Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences

Hong Yu   Vasileios Hatzivassiloglou

Columbia University, New York

24. November, 2008

Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP-03).

Presented by Lasse Soelberg

Layout

1. Introduction
   - Introduction
   - Goals of the Paper
2. The Approach
   - Document Classification
   - Finding Opinion Sentences
   - Identifying the Polarity
3. Experiments
   - Data
   - Evaluation
   - Results
4. Conclusion
   - Conclusion
   - Related Work
   - Article Evaluation

Introduction

Towards Answering Opinion Questions

- Question-answering systems.
- Easier to use factual statements.
- Extend to also use subjective opinion statements.

Introduction

Towards Answering Opinion Questions

- Question-answering systems.
- Easier to use factual statements.
- Extend to also use subjective opinion statements.

Simple Question

Who was elected as the new US President in 2008?

Introduction

Towards Answering Opinion Questions

- Question-answering systems.
- Easier to use factual statements.
- Extend to also use subjective opinion statements.

Simple Question

Who was elected as the new US President in 2008?

Complex Question

What has caused the current financial crisis?

Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences

Document Classification

Classifying articles as either subjective or objective

Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences

Document Classification

Classifying articles as either subjective or objective

Finding Opinion Sentences

In both subjective and objective articles

Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences

Document Classification

Classifying articles as either subjective or objective

Finding Opinion Sentences

In both subjective and objective articles

Identify the Polarity of Opinion Sentences

Determine if the opinions are positive or negative

Layout

1. Introduction
   - Introduction
   - Goals of the Paper
2. The Approach
   - Document Classification
   - Finding Opinion Sentences
   - Identifying the Polarity
3. Experiments
   - Data
   - Evaluation
   - Results
4. Conclusion
   - Conclusion
   - Related Work
   - Article Evaluation

Document Types

Training Sets

Articles from Wall Street Journal, which is annotated with document types.

Document Types

Training Sets

Articles from Wall Street Journal, which is annotated with document types.

Subjective Articles (Opinion)

- Editorials
- Letter to the Editor

Document Types

Training Sets

Articles from Wall Street Journal, which is annotated with document types.

Subjective Articles (Opinion)

- Editorials
- Letter to the Editor

Objective Articles (Fact)

- News
- Business

Classification

Naive Bayes

Calculating the likelihood that the document is either subjective or objective.

Classification

Naive Bayes

Calculating the likelihood that the document is either subjective or objective.

Bayes Rule

$$P(c|d) = \frac{P(c)P(d|c)}{P(d)}$$

where c is a class, d is a document and single words are used as feature.

Three Different Approaches

Rely on Expectation

Documents classified as opinions tends to have mostly opinion sentences, and documents classified as facts tends to have more factual sentences.

Three Different Approaches

Rely on Expectation

Documents classified as opinions tends to have mostly opinion sentences, and documents classified as facts tends to have more factual sentences.

The Three Approaches

- Similarity Approach
- Naive Bayes Classifier
- Multiple Naive Bayes Classifier

Similarity Approach

Hypothesis

Opinion sentences within a given topic will be more similar to other opinion sentences than to factual sentences.

Similarity Approach

Hypothesis

Opinion sentences within a given topic will be more similar to other opinion sentences than to factual sentences.

SimFinder

Measures sentence similarity based on shared words, phrases and WordNet synsets.

Variants

The score variant

- Select documents with the same topic as the sentence.
- Average the similarities with each sentence in the documents.
- Assign the sentence to the category with the highest average.

Variants

The score variant

- Select documents with the same topic as the sentence.
- Average the similarities with each sentence in the documents.
- Assign the sentence to the category with the highest average.

The frequency variant

Count how many of the sentences, for each category, that exceeds a predetermined threshold (set to 0.65).

Naive Bayes Classifier

Bayes Rule

$$P(c|d) = \frac{P(c)P(d|c)}{P(d)}$$

Naive Bayes Classifier

Bayes Rule

$$P(c|d) = \frac{P(c)P(d|c)}{P(d)}$$

Some of Features Used

- Words
- Bigrams
- Trigrams
- Parts of Speech
- Counts of positive and negative words
- Counts of the polarities of semantically oriented words
- Average semantic orientation score of the words

Multiple Naive Bayes Classifier

Problem

The designation of all sentences as opinions or facts is an approximation.

Multiple Naive Bayes Classifier

Problem

The designation of all sentences as opinions or facts is an approximation.

