COM3110/4115/6150: Text Processing

Sentiment Analysis: Approaches and Evaluation

Rob Gaizauskas

Department of Computer Science
University of Sheffield

Learning Outcomes

By the end of the SA sessions, you will be able to:

- Explain the relevance of the topic
- Differentiate between objective and subjective texts
- List the main elements in a sentiment analysis system
- Provide a critical summary of the main approaches for the problem
- Explain how sentiment analysis systems are evaluated.

Overview

- Definition of the problem of sentiment analysis
- **Approaches to sentiment analysis**
- **Evaluation of sentiment analysis approaches**

Two approaches to SA

- Lexicon-based
  - Binary
  - Gradable
- **Corpus-based (machine learning)**

A corpus-based approach to SA - Machine Learning

Naive Bayes classifier: estimate the probability of each class given a text:

- Compute the posterior probability (Bayes rule) of each class c_i for text segment T

$$P(c_i|T) = \frac{P(T|c_i)P(c_i)}{P(T)}$$

- Assumption of independence between features ("naive" assumption)

$$P(T|c_i) = P(t_1, t_2, ..., t_j|c_i) \approx \prod_{j=1}^{n} P(t_j|c_i)$$

where T is described by a number of attributes or features $t_1, ..., t_j$

I.e. joint probability of the features given the class is approximated by the product of the probabilities of each feature given the class.

A Naive Bayes classifier (ctd)

- **Likelihood**: product of probabilities of each feature value of segment occurring with class c_i

$$\prod_{j=1}^{n} P(t_j|c_i)$$

- **Prior**: probability of segment having class c_i

$$P(c_i)$$

- **Evidence**: product of probabilities of features of segment – constant term for all classes, so can be disregarded:

$$\prod_{j=1}^{n} P(t_j)$$

Final decision:

$$\underset{c_i}{\operatorname{argmax}} \prod_{j=1}^{n} P(t_j|c_i)P(c_i) \quad (= \underset{c_i}{\operatorname{argmax}}\, P(c_i) \prod_{j=1}^{n} P(t_j|c_i))$$

A corpus-based approach to SA - Machine Learning

A Naive Bayes classifier - a worked out example

- Corpus of movie reviews: 7 examples for **training**

Doc	Words	Class
1	Great movie, excellent plot, renowned actors	Positive
2	I had not seen a fantastic plot like this in good 5 years. Amazing!!!	Positive
3	Lovely plot, amazing cast, somehow I am in love with the bad guy	Positive
4	Bad movie with great cast, but very poor plot and unimaginative ending	Negative
5	I hate this film, it has nothing original	Negative
6	Great movie, but not...	Negative
7	Very bad movie, I have no words to express how I dislike it	Negative

A corpus-based approach to SA - Machine Learning

A Naive Bayes classifier - a worked out example (ctd)

- **Features**: adjectives (bag-of-words)

Doc	Words	Class
1	Great movie, excellent plot, renowned actors	Positive
2	I had not seen a fantastic plot like this in good 5 years. amazing !!!	Positive
3	Lovely plot, amazing cast, somehow I am in love with the bad guy	Positive
4	Bad movie with great cast, but very poor plot and unimaginative ending	Negative
5	I hate this film, it has nothing original. Really bad	Negative
6	Great movie, but not...	Negative
7	Very bad movie, I have no words to express how I dislike it	Negative

A corpus-based approach to SA - Machine Learning

Relative frequency in corpus is the simplest approach to estimating probabilities:

Priors:

$$P(positive) = count(positive)/N = 3/7 = 0.43$$

$$P(negative) = count(negative)/N = 4/7 = 0.57$$

where N = total training examples

Assume standard pre-processing: tokenisation, lowercasing, punctuation removal (except special punctuation like !!!)

