Roadmap

- Opinion Mining Problem
- **Document sentiment classification**
- Sentence subjectivity & sentiment classification
- Aspect-based sentiment analysis
- Aspect-based opinion summarization
- Opinion lexicon generation
- Mining comparative opinions
- Some other problems
- Opinion spam detection
- Utility or helpfulness of reviews
- Summary

Sentiment classification

- Classify a whole opinion document (e.g., a review) based on the overall sentiment of the opinion holder (Pang et al 2002; Turney 2002)
  - Classes: Positive, negative (possibly neutral)
  - Neutral or no opinion is hard. Most papers ignore it.
- An example review:
  - *"I bought an iPhone a few days ago. It is such a nice phone, although a little large. The touch screen is cool. The voice quality is clear too. I simply love it!"*
  - Classification: positive or negative?
- Perhaps the most widely studied problem.

A text classification task

- It is basically a text classification problem
- But different from topic-based text classification.
  - In topic-based text classification (e.g., computer, sport, science), topic words are important.
  - But in sentiment classification, opinion/sentiment words are more important, e.g., great, excellent, horrible, bad, worst, etc.
- Opinion/sentiment words
  - Words and phrases that express desired or undesired states or qualities.

Assumption and goal

- Assumption: The doc is written by a single person and express opinion/sentiment on a single entity.
- Goal: discover (_, _, *so*, _, _),

  where e, a, h, and t are ignored
- Reviews usually satisfy the assumption.
  - Almost all papers use reviews
  - Positive: 4 or 5 stars, negative: 1 or 2 stars
- Many forum postings and blogs do not
  - They can mention and compare multiple entities
  - Many such postings express no sentiments

Some Amazon reviews

248 of 263 people found the following review helpful:

This is one to get if you want 5MP, April 14, 2004

By Gadgester "No Time, No Money" (Mother Earth) - See all my reviews

TOP 100 REVIEWER

Amazon Verified Purchase (What's this?)

This review is from: **Canon PowerShot S500 5MP Digital Elph with 3x Optical Zoom (Electronics)**

The new Canon PowerShot S500 is a 5MP upgrade to the immensely popular S400 model, which was a 4MP digital camera. The S500 produces excellent images, is easy to use, and is compact enough to carry in a pocket. 3X optical zoom is standard on these cameras. Besides shooting still photos, you can record low-res video clips as well as audio clips, but don't expect high quality on either.

For a hundred bux less, you can get the 4MP S410 model which is otherwise identical to the S500. Should you go for this or the S410? I think for most consumers 4MP is plenty enough, with room for cropping and enlargements. 5MP is only necessary if you really crop a lot *and* plan to blow up the cropped images. The S410 strikes a great balance between pixel count and price -- it's a better value.

Help other customers find the most helpful reviews

Was this review helpful to you? Yes No

Report abuse | Permalink

Comment

41 of 41 people found the following review helpful:

E18 Error / problem with the lens, September 29, 2004

By Johnathan Parker (Springdale, AR USA) - See all my reviews

REAL NAME

This review is from: **Canon PowerShot S500 5MP Digital Elph with 3x Optical Zoom (Electronics)**

This is my second Canon digital elph camera. Both were great cameras. Recently upgraded to the S500. About 6 months later I get the dreaded E18 error. I searched the Internet and found numerous people having problems. When I determined the problem to be the lens not fully extending I decided to give it a tug. It clicked and the camera came on,

Unsupervised classification
(Turney, 2002)

- Data: reviews from epinions.com on automobiles, banks, movies, and travel destinations.
- The approach: Three steps
- Step 1:
  - Part-of-speech (POS) tagging
  - Extracting two consecutive words (two-word phrases) from reviews if their tags conform to some given patterns, e.g., (1) JJ, (2) NN.