Solution

Use multiple Naive Bayes classifiers, each using a different subset of the features.

Multiple Naive Bayes Classifier

Problem

The designation of all sentences as opinions or facts is an approximation.

Solution

Use multiple Naive Bayes classifiers, each using a different subset of the features.

The Goal

Reduce the training set to the sentences most likely to be correctly labelled.

Multiple Naive Bayes Classifier

The Approach

- Train separate classifiers $C_1, C_2, \ldots, C_m$ given separate feature sets $F_1, F_2, \ldots, F_m$.

Multiple Naive Bayes Classifier

The Approach

- Train separate classifiers $C_1, C_2, ..., C_m$ given separate feature sets $F_1, F_2, ..., F_m$.
- Assume sentences inherit the document classification.

Multiple Naive Bayes Classifier

The Approach

- Train separate classifiers $C_1, C_2, \ldots, C_m$ given separate feature sets $F_1, F_2, \ldots, F_m$.
- Assume sentences inherit the document classification.
- Train C_1 on the entire training set, and use it to predict labels for the training set.

Multiple Naive Bayes Classifier

The Approach

- Train separate classifiers $C_1, C_2, \ldots, C_m$ given separate feature sets $F_1, F_2, \ldots, F_m$.
- Assume sentences inherit the document classification.
- Train C_1 on the entire training set, and use it to predict labels for the training set.
- Remove sentences with labels different from the assumption, and train C_2 on the remaining sentences.

Multiple Naive Bayes Classifier

The Approach

- Train separate classifiers $C_1, C_2, \ldots, C_m$ given separate feature sets $F_1, F_2, \ldots, F_m$.
- Assume sentences inherit the document classification.
- Train C_1 on the entire training set, and use it to predict labels for the training set.
- Remove sentences with labels different from the assumption, and train C_2 on the remaining sentences.
- Continue iteratively until no more sentences can be removed.

Multiple Naive Bayes Classifier

The Approach

- Train separate classifiers $C_1, C_2, \ldots, C_m$ given separate feature sets $F_1, F_2, \ldots, F_m$.
- Assume sentences inherit the document classification.
- Train C_1 on the entire training set, and use it to predict labels for the training set.
- Remove sentences with labels different from the assumption, and train C_2 on the remaining sentences.
- Continue iteratively until no more sentences can be removed.

Five Feature Sets

Starting with only words and adding in bigrams, trigrams, part of speech and polarity.

Identifying the Polarity of Opinion Sentences

What We Have

Sentences that are distinguished as either opinions or facts.

Identifying the Polarity of Opinion Sentences

What We Have

Sentences that are distinguished as either opinions or facts.

What We Want

Separate the opinion sentences into three classes

- Positive sentences.
- Negative Sentences.
- Neutral sentences.

Identifying the Polarity of Opinion Sentences

What We Have

Sentences that are distinguished as either opinions or facts.

What We Want

Separate the opinion sentences into three classes

- Positive sentences.
- Negative Sentences.
- Neutral sentences.

How We Do It

By the number and strength of semantically oriented words (either positive or negative) in the sentence.

Semantically Oriented Words

Hypothesis

Positive words co-occur more than expected by chance, and so do negative words.

Semantically Oriented Words

Hypothesis

Positive words co-occur more than expected by chance, and so do negative words.

Approach

Measure the words co-occurence with words from a known seed set of semantically oriented words.

Semantically Oriented Words

Log-likelihood ratio

$$L(W_i, POS_j) = log \left(\frac{\frac{Freq(W_i, POS_j, ADJp) + \epsilon}{Freq(W_{all}, POS_j, ADJp)}}{\frac{Freq(W_i, POS_j, ADJn) + \epsilon}{Freq(W_{all}, POS_j, ADJn)}} \right)$$

Where W_i is a word in the sentence, ADJ_p is positive seed word set, ADJ_n is negative seed word set, POS_j is part of speech collocation frequency ratio with ADJ_p and ADJ_n and ϵ is a smoothing constant (0.5).

Sentence Polarity Tagging

Determine the orientation of an opinion sentence

- Specify cutoffs t_p and t_n.
- Calculate the sentences average log-likelihood score.
- Positive sentences have average scores greater than t_p.
- Negative sentences have average scores lower than t_n.
- Neutral sentences have average scores between t_p and t_n.

Sentence Polarity Tagging

Determine the orientation of an opinion sentence

- Specify cutoffs t_p and t_n.
- Calculate the sentences average log-likelihood score.
- Positive sentences have average scores greater than t_p.
- Negative sentences have average scores lower than t_n.
- Neutral sentences have average scores between t_p and t_n.

