Thumbs Up? Sentiment Classification using Machine Learning Techniques

Pang, Lee, Vaithyanathan - EMNLP 2002

Slides by: Dor Cohen, Itai Gat

IE @ Technion

October 16, 2018

Code: http://github.com/dorcoh/sent-emnlp

Agenda

Introduction

Topic classification

- Recent (2002) works sort documents according to their **subject**
  - e.g., sports vs. politics
  - spam vs. no-spam

Topic classification

- Recent (2002) works sort documents according to their **subject**
  - e.g., sports vs. politics
  - spam vs. no-spam
- Yet crucial part of online posted articles is their **sentiment** (or overall opinion)
  - provide useful insights for readers automatically
  - e.g., product review is negative or positive

Topic classification

Current (2002) techniques for non-topic text categorization

- Source style with features as stylistic variation (Biber, 1988)
  - e.g., author, publisher (NY times vs. Daily News)
- Genre of text
  - e.g., editorial 'subjective' genre
- Is subjective language used?
- Does text contains opinion expressing?

Sentiment analysis

- This work: apply topic classification techniques on sentiment analysis
  - Q: What are our expected challenges?

Sentiment analysis

- This work: apply topic classification techniques on sentiment analysis
  - Q: What are our expected challenges?
  - A: Topics are identifiable by key words alone, while sentiment requires more **understanding**

Sentiment analysis

- This work: apply topic classification techniques on sentiment analysis
  - Q: What are our expected challenges?
  - A: Topics are identifiable by key words alone, while sentiment requires more **understanding**
- e.g., "How could anyone sit through this movie?"
  - Can you mark any negative word?

Sentiment analysis

- This work: apply topic classification techniques on sentiment analysis
  - Q: What are our expected challenges?
  - A: Topics are identifiable by key words alone, while sentiment requires more **understanding**
- e.g., "How could anyone sit through this movie?"
  - Can you mark any negative word?
- Previous (2002) techniques:
  - Semantic orientation of words (Turney and Littman, 2002)
  - Cognitive linguistic models (Sack, 1994)
  - Discriminant word lexicons (Tong, 2001)

Problem definition

Problem definition

Data: IMDB Movie Reviews

- Lucky for us: user rating provides us **supervised** learning
- Converted into three categories (or topics):
  - *Positive*, *negative*, (and *neutral* - not used)
- Avoid bias issues:
  - 20 reviews per author per sentiment
  - 752 negative vs 1301 positive
  - total of 144 reviewers
- More preprocess steps: TODO!

Human based sentiment classifiers

- In contrast to topics, detecting sentiment is easier for us (why?)

Human based sentiment classifiers

- In contrast to topics, detecting sentiment is easier for us (why?)
  - Topics can be related, while with opinions people tend to express strong feelings

Human based sentiment classifiers

- In contrast to topics, detecting sentiment is easier for us (why?)
  - Topics can be related, while with opinions people tend to express strong feelings
- Hypothesis: certain words indicate on sentiment type

Human based sentiment classifiers

- In contrast to topics, detecting sentiment is easier for us (why?)
  - Topics can be related, while with opinions people tend to express strong feelings
- Hypothesis: certain words indicate on sentiment type
- Test: decision procedure - count positive vs. negative words

Human based sentiment classifiers

- In contrast to topics, detecting sentiment is easier for us (why?)
  - Topics can be related, while with opinions people tend to express strong feelings
- Hypothesis: certain words indicate on sentiment type
- Test: decision procedure - count positive vs. negative words

Human	Proposed words	Accuracy	Ties [1]
1	positive (5): dazzling, brilliant.. negative (5): suck, terrible..	58%	75%
2	positive (11): gripping, spectacular.. negative (6): cliched, boring..	64%	39%

Table: Baseline results for human word lists, data is balanced (700 vs. 700)

[1] Documents percentage where sentiments rated equally

Human based sentiment classifiers

Should we worry about high rate of ties?

- Proposed words list is relatively short (effect is 0 vs. 0 ties)

Human based sentiment classifiers

Should we worry about high rate of ties?

- Proposed words list is relatively short (effect is 0 vs. 0 ties)
  - Not necessarily the reason for low accuracy!

Human based sentiment classifiers

Should we worry about high rate of ties?