Unsupervised classification

Tag	Description	Tag	Description
CC	Coordinating conjunction	PRP$	Possessive pronoun
CD	Cardinal number	RB	Adverb
DT	Determiner	RBR	Adverb, comparative
EX	Existential *there*	RBS	Adverb, superlative
FW	Foreign word	RP	Particle
IN	Preposition or subordinating conjunction	SYM	Symbol
JJ	Adjective	TO	*to*
JJR	Adjective, comparative	UH	Interjection
JJS	Adjective, superlative	VB	Verb, base form
LS	List item marker	VBD	Verb, past tense
MD	Modal	VBG	Verb, gerund or present participle
NN	Noun, singular or mass	VBN	Verb, past participle
NNS	Noun, plural	VBP	Verb, non-3rd person singular present
NNP	Proper noun, singular	VBZ	Verb, 3rd person singular present
NNPS	Proper noun, plural	WDT	Wh-determiner
PDT	Predeterminer	WP	Wh-pronoun
POS	Possessive ending	WP$	Possessive wh-pronoun
PRP	Personal pronoun	WRB	Wh-adverb

Patterns of POS tags

	First word	Second word	Third word (Not Extracted)
1.	JJ	NN or NNS	anything
2.	RB, RBR, or RBS	JJ	not NN nor NNS
3.	JJ	JJ	not NN nor NNS
4.	NN or NNS	JJ	not NN nor NNS
5.	RB, RBR, or RBS	VB, VBD, VBN, or VBG	anything

Two consecutive words are extracted if their POS tags confirm to any of the above pattern.

- Step 2: Estimate the sentiment orientation (SO) of the extracted phrases
  - Use Pointwise mutual information

$$PMI(word_1, word_2) = \log_2\left(\frac{P(word_1 \wedge word_2)}{P(word_1)P(word_2)}\right)$$

  - Semantic orientation (SO):

SO(phrase) = PMI(phrase, "excellent")
- PMI(phrase, "poor")

  - Using AltaVista near operator to do search to find the number of hits to compute PMI and SO.

- Step 3: Compute the average SO of all phrases
  - classify the review as positive if average SO is positive, negative otherwise.

- Final classification accuracy:
  - automobiles - 84%
  - banks - 80%
  - movies - 65.83
  - travel destinations - 70.53%

Supervised learning (Pang et al, 2002)

- Directly apply supervised learning techniques to classify reviews into positive and negative.
  - Like a text classification problem
- Three classification techniques were tried:
  - Naïve Bayes
  - Maximum entropy
  - Support vector machines
- Pre-processing:
  - Features: negation tag, unigram (single words), bigram, POS tag, position.

Supervised learning

- Training and test data
  - Movie reviews with star ratings
    - 4-5 stars as positive
    - 1-2 stars as negative
- Neutral is ignored.
- SVM gives the best classification accuracy based on balance training data
  - 83%
  - Features: unigrams (bag of individual words)

Features for supervised learning

- The problem has been studied by numerous researchers subsequently
  - Probably the most extensive studied problem
    - Including domain adaption and cross-lingual, etc.
- Key: feature engineering. A large set of features have been tried by researchers. E.g.,
  - Terms frequency and different IR weighting schemes
  - Part of speech (POS) tags
  - Opinion words and phrases
  - Negations
  - Syntactic dependency

A large number of related papers

- Bickerstaffe and Zukerman (2010) used a hierarchical multi-classifier considering inter-class similarity
- Burfoot, Bird and Baldwin (2011) sentiment-classified congressional floor debates
- Cui et al. (2006) evaluated some sentiment classification algorithms
- Das and Chen (2001) extracted market sentiment from stock message boards
- Dasgupta and Ng (2009) used semi-supervised learning
- Dave, Lawrence & Pennock (2003) designed a custom function for classification
- Gamon (2004) classified customer feedback data

A large number of related papers

- Goldberg and Zhu (2006) used semi-supervised learning.
- Kim, Li and Lee (2009) and Paltoglou and Thelwall (2010) studied different IR term weighting schemes
- Li, Lee, et al (2010) made use of different polarity shifting.
- Li, Huang, Zhou and Lee (2010) used personal (I, we) and impersonal (they, it, this product) sentences to help
- Maas et al (2011) used word vectors which are latent aspects of the words.
- Mullen and Collier (2004) used PMI, syntactic relations and other attributes with SVM.
- Nakagawa, Inui and Kurohashi (2010) used dependency relations and CRF.