A corpus-based approach to SA - Machine Learning

Likelihoods:

$$P(t_j|c_i) = \frac{count(t_j, c_i)}{count(c_i)}$$

Count word t_j in class c_i / total words in that class

$$
\begin{aligned}
P(amazing|positive) &= 2/10 & P(amazing|negative) &= 0/8 \\
P(bad|positive) &= 1/10 & P(bad|negative) &= 3/8 \\
P(excellent|positive) &= 1/10 & P(excellent|negative) &= 0/8 \\
P(fantastic|positive) &= 1/10 & P(fantastic|negative) &= 0/8 \\
P(good|positive) &= 1/10 & P(good|negative) &= 0/8 \\
P(great|positive) &= 1/10 & P(great|negative) &= 2/8 \\
P(lovely|positive) &= 1/10 & P(lovely|negative) &= 0/8 \\
P(original|positive) &= 0/10 & P(original|negative) &= 1/8 \\
P(poor|positive) &= 0/10 & P(poor|negative) &= 1/8 \\
P(renowned|positive) &= 1/10 & P(renowned|negative) &= 0/8 \\
P(unimaginative|positive) &= 0/10 & P(unimaginative|negative) &= 1/8 \\
P(!!!|positive) &= 1/10 & P(!!!|negative) &= 0/8
\end{aligned}
$$

A corpus-based approach to SA - Machine Learning

- Relative frequencies for prior ($P(c_i)$) and likelihood ($P(t_j|c_i)$) make the **model** in a Naive Bayes classifier.
- At decision (test) time, given a new segment to classify, this model is applied to find the most likely class for the segment:

$$\underset{c_i}{\operatorname{argmax}}\, P(c_i) \prod_{j=1}^{n} P(t_j|c_i)$$

A corpus-based approach to SA - Machine Learning

Given a new segment to classify (**test time**):

Doc	Words	Class
8	This was a fantastic story, good, lovely	???

Final decision

$$\underset{c_i}{\text{argmax}}\, P(c_i) \prod_{j=1}^{n} P(t_j|c_i)$$

$$P(positive) * P(fantastic|positive) * P(good|positive) * P(lovely|positive)$$

$$3/7 * 1/10 * 1/10 * 1/10 = 0.00043$$

$$P(negative) * P(fantastic|negative) * P(good|negative) * P(lovely|negative)$$

$$4/7 * 0/8 * 0/8 * 0/8 = 0$$

So: *sentiment = positive*

A corpus-based approach to SA - Machine Learning

What if the new segment to classify (**test time**) is:

Doc	Words	Class
10	Lovely plot, excellent cast, amazing everything	???

Final decision

$$\frac{P(positive) * P(lovely|positive) * P(excellent|positive) * P(amazing|positive) \qquad 3/7 * 1/10 * 1/10 * 1/10 = 0.00043}{P(negative) * P(lovely|negative) * P(excellent|negative) * P(amazing|negative) \qquad 4/7 * 0/8 * 0/8 * 0/8 = 0}$$

So: $sentiment = positive$

A corpus-based approach to SA - Machine Learning

Given a new segment to classify (**test time**):

Doc	Words	Class
9	Great plot, great cast, great everything	???

Final decision

$$P(positive) * P(great|positive) * P(great|positive) * P(great|positive)$$

$$3/7 * 1/10 * 1/10 * 1/10 = 0.00043$$

$$P(negative) * P(great|negative) * P(great|negative) * P(great|negative)$$

$$4/7 * 2/8 * 2/8 * 2/8 = 0.00893$$

So: $sentiment = negative$

A corpus-based approach to SA - Machine Learning

But if the new segment to classify (**test time**) is:

Doc	Words	Class
11	Boring movie, annoying plot, unimaginative ending	???

Final decision

$$P(positive) * P(boring|positive) * P(annoying|positive) * P(unimaginative|positive)$$

$$3/7 * 0/10 * 0/10 * 0/10 = 0$$

$$P(negative) * P(boring|negative) * P(annoying|negative) * P(unimaginative|negative)$$

$$4/7 * 0/8 * 0/8 * 1/8 = 0$$

So: $sentiment = ???$

A corpus-based approach to SA - Machine Learning

Add smoothing to feature counts (add 1 to every count). **Likelihoods** $=$

$$P(t_j|c_i) = \frac{count(t_j, c_i) + 1}{count(c_i) + |V|}$$

where $|V|$ is the number of distinct attributes in training (all classes) $=$ **12**

Doc	Words	Class
12	Boring movie, annoying plot, unimaginative ending	???