- Proposed words list is relatively short (effect is 0 vs. 0 ties)
  - Not necessarily the reason for low accuracy!
- Authors propose their own list of words
  - Backed up with preliminary **data analysis** (including test set)

Human based sentiment classifiers

Should we worry about high rate of ties?

- Proposed words list is relatively short (effect is 0 vs. 0 ties)
  - Not necessarily the reason for low accuracy!
- Authors propose their own list of words
  - Backed up with preliminary **data analysis** (including test set)

Human	Proposed words	Accuracy	Ties
3+Stats	positive (7): love, wonderful.. negative (7): bad, worst, '?', '!',..	69%	16%

Table: Results where words (total 14) were chosen based on data statistics

Human based sentiment classifiers

Should we worry about high rate of ties?

- Proposed words list is relatively short (effect is 0 vs. 0 ties)
  - Not necessarily the reason for low accuracy!
- Authors propose their own list of words
  - Backed up with preliminary **data analysis** (including test set)

Human	Proposed words	Accuracy	Ties
3+Stats	positive (7): love, wonderful.. negative (7): bad, worst, '?', '!',..	69%	16%

Table: Results where words (total 14) were chosen based on data statistics

Reproduce data analysis (2018), feature occurrences binarized:

- $P(love|\widehat{d = positive}) = 0.1796$, $P(love|\widehat{d = negative}) = 0.1334$

Human based sentiment classifiers

Should we worry about high rate of ties?

- Proposed words list is relatively short (effect is 0 vs. 0 ties)
  - Not necessarily the reason for low accuracy!
- Authors propose their own list of words
  - Backed up with preliminary **data analysis** (including test set)

Human	Proposed words	Accuracy	Ties
3+Stats	positive (7): love, wonderful.. negative (7): bad, worst, '?', '!',..	69%	16%

Table: Results where words (total 14) were chosen based on data statistics

Reproduce data analysis (2018), feature occurrences binarized:

- $P(love|\widehat{d = positive}) = 0.1796$, $P(love|\widehat{d = negative}) = 0.1334$
- $P(worst|\widehat{d = positive}) = 0.0252$, $P(worst|\widehat{d = negative}) = 0.0937$

Methods

Bag of words

- d_1: "The audio quality really stinks."

Bag of words

- d_1: "The audio quality really stinks."
- Extract features:
  - Unigrams: $\{$the,audio,.. $\}$
  - Bigrams: $\{$the audio, audio quality, really stinks,.. $\}$
  - N-gram!

Bag of words

- d_1: "The audio quality really stinks."
- Extract features:
  - Unigrams: {the,audio,.. }
  - Bigrams: {the audio, audio quality, really stinks,.. }
  - N-gram!

Definition (Bag of words framework)

Let $\{f_1, ..f_m\}$ denote set of m features that can appear in document.
Let $n_i(d)$ be the number of times f_i occurs in document d.
Then each document d is represented by $\overrightarrow{d} := (n_1(d), ..., n_m(d))$.

Bag of words

- d_1: "The audio quality really stinks."
- Extract features:
  - Unigrams: {the,audio,.. }
  - Bigrams: {the audio, audio quality, really stinks,.. }
  - N-gram!

Definition (Bag of words framework)

Let $\{f_1, ..f_m\}$ denote set of m features that can appear in document.
Let $n_i(d)$ be the number of times f_i occurs in document d.
Then each document d is represented by $\overrightarrow{d} := (n_1(d), ..., n_m(d))$.

Unigram example

- $\overrightarrow{d_1} = (The : 1, audio : 1, quality : 1, really : 1, stinks : 1)$

Bag of words

- d_1: "The audio quality really stinks."
- Extract features:
  - Unigrams: {the,audio,.. }
  - Bigrams: {the audio, audio quality, really stinks,.. }
  - N-gram!

Definition (Bag of words framework)

Let $\{f_1, ..f_m\}$ denote set of m features that can appear in document.
Let $n_i(d)$ be the number of times f_i occurs in document d.
Then each document d is represented by $\overrightarrow{d} := (n_1(d), ..., n_m(d))$.

Unigram example

- $\overrightarrow{d_1} = (The : 1, audio : 1, quality : 1, really : 1, stinks : 1)$
- d_2: "Stinking quality, really stinks". *Question:* $\overrightarrow{d_2} =?$

Bag of words

- d_1: "The audio quality really stinks."
- Extract features:
  - Unigrams: {the,audio,.. }
  - Bigrams: {the audio, audio quality, really stinks,.. }
  - N-gram!