A large number of related papers

- Ng, Dasgupta and Arifin (2006) identified reviews and classified sentiments of reviews
- Pang and Lee (2004) used minimum cuts
- Qiu, Zhang, Hu and Zhao (2009) proposed a lexicon-based and self-supervision approach
- Tong (2001) used a set of domain specific phrases
- Yessenalina, Choi and Cardie (2010) automatically generated annotator rationales to help classification
- Yessenalina, Yue and Cardie (2010) found subjective sentences and then used them for model building
- Zhou, Chen and Wang (2010) used semi-supervised and active learning

Review rating prediction

- Apart from classification of positive or negative sentiments,
  - research has also been done to predict the rating scores (e.g., 1–5 stars) of reviews (Pang and Lee, 2005; Liu and Seneff 2009; Qu, Ifrim and Weikum 2010; Long, Zhang and Zhu, 2010).
  - Training and testing are reviews with star ratings.
- Formulation: The problem is formulated as regression since the rating scores are ordinal.
- Again, feature engineering and model building.

Domain adaptation (transfer learning)

- Sentiment classification is sensitive to the domain of the training data.
  - A classifier trained using reviews from one domain often performs poorly in another domain.
    - words and even language constructs used in different domains for expressing opinions can be quite different.
    - same word in one domain may mean positive but negative in another, e.g., "*this vacuum cleaner really sucks.*"
- Existing research has used labeled data from one domain and unlabeled data from the target domain and general opinion words for learning (Aue and Gamon 2005; Blitzer et al 2007; Yang et al 2006; Pan et al 2010; Wu, Tan and Cheng 2009; Bollegala, Wei and Carroll 2011; He, Lin and Alani 2011).

Cross-lingual sentiment classification

- Useful in the following scenarios:
  - E.g., there are many English sentiment corpora, but for other languages (e.g. Chinese), the annotated sentiment corpora may be limited.
  - Utilizing English corpora for Chinese sentiment classification can relieve the labeling burden.
- Main approach: use available language corpora to train sentiment classifiers for the target language data. Machine translation is typically employed
  - (Banea et al 2008; Wan 2009; Wei and Pal 2010; Kim et al. 2010; Guo et al 2010; Mihalcea & Wiebe 2010; Boyd-Graber and Resnik 2010; Banea et al 2010; Duh, Fujino & Nagata 2011; Lu et al 2011)

Roadmap

- Opinion Mining Problem
- Document sentiment classification
- **Sentence subjectivity & sentiment classification**
- Aspect-based sentiment analysis
- Aspect-based opinion summarization
- Opinion lexicon generation
- Mining comparative opinions
- Some other problems
- Opinion spam detection
- Utility or helpfulness of reviews
- Summary

Subjectivity classification

- Document-level sentiment classification is too coarse for most applications.
- We now move to the sentence level.
- Much of the early work on sentence level analysis focuses on identifying subjective sentences.
- Subjectivity classification: classify a sentence into one of the two classes (Wiebe et al 1999)
  - Objective and subjective.
- Most techniques use supervised learning.
  - E.g., a naïve Bayesian classifier (Wiebe et al. 1999).

Sentence sentiment analysis

- Usually consist of two steps
  - Subjectivity classification
    - To identify subjective sentences
  - Sentiment classification of subjective sentences
    - Into two classes, positive and negative
- But bear in mind
  - Many objective sentences can imply sentiments
  - Many subjective sentences do not express positive or negative sentiments/opinions
    - E.g.,”I believe he went home yesterday.”

As an intermediate step

- We do not use the quintuple (*e*, *a*, *so*, *h*, *t*) to define the problem here because
  - sentence classification is an intermediate step.
- Knowing that some sentences have positive or negative opinions are not sufficient.
- However, it helps
  - filter out sentences with no opinions (mostly)
  - determine (to some extend) if sentiments about entities and their aspects are positive or negative.
    - But not enough

Assumption

- Assumption: Each sentence is written by a single person and expresses a single positive or negative opinion/sentiment.
- True for simple sentences, e.g.,
  - "I like this car"
- But not true for compound and "complex" sentences, e.g.,
  - "I like the picture quality but battery life sucks."
  - "Apple is doing very well in this lousy economy."