Final decision

$$P(positive) * P(boring|positive) * P(annoying|positive) * P(unimaginative|positive)$$

$$3/7 * ((0+1)/(10+12)) * ((0+1)/(10+12)) * ((0+1)/(10+12)) = 0.000040$$

$$P(negative) * P(boring|negative) * P(annoying|negative) * P(unimaginative|negative)$$

$$4/7 * ((0+1)/(8+12)) * ((0+1)/(8+12)) * ((1+1)/(8+12)) = 0.000143$$

So: *sentiment = negative*

A corpus-based approach to SA - Machine Learning

Given a trained classifier that classifies arbitrary segments of text we can use it to:

- Classify entire documents, e.g an entire review.
- Classify sentences in a document (perhaps just those identified as subjective) and then compute a classification of the document by aggregating the sentiments of individual sentences, according to some function.
- Classify sentences or phrases identified as discussing an aspect/feature of a target object (e.g. a sentence discussing battery life of a phone) and interpret the sentiment as the sentiment of opinion holder towards the specific aspect under discussion

Questions:

- Is this a good solution? Is it robust?
- What is the role of the prior?
- How can we improve this solution?
  - Other features? Are we missing out critical information?
  - Other algorithms?
- What about non-binary classification (e.g. 5-grades of sentiment)?

A corpus-based approach to SA - Machine Learning

Questions:

- Is this a good solution? Is it robust?
  - → It's simple and will work well if data is not sparse
- What is the role of the prior?
  - → Prior is very important esp. on biased cases
- How can we improve this solution?
  - ◇ Other features? Are we missing out critical information?
    - → Using all words (in Naive Bayes) works well in some tasks
    - → Finding subsets of words may help in other tasks
    - → Using only adjectives can be limiting. Verbs like hate, dislike; nouns like love; words for inversion like not; intensifiers like very
    - → Pre-built polarity lexicons can be helpful
    - → Negation is important
  - ◇ Other algorithms?
    - → MaxEnt & SVM tend to do better than Naive Bayes
- What about non-binary classification (e.g. 5-grades of sentiment)?
  - → 5-class ordinal classification or regression algorithms can be used

Comparative SA

Can contrast **direct opinions** versus more complex **comparative opinions**:

- Direct sentiment expressions on target objects
  - e.g., *"the picture quality of this camera is great."*
- Comparisons expressing similarities or differences between objects,
  - e.g., *"car x is cheaper than car y."*

Comparative SA – Linguistic Forms of Expression

Comparatives and superlatives in English are expressed in one of three ways:

- Short regulars: short "regular" adjectives/adverbs form comparatives by adding "er" and superlatives by adding "est": long, longer, longest, fast, faster, fastest, etc.
- Longer regulars: adjectives/adverbs longer than 2 syllables and not ending in "y" form comparatives and superlatives by adding the words "more" and "most"' before them: more expensive, most expensive.
- Irregulars: more, most, less, least, better, best, worse, worst, further/farther, furthest/farthest

Comparative SA – Comparative Relations

Bing Liu distinguishes 4 types of comparative relations:

1. Gradable Non-equal gradable: Relations of the type greater or less than. E.g.: *"lenses of camera A are better than those of camera B"*
2. Equative: Relations of the type equal to. E.g.: *"camera A and camera B both come in 7MP"*
3. Superlative: Relations of the type greater or less than all others. E.g.: *"camera A is the cheapest camera available in market"*
4. Non-gradable comparisons: Relations that compare aspects of two or more entities, but do not grade them. There are 3 main sub-types:
   - Entity A is similar to or different from entity B with regard to some of their shared aspects, e.g., *"Coke tastes differently from Pepsi."*
   - Entity A has aspect a1, and entity B has aspect a2, e.g., *"Desktop PCs use external speakers but laptops use internal speakers."*
   - Entity A has aspect a, but entity B does not, e.g., *"Phone-x has an earphone, but Phone-y does not."*