Definition (Bag of words framework)

Let $\{f_1, ..f_m\}$ denote set of m features that can appear in document.
Let $n_i(d)$ be the number of times f_i occurs in document d.
Then each document d is represented by $\overrightarrow{d} := (n_1(d), ..., n_m(d))$.

Unigram example

- $\overrightarrow{d_1} = (The : 1, audio : 1, quality : 1, really : 1, stinks : 1)$
- d_2: "Stinking quality, really stinks". *Question:* $\overrightarrow{d_2} =?$
- $\overrightarrow{d_2} = (0, 0, 1, 1, 2)$

Naive Bayes classifier

- Assign class which maximizes probability: $c^* = argmax_c P(c|d)$

Naive Bayes classifier

- Assign class which maximizes probability: $c^* = argmax_c P(c|d)$
- Recap:

Definition (Bayes theorem)

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

Naive Bayes classifier

- Assign class which maximizes probability: $c^* = argmax_c P(c|d)$
- Recap:

Definition (Bayes theorem)

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

- In practice to estimate $P(d|c)$, we **naively** assume f_i are conditionally independent.

Naive Bayes classifier

- Assign class which maximizes probability: $c^* = argmax_c P(c|d)$
- Recap:

Definition (Bayes theorem)

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

- In practice to estimate $P(d|c)$, we **naively** assume f_i are conditionally independent.
  - Hence $\widehat{P(d|c)} = \prod_{i=1}^{m} P(f_i|c)^{n_i(d)}$

Naive Bayes classifier

- Assign class which maximizes probability: $c^* = argmax_c P(c|d)$
- Recap:

Definition (Bayes theorem)

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

- In practice to estimate $P(d|c)$, we **naively** assume f_i are conditionally independent.
  - Hence $\widehat{P(d|c)} = \prod_{i=1}^{m} P(f_i|c)^{n_i(d)}$
- Q: Any numeric issues you can think about?

Naive Bayes classifier

- Assign class which maximizes probability: $c^* = argmax_c P(c|d)$
- Recap:

Definition (Bayes theorem)

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

- In practice to estimate $P(d|c)$, we **naively** assume f_i are conditionally independent.
  - Hence $\widehat{P(d|c)} = \prod_{i=1}^{m} P(f_i|c)^{n_i(d)}$
- Q: Any numeric issues you can think about?
- A_1: Some estimates are zero, can smooth (e.g., add-one smoothing)

Naive Bayes classifier

- Assign class which maximizes probability: $c^* = argmax_c P(c|d)$
- Recap:

Definition (Bayes theorem)

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

- In practice to estimate $P(d|c)$, we **naively** assume f_i are conditionally independent.
  - Hence $\widehat{P(d|c)} = \prod_{i=1}^{m} P(f_i|c)^{n_i(d)}$
- Q: Any numeric issues you can think about?
- A_1: Some estimates are zero, can smooth (e.g., add-one smoothing)
- A_2: Short documents vs. long documents (propose: tf-idf)

Maximum entropy classifier

aka logistic regression

Definition (MaxEnt estimator)

$P(c|d) = \frac{1}{Z(d)} exp(\sum_i \lambda_{i,c} F_{i,c}(d, c))$

Maximum entropy classifier

aka logistic regression

Definition (MaxEnt estimator)

$P(c|d) = \frac{1}{Z(d)} exp(\sum_i \lambda_{i,c} F_{i,c}(d,c))$

- $Z(d)$ - normalization function

Maximum entropy classifier

aka logistic regression

Definition (MaxEnt estimator)

$P(c|d) = \frac{1}{Z(d)} exp(\sum_i \lambda_{i,c} F_{i,c}(d, c))$

- $Z(d)$ - normalization function
- $F_{i,c}$ is a *feature/class function*

Maximum entropy classifier

aka logistic regression

Definition (MaxEnt estimator)

$P(c|d) = \frac{1}{Z(d)} exp(\sum_i \lambda_{i,c} F_{i,c}(d,c))$

- $Z(d)$ - normalization function
- $F_{i,c}$ is a *feature/class function*
  - Defined: $F_{i,c}(d,c') = 1$ if feature i **appears** on document d and its estimated class is c, o.w. $F_{i,c}(d,c) = 0$

Maximum entropy classifier

aka logistic regression

Definition (MaxEnt estimator)