Subjectivity classification using patterns (Rilloff and Wiebe, 2003)

- A bootstrapping approach.
  - A high precision classifier is first used to automatically identify some subjective and objective sentences.
    - Two high precision (but low recall) classifiers are used,
      - a high precision subjective classifier
      - A high precision objective classifier
      - Based on manually collected lexical items, single words and n-grams, which are good subjective clues.
  - A set of patterns are then learned from these identified subjective and objective sentences.
    - Syntactic templates are provided to restrict the kinds of patterns to be discovered, e.g., <subj> passive-verb.
  - The learned patterns are then used to extract more subject and objective sentences (the process can be repeated).

Subjectivity and sentiment classification
(Yu and Hazivassiloglou, 2003)

- Subjective sentence identification: a few methods were tried, e.g.,
  - Sentence similarity.
  - Naïve Bayesian classification.
- Sentiment classification (positive, negative or neutral) (also called polarity): it uses a similar method to (Turney, 2002), but
  - with more seed words (rather than two) and based on log-likelihood ratio (LLR).
  - For classification of each word, it takes the average of LLR scores of words in the sentence and use cutoffs to decide positive, negative or neutral.

Segmentation and classification

- Since a single sentence may contain multiple opinions and subjective and factual clauses
- A study of automatic clause sentiment classification was presented in (Wilson et al 2004)
  - to classify clauses of every sentence by the *strength* of opinions being expressed in individual clauses, down to four levels
    - *neutral*, *low*, *medium*, and *high*
- Clause-level may not be sufficient
  - "Apple is doing very well in this lousy economy."

Some other related work

- Abdul-Mageed, Diab and Korayem (2011) carried out subjectivity and sentiment analysis of Arabic sentences
- Alm (2011) analyzed subjectivity research motivations, applications, characteristics, etc
- Barbosa and Feng (2010) and Davidov, Tsur and Rappoport (2010) performed Twitter subjectivity and sentiment classification using many features, hashtags, and smileys
- Eguchi and Lavrendo (2006) studied sentiment sentence retrieval
- Gamon et al. (2005) used semi-supervised learning
- Hassan, Qazvinian, Radev (2010) found attitude sentences
- Kim and Hovy (2004) summed up orientations of opinion words in a sentence (or within some word window).
- Hatzivassiloglou & Wiebe (2000) considered gradable adjectives

Some other related work

- Johansson and Moschitti (2011) extracted opinion expressions and sentiments
- Joshi and Penstein-Rose (2009) used dependency triples with "back-off" using POS tags rather than words
- Kim and Hovy (2006a) automatically identified pro and con reasons
- Kim and Hovy (2006b) Identified judgment opinions
- Kim and Hovy (2007) mined predictive opinions in election postings
- Kim, Li and Lee (2010) compared subjectivity analysis tools
- McDonald et al (2007) performed sentence to document sentiment classification
- Mukund and Srihari (2010) performed subjectivity classification with co-training

Some other related work

- Nasukawa and Yi (2003) captured favorability
- Nigam and Hurst (2005) classified subjective and topic sentences
- Tackstrom & McDonald (2011) performed sentence sentiment classification
- Wiebe et al (2004) learned subjective language
- Wiebe and Riloff (2005) used semi-supervised learning with a initial training set identified by some strong patterns
- Wiebe and Mihalcea (2006) studied word sense and subjectivity
- Wilson, Wiebe and Hwa (2006) recognized strong and weak opinion clauses
- Wilson et al. (2004, 2005) found strength of sentiments/opinions in clauses

Roadmap

- Opinion Mining Problem
- Document sentiment classification
- Sentence subjectivity & sentiment classification
- **Aspect-based sentiment analysis**
- Aspect-based opinion summarization
- Opinion lexicon generation
- Mining comparative opinions
- Some other problems
- Opinion spam detection
- Utility or helpfulness of reviews
- Summary

We need to go further

- Sentiment classification at both the document and sentence (or clause) levels are useful, but
  - They do not find what people liked and disliked.
- They do not identify the targets of opinions, i.e.,
  - Entities and their aspects
  - Without knowing targets, opinions are of limited use.
- We need to go to the entity and aspect level.
  - *Aspect-based opinion mining and summarization* (Hu and Liu 2004).
  - We thus need the full opinion definition.

