Thumbs up? Sentiment Classification using Machine Learning Techniques

By Ronald de Langen
Research seminar, 2016

change perspective

Radboud University

Objective of the paper

- Sentiment analysis
  - Is a review negative or positive

- Difference between Topic-based categorization and sentiment analysis

Proposal of the paper

- Use Naïve Bayes, Maximum Entropy and Support Vector Machine (SVM)

- Different features are necessary for Sentiment Analysis compared to Topic Categorization

Model

- 1400 movie reviews with as much positive as negative reviews

- NOT_ before words

- 3-fold cross validation

Features
unigrams
unigrams
unigrams+bigrams
bigrams
unigrams+POS
adjectives
top 2633 unigrams
unigrams+position

Evidence

- Baseline: 58, 64% humans. 69% from data

- Other work: 66%

	Features	# of features	frequency or presence?	NB	ME	SVM
(1)	unigrams	16165	freq.	**78.7**	N/A	72.8
(2)	unigrams	"	pres.	81.0	80.4	**82.9**
(3)	unigrams+bigrams	32330	pres.	80.6	80.8	**82.7**
(4)	bigrams	16165	pres.	77.3	**77.4**	77.1
(5)	unigrams+POS	16695	pres.	81.5	80.4	**81.9**
(6)	adjectives	2633	pres.	77.0	**77.7**	75.1
(7)	top 2633 unigrams	2633	pres.	80.3	81.0	**81.4**
(8)	unigrams+position	22430	pres.	81.0	80.1	**81.6**

Figure 3: Average three-fold cross-validation accuracies, in percent. Boldface: best performance for a given setting (row). Recall that our baseline results ranged from 50% to 69%.

Radboud University

- Based upon the idea of topic classification

- Sentence opinion in relation to a linguistic model:
  Hearst (1992)

- Sentiment analysis with a lexicon of two words:
  Turney (2002)

Impact

- Started off the field of sentiment analysis using machine learning

- 5000 times cites total, cited 50 times last 2 weeks

- David Zimbra, M. Ghiassi , Sean Lee, *"Brand-Related Twitter Sentiment Analysis Using Feature Engineering and the Dynamic Architecture for Artificial Neural Networks"*

Discussion

- Do you think bigrams not being effective for capturing context has to do with the data (eg. only 700 reviews per class) or with the models (eg. independence assumption of Naive Bayes)?

- Do you think that choosing from the IMDB review database was a bad choice, as these reviews can sometimes be of poor quality? Do you think it would perhaps have been a better idea to take reviews from newspapers or something more professional?

- Have you thought about other features sets that might be useful to increase the score?

- Why not cut out the top 10% of words, since words like "The" add very little information.

- Has there been any similar sentiment-classification research on data sets where thwarted expectation kind of reviews are less likely (product/restaurant reviews)? If so, what is the performance on these data sets?

Discussion

- I was wondering what the current state of the art in sentiment analysis is and whether the authors did extend their work as indicated at the end of section 7. Were they able to improve their results?

- What was the future research (after this paper) on the domain of sentiment analysis of documents and what is the current state of art method?

- The authors speculate that the difference in result between feature presence and feature frequency might indicate a difference between sentiment and topic categorization (section 6.2 *Feature frequency vs presence*). Were you able to find papers that reproduced these results and (if this is the case) did they provide a deeper analysis on why this might be the case?

Radboud University