Optimal t_p and t_n values

Are obtained from the training data via density estimation, using a small subset of hand-labeled sentences.

Seed Set

Seed words used

The seed words were subsets of 1.336 adjectives that were manually classified as either positive or negative.

Seed Set

Seed words used

The seed words were subsets of 1.336 adjectives that were manually classified as either positive or negative.

Seed Set Size

To see whether seed set sizes would influence the result, seed sets of 1, 20, 100 and over 600 positive and negative pairs of adjectives were used.

Layout

1. Introduction
   - Introduction
   - Goals of the Paper
2. The Approach
   - Document Classification
   - Finding Opinion Sentences
   - Identifying the Polarity
3. Experiments
   - Data
   - Evaluation
   - Results
4. Conclusion
   - Conclusion
   - Related Work
   - Article Evaluation

Data

Data Used

The data is from the TREC 8,9 and 11 collections, which consists of more than 1.7 million newswire articles from six different sources.

Data

Data Used

The data is from the TREC 8,9 and 11 collections, which consists of more than 1.7 million newswire articles from six different sources.

Wall Street journal

Some articles are marked with document type

- Editorial (2,877)
- Letter to Editor (1,695)
- Business (2,009)
- News (3,714)

2,000 articles from each type is randomly selected.

Evaluation Metrics

Recall

The fraction of the relevant documents that are retrieved.

$$recall = \frac{|\{relevant\ \ documents\} \cap \{retrieved\ \ documents\}|}{|\{relevant\ \ documents\}|}$$

Evaluation Metrics

Recall

The fraction of the relevant documents that are retrieved.

$$recall = \frac{|\{relevant\ \ documents\} \cap \{retrieved\ \ documents\}|}{|\{relevant\ \ documents\}|}$$

Precision

The fraction of the retrieved documents that are relevant.

$$precision = \frac{|\{relevant\ \ documents\} \cap \{retrieved\ \ documents\}|}{|\{retrieved\ \ documents\}|}$$

Evaluation Metrics

Recall

The fraction of the relevant documents that are retrieved.

$$recall = \frac{|\{relevant\ \ documents\} \cap \{retrieved\ \ documents\}|}{|\{relevant\ \ documents\}|}$$

Precision

The fraction of the retrieved documents that are relevant.

$$precision = \frac{|\{relevant\ \ documents\} \cap \{retrieved\ \ documents\}|}{|\{retrieved\ \ documents\}|}$$

F-measure

The weighted harmonic mean of recall and precision.

$$F = \frac{2 \cdot precision \cdot recall}{(precision + recall)}$$

Examples

Common Attributes

- Body of 1,000 documents.
- 100 relevant documents.

Examples

Common Attributes

- Body of 1,000 documents.
- 100 relevant documents.

Example 1

- 50 retrieved documents, all relevant.

Examples

Common Attributes

- Body of 1,000 documents.
- 100 relevant documents.

Example 1

- 50 retrieved documents, all relevant.
- Precision $= 1.00$

Examples

Common Attributes

- Body of 1,000 documents.
- 100 relevant documents.

Example 1

- 50 retrieved documents, all relevant.
- Precision = 1.00, Recall = 0.5

Examples

Common Attributes

- Body of 1,000 documents.
- 100 relevant documents.

Example 1

- 50 retrieved documents, all relevant.
- Precision = 1.00, Recall = 0.5, F-Measure = 0.67

Examples

Common Attributes

- Body of 1,000 documents.
- 100 relevant documents.

Example 1

- 50 retrieved documents, all relevant.
- Precision $=1.00$, Recall $=0.5$, F-Measure $=0.67$

Example 2

- Retrieves all documents.

Examples

Common Attributes

- Body of 1,000 documents.
- 100 relevant documents.

Example 1

- 50 retrieved documents, all relevant.
- Precision $= 1.00$, Recall $= 0.5$, F-Measure $= 0.67$

Example 2

- Retrieves all documents.
- Recall $= 1.00$

Examples

Common Attributes

- Body of 1,000 documents.
- 100 relevant documents.

Example 1

- 50 retrieved documents, all relevant.
- Precision $= 1.00$, Recall $= 0.5$, F-Measure $= 0.67$

Example 2

- Retrieves all documents.
- Recall $= 1.00$, precision $= 0.1$

Examples

Common Attributes

- Body of 1,000 documents.
- 100 relevant documents.