Comparative SA

Earlier quintuple-based model of opinion needs to be modified for comparative opinions:

Comparative SA Model: Given an opinionated document d, extract comparative opinions: (O_1, O_2, F, PO, h, t), where

- O_1 and O_2 are the object sets being compared based on their shared features F, PO is the preferred object set of the opinion holder h, and t is the time when the comparative opinion is expressed.
- No positive/negative opinions.

Example: Canons optics is better than those of Sony and Nikon. John, 2010

O_1	$\{Canon\}$
O_2	$\{Sony, Nikon\}$
F	$\{optics\}$
PO	$\{Canon\}$
h	$John$
t	2010

Evaluation

How do we quantify how well our Sentiment Analysis systems work?

- Create experimental datasets (aka test corpora): i.e., text segments that have been classified by humans, e.g. positive vs negative
- Compare (positive vs negative) system to human classifications
- Compute metrics like

$$\text{Accuracy} = \frac{\#\text{ correctly classified texts}}{\#\text{ texts}}$$

$$\text{Precision Pos} = \frac{\#\text{ texts correctly classified as positive}}{\#\text{ texts classified as positive}}$$

$$\text{Recall Pos} = \frac{\#\text{ texts correctly classified as positive}}{\#\text{ positive texts}}$$

$$\text{F-measure Pos} = \frac{2 * \text{Precision Pos} * \text{Recall Pos}}{\text{Precision Pos} + \text{Recall Pos}}$$

Same for negative class.

Baseline: most frequent class in the training set.

Conclusions

- Exciting topic, many applications, huge market for systems, particularly in focused domains.
- Promising results with simple techniques, but many interesting research challenges to be addressed for high accuracy.

- Trends:
  - Use subjectivity/polarity filtering in pre-processing of NLP tasks, like summarisation.
  - Joint-emotion analysis task: each fragment is classified with a number of emotions. E.g. SemEval-2007 Affective Text task `http://nlp.cs.swarthmore.edu/semeval/tasks/task14/summary.shtml`:
    Predict six emotions in a news headline: Anger, Disgust, Fear, Joy, Sadness and Surprise.
    "Amount" of emotion given by a score in [0,100], where 0 = total lack of emotion and 100 = maximum emotional load.

Conclusions

- Trends:
  - Robust systems for sentiment analysis on challenging types of data at feature-level (also called aspect-based sentiment analysis).
    SemEval-2014-16 tasks, e.g.
    `http://alt.qcri.org/semeval2014/task4/`:
    Aspect **term** extraction and polarity
    "I loved their fajitas" → {fajitas: positive}
    "I hated their fajitas, but their salads were great" → {fajitas: negative, salads: positive}
    "The fajitas are their first plate" → {fajitas: neutral}
    "The fajitas were great to taste, but not to see" → {fajitas: conflict}
    Aspect **category** extraction and polarity
    "The restaurant was too expensive" → {price: negative}
    "The restaurant was expensive, but the menu was great" → {price: negative, food: positive}

Conclusions

- Trends:
  - Sentiment analysis in social media like Twitter: SemEval-2015-16 tasks
    `http://alt.qcri.org/semeval2015/task10/`
  - New related tasks at SemEval 2017-2019 such as
    - affect detection
    - humour and irony detection
    - hate and abuse detection

    Mostly focussed on social media, esp. Twitter

Extra reading

Bing Liu and Lei Zhang (2012). A survey on opinion mining and sentiment analysis. Kluwer Academic Publishers:
http://www.cs.uic.edu/~lzhang3/paper/opinion_survey.pdf

Bing Liu (2012). Sentiment Analysis and Opinion Mining. Morgan and Claypool Publishers. Draft on line at: https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf

Article on SemEval in Wikipedia:
https://en.wikipedia.org/wiki/SemEval.