$P(c|d) = \frac{1}{Z(d)} exp(\sum_i \lambda_{i,c} F_{i,c}(d,c))$

- $Z(d)$ - normalization function
- $F_{i,c}$ is a *feature/class function*
  - Defined: $F_{i,c}(d,c') = 1$ if feature i **appears** on document d and its estimated class is c, o.w. $F_{i,c}(d,c) = 0$
- $\lambda_{i,c}$ - feature-weight parameters (large values imply f_i is a strong indicator for class c)

Maximum entropy classifier

aka logistic regression

Definition (MaxEnt estimator)

$P(c|d) = \frac{1}{Z(d)} exp(\sum_i \lambda_{i,c} F_{i,c}(d,c))$

- $Z(d)$ - normalization function
- $F_{i,c}$ is a *feature/class function*
  - Defined: $F_{i,c}(d, c') = 1$ if feature i **appears** on document d and its estimated class is c, o.w. $F_{i,c}(d, c) = 0$
- $\lambda_{i,c}$ - feature-weight parameters (large values imply f_i is a strong indicator for class c)
- Aside: reminds neuron activation function

Maximum entropy classifier

aka logistic regression

Definition (MaxEnt estimator)

$P(c|d) = \frac{1}{Z(d)} exp(\sum_i \lambda_{i,c} F_{i,c}(d,c))$

- $Z(d)$ - normalization function
- $F_{i,c}$ is a *feature/class function*
  - Defined: $F_{i,c}(d,c') = 1$ if feature i **appears** on document d and its estimated class is c, o.w. $F_{i,c}(d,c) = 0$
- $\lambda_{i,c}$ - feature-weight parameters (large values imply f_i is a strong indicator for class c)
- Aside: reminds neuron activation function

Fit procedure

- Training data used to estimate distribution F
- λ's are set to maximize entropy of induced distribution

Support vector machines

- Goal: Find hyperplane w which separates classes with margin large as possible

Definition (SVM hyperplane)

Let $c_j \in \{1, -1\}$ be the class of document d_j then:

$$w := \sum_j \alpha_j c_j \overrightarrow{d_j},\ \alpha_j \geq 0$$

- α_j are obtained by solving dual optimization problem.

Support vector machines

- Goal: Find hyperplane w which separates classes with margin large as possible

Definition (SVM hyperplane)

Let $c_j \in \{1, -1\}$ be the class of document d_j then:

$$w := \sum_j \alpha_j c_j \overrightarrow{d_j}, \ \alpha_j \geq 0$$

- α_j are obtained by solving dual optimization problem.

Alternative fitting procedure:

Definition (Gradient descent)

$$w = w - \alpha * \frac{\partial L(X,w)}{\partial w} \text{ , Update till convergence}$$

Results

Results and discussion

ID	Features	count	freq/pres	NB	ME	SVM
1	unigrams	16165	freq	78.7%	NA	72.8%
2	unigrams	16165	pres	81.0%	80.4%	**82.9%**

Table: 3-fold average accuracies, unigrams appear at least 4 times on corpus.

Results and discussion

ID	Features	count	freq/pres	NB	ME	SVM
1	unigrams	16165	freq	78.7%	NA	72.8%
2	unigrams	16165	pres	81.0%	80.4%	**82.9%**

Table: 3-fold average accuracies, unigrams appear at least 4 times on corpus.

- Recall human baseline ranges between $50\% - 69\%$

Results and discussion

ID	Features	count	freq/pres	NB	ME	SVM
1	unigrams	16165	freq	78.7%	NA	72.8%
2	unigrams	16165	pres	81.0%	80.4%	**82.9%**

Table: 3-fold average accuracies, unigrams appear at least 4 times on corpus.

- Recall human baseline ranges between $50\% - 69\%$
- Topic-based classification reached $90\%+$ accuracy

Results and discussion

ID	Features	count	freq/pres	NB	ME	SVM
1	unigrams	16165	freq	78.7%	NA	72.8%
2	unigrams	16165	pres	81.0%	80.4%	**82.9%**

Table: 3-fold average accuracies, unigrams appear at least 4 times on corpus.

- Recall human baseline ranges between $50\% - 69\%$
- Topic-based classification reached $90\%+$ accuracy
  - Settings were multi-class
  - We conclude that sentiment analysis is harder

Results and discussion

ID	Features	count	freq/pres	NB	ME	SVM
1	unigrams	16165	freq	78.7%	NA	72.8%
2	unigrams	16165	pres	81.0%	80.4%	**82.9%**

Table: 3-fold average accuracies, unigrams appear at least 4 times on corpus.