Example 1

- 50 retrieved documents, all relevant.
- Precision = 1.00, Recall = 0.5, F-Measure = 0.67

Example 2

- Retrieves all documents.
- Recall = 1.00, precision = 0.1, F-Measure = 0.18

Gold Standards

Document-level Standard

Already available from Wall Streel Journal.

- News and Business is mapped to facts.
- Editorial and Letter to the Editor is mapped to opinions.

Gold Standards

Document-level Standard

Already available from Wall Street Journal.

- News and Business is mapped to facts.
- Editorial and Letter to the Editor is mapped to opinions.

Sentence-level Standard

- There is no automated standard that can distinguish between facts and opinions, or between positive and negative opinions.

Gold Standards

Document-level Standard

Already available from Wall Street Journal.

- News and Business is mapped to facts.
- Editorial and Letter to the Editor is mapped to opinions.

Sentence-level Standard

- There is no automated standard that can distinguish between facts and opinions, or between positive and negative opinions.
- Human evaluators classify a set of sentences between facts and opinions and determine the type of opinions.

Topics and Articles

Topics

Four topics are chosen for the evaluation

- Gun control
- Illegal aliens
- Social security
- Welfare reform

Topics and Articles

Topics

Four topics are chosen for the evaluation

- Gun control
- Illegal aliens
- Social security
- Welfare reform

Articles

25 articles were randomly chosen for each topic from the TREC corpus. The articles were found using the Lucene search engine.

Sentences

Selection of Sentences

- Four sentences chosen from each document.
- The sentences were grouped into ten 50-sentence blocks.
- Each block shares ten sentences with the preceding and following block.

Sentences

Selection of Sentences

- Four sentences chosen from each document.
- The sentences were grouped into ten 50-sentence blocks.
- Each block shares ten sentences with the preceding and following block.

Standard A

The 300 sentences appearing once, and one judgement from the remaining 100 sentences.

Sentences

Selection of Sentences

- Four sentences chosen from each document.
- The sentences were grouped into ten 50-sentence blocks.
- Each block shares ten sentences with the preceding and following block.

Standard A

The 300 sentences appearing once, and one judgement from the remaining 100 sentences.

Standard B

The subset of the 100 sentences appearing twice, which were given identical labels.

Document Classification

Training

The classifier was trained on 4,000 articles from WSJ and evaluated on other 4,000 articles.

Document Classification

Training

The classifier was trained on 4,000 articles from WSJ and evaluated on other 4,000 articles.

The result

	F-measure
News vs. *Editorial*	0.96
News+Business vs. *Editorial+Letter*	0.97

Sentence Classification

Three Approaches

- Similarity approach
- Bayes classifier
- Multiple Bayes classifiers

Sentence Classification

Three Approaches

- Similarity approach
- Bayes classifier
- Multiple Bayes classifiers

The Similarity Approach

Variant	Class	Standard A	Standard B
Score	Fact	{0.61,0.34}	{1.00,0.27}
	Opinion	{0.30,0.49}	{0.16,0.64}
Frequency	Fact	{0.82,0.32}	{0.89,0.19}
	Opinion	{0.17,0.55}	{0.28,0.55}