Recall an opinion is a quintuple

- An *opinion* is a quintuple

  (e_j, a_{jk}, so_{ijkl}, h_i, t_l),

  where

  - e_j is a target entity.
  - a_{jk} is an aspect/feature of the entity e_j.
  - so_{ijkl} is the sentiment value of the opinion of the opinion holder h_i on feature a_{jk} of entity e_j at time t_l. so_{ijkl} is +ve, -ve, or neu, or a more granular rating.
  - h_i is an opinion holder.
  - t_l is the time when the opinion is expressed.

Aspect-based sentiment analysis

- Much of the research is based on online reviews
- For reviews, aspect-based sentiment analysis is easier because the entity (i.e., product name) is usually known
  - Reviewers simply express positive and negative opinions on different aspects of the entity.
- For blogs, forum discussions, etc., it is harder:
  - both entity and aspects of entity are unknown,
  - there may also be many comparisons, and
  - there is also a lot of irrelevant information.

Find entities (entity set expansion)

- Although similar, it is somewhat different from the traditional named entity recognition (NER).
- E.g., one wants to study opinions on phones
  - given Motorola and Nokia, find all phone brands and models in a corpus, e.g., Samsung, Moto,
- Formulation: Given a set *Q* of seed entities of class *C*, and a set *D* of candidate entities, we wish to determine which of the entities in *D* belong to *C*.
  - A classification problem. It needs a binary decision for each entity in *D* (belonging to *C* or not)
  - But it's often solved as a ranking problem

Some methods (Li, Zhang et al 2010, Zhang and Liu 2011)

- Distributional similarity: This is the traditional method used in NLP. It compares the surrounding text of candidates using cosine or PMI.
  - It performs poorly.
- PU learning: learning from positive and unlabeled examples.
  - S-EM algorithm (Liu et al. 2002)
- Bayesian Sets: We extended the method given in (Ghahramani and Heller, 2006).

Aspect extraction

- Goal: Given an opinion corpus, extract all aspects
- A frequency-based approach (Hu and Liu, 2004): nouns (NN) that are frequently talked about are likely to be true aspects (called frequent aspects) .
- Why the frequency based approach?
  - Different reviewers tell different stories (irrelevant)
  - When product aspects/features are discussed, the words they use converge.
  - They are the main aspects.
- Sequential/association pattern mining finds frequent nouns and noun phrases.

An example review

GREAT Camera., Jun 3, 2004
Reviewer: **jprice174** from Atlanta, Ga.

I did a lot of research last year before I bought this camera... It kinda hurt to leave behind my beloved nikon 35mm SLR, but I was going to Italy, and I needed something smaller, and digital.

The pictures coming out of this camera are amazing. The 'auto' feature takes great pictures most of the time. And with digital, you're not wasting film. …

….

Infrequent aspect extraction

- To improve recall due to loss of infrequent aspects. It uses opinions words to extract them
- Key idea: opinions have targets, i.e., opinion words are used to modify aspects and entities.
  - "The pictures are absolutely amazing."
  - "This is an amazing software."
- The modifying relation was approximated with the nearest noun to the opinion word.
- The idea was generalized to dependency in (Zhuang et al 2006) and double propagation in (Qiu et al 2009;2011).
  - It has been used in many papers and practical systems

Using part-of relationship and the Web (Popescu and Etzioni, 2005)

- Improved (Hu and Liu, 2004) by removing those frequent noun phrases that may not be aspects: better precision (a small drop in recall).
- It identifies part-of relationship
  - Each noun phrase is given a pointwise mutual information score between the phrase and **part discriminators** associated with the product class, e.g., a scanner class.
  - E.g., “of scanner”, “scanner has”, etc, which are used to find parts of scanners by searching on the Web:

$$PMI(a,d) = \frac{hits(a \wedge d)}{hits(a)hits(d)},$$