- Recall human baseline ranges between $50\% - 69\%$
- Topic-based classification reached $90\%+$ accuracy
  - Settings were multi-class
  - We conclude that sentiment analysis is harder
- The frequency vs. presence of features seems to make the difference

Results and discussion

ID	Features	count	freq/pres	NB	ME	SVM
1	unigrams	16165	freq	78.7%	NA	72.8%
2	unigrams	16165	pres	81.0%	80.4%	**82.9%**

Table: 3-fold average accuracies, unigrams appear at least 4 times on corpus.

- Recall human baseline ranges between $50\% - 69\%$
- Topic-based classification reached $90\%+$ accuracy
  - Settings were multi-class
  - We conclude that sentiment analysis is harder
- The frequency vs. presence of features seems to make the difference
  - Hence from this point authors use presence (**binarized occurrences**)

Results and discussion

Using feature presence

ID	Features	count	NB	ME	SVM
2	unigrams	16165	81.0%	80.4%	**82.9%**
3	uni+bigrams	32330	80.6%	80.8	**82.7%**
4	bigrams	16165	77.3%	77.4%	77.1%

Results and discussion

Using feature presence

ID	Features	count	NB	ME	SVM
2	unigrams	16165	81.0%	80.4%	**82.9%**
3	uni+bigrams	32330	80.6%	80.8	**82.7%**
4	bigrams	16165	77.3%	77.4%	77.1%
5	unigrams+POS	16695	81.5%	80.4%	**81.9%**

Results and discussion

Using feature presence

ID	Features	count	NB	ME	SVM
2	unigrams	16165	81.0%	80.4%	**82.9%**
3	uni+bigrams	32330	80.6%	80.8	**82.7%**
4	bigrams	16165	77.3%	77.4%	77.1%
5	unigrams+POS	16695	81.5%	80.4%	**81.9%**
6	adjectives	2633	77.0%	77.7%	75.1%

Results and discussion

Using feature presence

ID	Features	count	NB	ME	SVM
2	unigrams	16165	81.0%	80.4%	**82.9%**
3	uni+bigrams	32330	80.6%	80.8	**82.7%**
4	bigrams	16165	77.3%	77.4%	77.1%
5	unigrams+POS	16695	81.5%	80.4%	**81.9%**
6	adjectives	2633	77.0%	77.7%	75.1%
7	top 2633 unigrams	2633	80.3%	81.0%	81.4%

Results and discussion

Using feature presence

ID	Features	count	NB	ME	SVM
2	unigrams	16165	81.0%	80.4%	**82.9%**
3	uni+bigrams	32330	80.6%	80.8	**82.7%**
4	bigrams	16165	77.3%	77.4%	77.1%
5	unigrams+POS	16695	81.5%	80.4%	**81.9%**
6	adjectives	2633	77.0%	77.7%	75.1%
7	top 2633 unigrams	2633	80.3%	81.0%	81.4%
8	unigram+position	22430	81.0%	80.1%	**81.6%**

Table: 3-fold average accuracies, bigrams appear at least 7 times on corpus.

Results and discussion

Using feature presence

ID	Features	count	NB	ME	SVM
2	unigrams	16165	81.0%	80.4%	**82.9%**
3	uni+bigrams	32330	80.6%	80.8	**82.7%**
4	bigrams	16165	77.3%	77.4%	77.1%
5	unigrams+POS	16695	81.5%	80.4%	**81.9%**
6	adjectives	2633	77.0%	77.7%	75.1%
7	top 2633 unigrams	2633	80.3%	81.0%	81.4%
8	unigram+position	22430	81.0%	80.1%	**81.6%**

Table: 3-fold average accuracies, bigrams appear at least 7 times on corpus.

- Adding bigrams doesn't improve results; Bigrams alone is worse

Results and discussion

Using feature presence

ID	Features	count	NB	ME	SVM
2	unigrams	16165	81.0%	80.4%	**82.9%**
3	uni+bigrams	32330	80.6%	80.8	**82.7%**
4	bigrams	16165	77.3%	77.4%	77.1%
5	unigrams+POS	16695	81.5%	80.4%	**81.9%**
6	adjectives	2633	77.0%	77.7%	75.1%
7	top 2633 unigrams	2633	80.3%	81.0%	81.4%
8	unigram+position	22430	81.0%	80.1%	**81.6%**

Table: 3-fold average accuracies, bigrams appear at least 7 times on corpus.