{recall, precision}

Sentence Classification

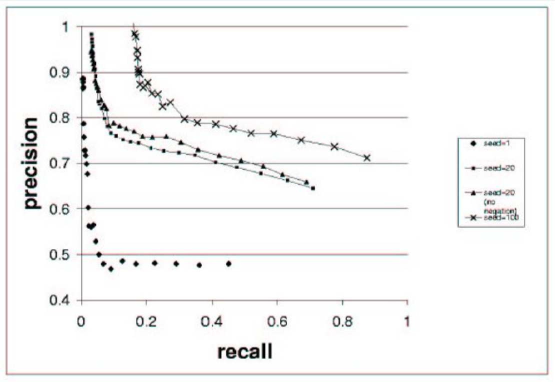
Bayes classifiers

Features	Class	Standard A		Standard B	
		Single	Multiple	Single	Multiple
Features from (Wiebe et al., 1999)	Fact	{0.03,0.38}	{0.03,0.38}	{0.06,1.00}	{0.06,1.00}
	Opinion	{0.97,0.69}	{0.97,0.69}	{1.00,0.80}	{1.00,0.80}
Words only	Fact	{0.14,0.39}	{0.12,0.42}	{0.28,0.42}	{0.28,0.45}
	Opinion	{0.90,0.69}	{0.92,0.69}	{0.90,0.82}	{0.91,0.83}
Words and Bigrams	Fact	{0.15,0.39}	{0.12,0.43}	{0.16,0.25}	{0.16,0.25}
	Opinion	{0.89,0.69}	{0.92,0.69}	{0.87,0.79}	{0.87,0.79}
Words, Bigrams, and Trigrams	Fact	{0.18,0.44}	{0.13,0.41}	{0.26,0.50}	{0.26,0.50}
	Opinion	{0.89,0.70}	{0.91,0.69}	{0.93,0.82}	{0.93,0.82}
Words, Bigrams, Trigrams, and Part-of-Speech	Fact	{0.17,0.42}	{0.13,0.40}	{0.18,0.49}	{0.27,0.44}
	Opinion	{0.89,0.70}	{0.91,0.69}	{0.92,0.70}	{0.85,0.84}
Words, Bigrams, Trigrams, Part-of-Speech, and Polarity	Fact	{0.15,0.43}	{0.13,0.42}	{0.44,0.50}	{0.44,0.53}
	Opinion	{0.91,0.69}	{0.92,0.70}	{0.88,0.86}	{0.91,0.86}

{recall, precision}

Sentence Classification

Seed Set Size



Polarity Classification

Accuracy of Sentence Polarity Tagging

Parts-of-speech Used	A	B
Adjectives	0.49	0.55
Adverbs	0.37	0.46
Nouns	0.54	0.52
Verbs	0.54	0.52
Adjectives and Adverbs	0.55	0.84
Adjectives, Adverbs, and Verbs	0.68	0.90
Adjectives, Adverbs, Nouns, and Verbs	0.62	0.74

Layout

1. Introduction
   - Introduction
   - Goals of the Paper
2. The Approach
   - Document Classification
   - Finding Opinion Sentences
   - Identifying the Polarity
3. Experiments
   - Data
   - Evaluation
   - Results
4. Conclusion
   - Conclusion
   - Related Work
   - Article Evaluation

Article Conclusion

Document Level

A fairly straightforward Bayesian classifier using lexical information can distinguish between mostly factual and opinion documents with very high precision and recall.

Article Conclusion

Document Level

A fairly straightforward Bayesian classifier using lexical information can distinguish between mostly factual and opinion documents with very high precision and recall.

Sentence Level

Three techniques were described for opinion/fact classification achieving up to 91% precision and recall on opinion sentences.

Article Conclusion

Document Level

A fairly straightforward Bayesian classifier using lexical information can distinguish between mostly factual and opinion documents with very high precision and recall.

Sentence Level

Three techniques were described for opinion/fact classification achieving up to 91% precision and recall on opinion sentences.

Polarity

Examined an automatic method for assigning polarity information (positive, negative or neutral), which assigns the correct polarity in 90% of the cases.

Related Work

Other work

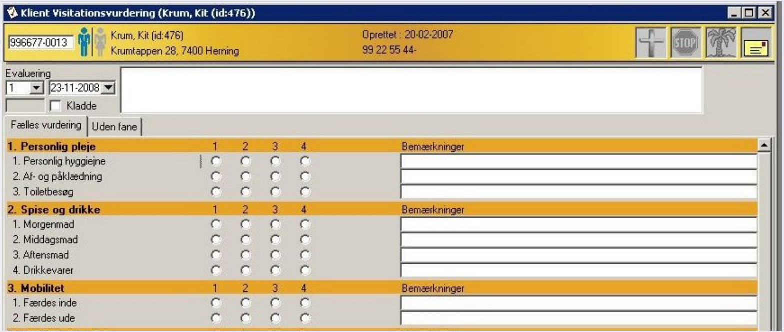
- There is a lot of research in the area of automated opinion detection.
- Prior works include SimFinder and classification of subjective words.
- Recent works includes Chinese web opinion mining and german news article.

Our Project - Herning Municipality

Citizens entering the homecare system gets a function evaluation, in order to establish their needs for help.

Relation to Our Project

Function Evaluation



Evaluation of the Article

The Good

- Good choice of titel.
- Good written description of the use of their methods.
- They keep a good flow through the article.

Evaluation of the Article

The Good

- Good choice of titel.
- Good written description of the use of their methods.
- They keep a good flow through the article.

The Not So Good

- No definition of recall and precision, not even a reference.
- SimFinder is presented as state-of-the-art. Made by one of the authors.