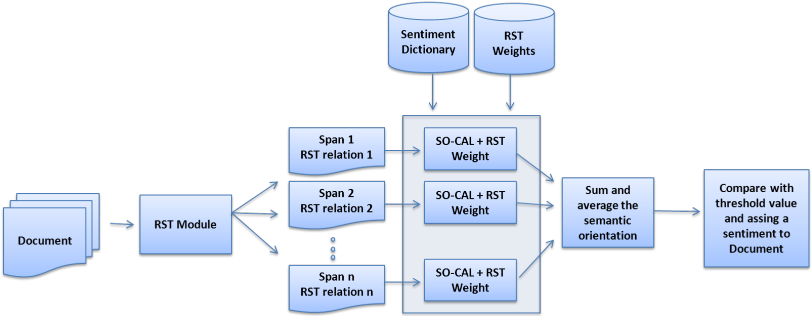
rule = 40
relation = “CIRCUMSTANCE”
pattern = “(?P<S>after/.+?,/,)(?P<N>.+)$”

Circumstance Nucleus: [an investigation disclosed that town officials regularly voted]
Circumstance Satellite: [After its previous mayor committed suicide last year,]

Rules matched by the SFU Reviews Corpus

Relation	Number of Rules	Number of Sentences Matched
Anthitesis	6	227
Background	2	1776
Cause	3	388
Circumstance	3	256
Concession	4	206
Condition	3	480
Elaboration	2	76
Means	1	134
Purpose	1	52
Unless	1	35
Total	**26**	**3630**

SO-RST Architecture with RST Module



Experiment

- Assigned the averaged weights learned from the previous experiment
- Two Corpora

- SFU Review corpus

Corpus	Accuracy
Baseline	74.81%
SO-RST - Scenario 1	**74.06%**
SO-RST - Scenario 2	**75.57%**

- Movie Reviews V2

Corpus	Accuracy
Baseline	71.90%
SO-RST - Scenario 1	**71.55%**
SO-RST - Scenario 2	**71.40%**

Discussion about the results

- The patterns crafted cover only a small set of the discourse phenomena which occurs in the text
- Some relations which received a high weight in the first experiment were not covered by the patterns or had few instances recognized
- The use of simple lexicon discourse markers may not be enough to improve sentiment classification

Conclusion

- This work demonstrated how to incorporate discourse knowledge in lexicon-based sentiment analysis
- The work presented the RST relations which most help in the process
- A proposal of shallow RST integration was discussed

Thank You