- Adding bigrams doesn't improve results; Bigrams alone is worse
- Part-of-speech: "I love this movie" vs. "This is a love story"

Results and discussion

Using feature presence

ID	Features	count	NB	ME	SVM
2	unigrams	16165	81.0%	80.4%	**82.9%**
3	uni+bigrams	32330	80.6%	80.8	**82.7%**
4	bigrams	16165	77.3%	77.4%	77.1%
5	unigrams+POS	16695	81.5%	80.4%	**81.9%**
6	adjectives	2633	77.0%	77.7%	75.1%
7	top 2633 unigrams	2633	80.3%	81.0%	81.4%
8	unigram+position	22430	81.0%	80.1%	**81.6%**

Table: 3-fold average accuracies, bigrams appear at least 7 times on corpus.

- Adding bigrams doesn't improve results; Bigrams alone is worse
- Part-of-speech: "I love this movie" vs. "This is a love story"
- Position based on dividing text into quarters.

Reproduce results

Reproduce results
(2018)

- We have tried to reproduce the experiment for the best settings reported

Features	count	NB	ME	SVM	MLP
unigrams	16165	81.0%	80.4%	**82.9%**	NA

Reproduce results
(2018)

- We have tried to reproduce the experiment for the best settings reported

Features	count	NB	ME	SVM	MLP
unigrams	16165	81.0%	80.4%	**82.9%**	NA
unigrams	16165	77.48%	**81.52%**	48.19%	**82.75%**

Table: Original vs. our results

Reproduce results
(2018)

- We have tried to reproduce the experiment for the best settings reported

Features	count	NB	ME	SVM	MLP
unigrams	16165	81.0%	80.4%	**82.9%**	NA
unigrams	16165	77.48%	**81.52%**	48.19%	**82.75%**

Table: Original vs. our results

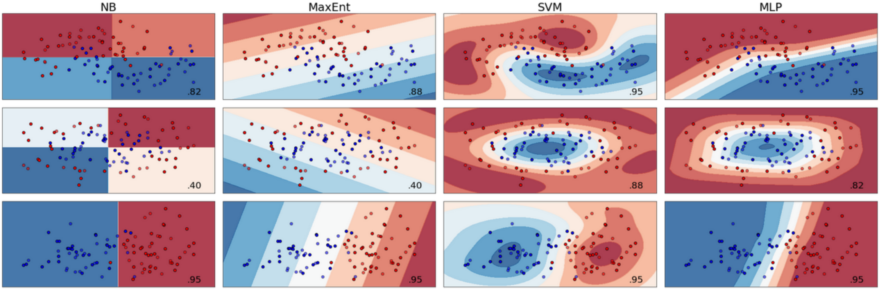
- **No tuning** was used at all; Plus not all described processing steps applied
- Notebook is available here(link)

Classifier comparison
(2018)

- Let's observe our classifiers decision boundaries for some toy datasets:

Classifier comparison
(2018)

- Let's observe our classifiers decision boundaries for some toy datasets:



Conclusions

Conclusions

- Unigrams presence setting achieves the best performance

Conclusions

- Unigrams presence setting achieves the best performance
- Applying selection algorithms for unigrams could improve performance

Conclusions

- Unigrams presence setting achieves the best performance
- Applying selection algorithms for unigrams could improve performance
- Contrarily, performance isn't comparable to topic classification

Conclusions

- Unigrams presence setting achieves the best performance
- Applying selection algorithms for unigrams could improve performance
- Contrarily, performance isn't comparable to topic classification

Review example

"This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up."

Conclusions

- Unigrams presence setting achieves the best performance
- Applying selection algorithms for unigrams could improve performance
- Contrarily, performance isn't comparable to topic classification

Review example

"This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up."

- Difficult for bag-of-words classifiers.

Conclusions

- Unigrams presence setting achieves the best performance
- Applying selection algorithms for unigrams could improve performance
- Contrarily, performance isn't comparable to topic classification

Review example

"This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up."

- Difficult for bag-of-words classifiers.
- Authors suggest determining the **focus** of each sentence, if is on/off topic.

Thank you for participating!
Questions?