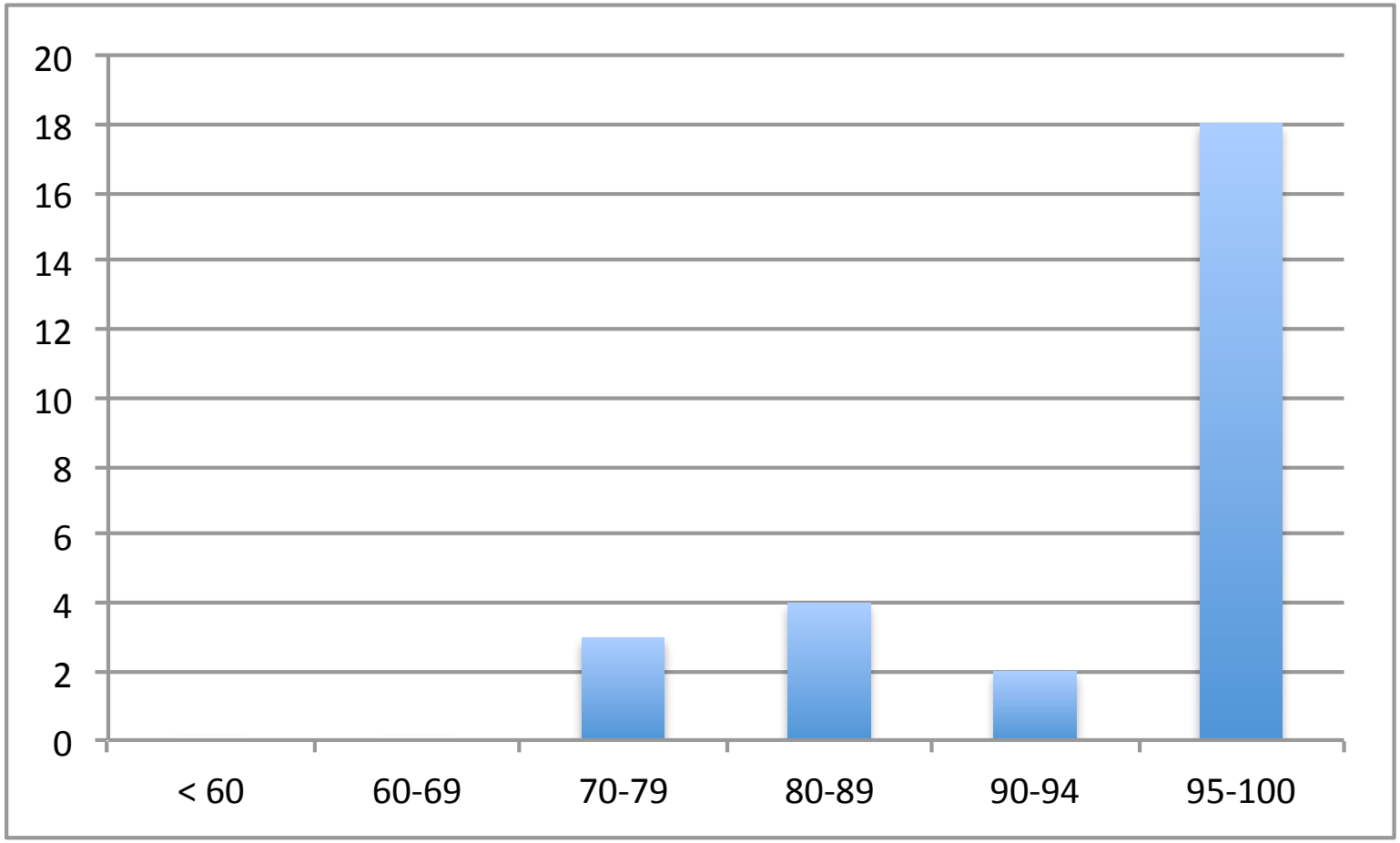
INLS 613 Text Mining
Annie T. Chen
November 4, 2013

Sentiment Analysis

Midterm



Road Map

- Sentiment Analysis I: Overview
  - Types of sentiment analysis
  - Features used in sentiment analysis
  - Opinion mining
  - Group activity

What is sentiment analysis?

Subjectivity analysis	Sentiment analysis ⇔ Opinion mining	Affect detection

- Detection of "private states" (Wiebe, 1994)
- Classification of text by whether it represents the opinions of the author or is merely stating fact
- Subjectivity resources http://www.cs.pitt.edu/mpqa/:
  - Opinion Finder
  - Political debate data
  - Subjectivity sense annotations
  - Subjectivity lexicon
  - MPQA Opinion Corpus
- Multilingual subjectivity analysis

What is sentiment analysis?

Subjectivity analysis	Sentiment analysis ⇔ Opinion mining	Affect detection

- Classification of reviews by polarity (positive, negative, neutral)
- Applications
  - Stock market trend prediction
  - Business intelligence
  - Poll prediction
  - Recommender systems
  - Summarization of product reviews
  - Detection of "flames" in email or other online communication
  - Advertisement placement

LEXALYTICS

Linguamatics

ebay

tripadvisor®

What is sentiment analysis?

Subjectivity analysis	Sentiment analysis ⇔ Opinion mining	Affect detection

- Automatic detection of emotions (e.g. anger, disgust, fear, guilt, joy, sadness, shame)
- Uses
  - Chat systems (e.g. Ma, Osherenko, Prendinger, & Ishizuka, 2005)
  - Discussion forums/message boards
  - Tracking changes in affect over time in online support groups (OSGs)

Motivation for Machine Learning

	Proposed word lists	Accuracy	Ties
Human 1	positive: *dazzling, brilliant, phenomenal, excellent, fantastic* negative: *suck, terrible, awful, unwatchable, hideous*	58%	75%
Human 2	positive: *gripping, mesmerizing, riveting, spectacular, cool, awesome, thrilling, badass, excellent, moving, exciting* negative: *bad, cliched, sucks, boring, stupid, slow*	64%	39%
Statistics-based	positive: *love, wonderful, best, great, superb, still, beautiful* negative: *bad, worst, stupid, waste, boring, ?, !*	69%	16%

Same as topic categorization?

- What are some characteristics of sentiment analysis that make it unique?

Same as topic categorization?

- Disjoint target classes
- Polarity Strength
- Negation
- Target resolution
- Sarcasm
- Attribute resolution/combination
- Topic-specific polarity of words
- Less term repetition
- ???

Features Used in Sentiment Analysis

- Term presence/frequency
- Parts of speech (POS)
  - POS tags
  - adjectives
  - adjectives + adverbs (Benamara et al., 2006)
- Punctuation
- Negation
  - I don’t like this movie.=> I don't NOT_like NOT_this NOT_movie.
- N-grams

N-grams

- N-gram: subsequence of N words from a given sequence of text or speech
- Unigrams: **cat, dog, said**
- Bigrams: **draw upon, drink in**
- Trigrams: **this holds true, squared off against**

- Example:
  - That looks like a killer movie!
  - What's your favorite serial killer movie?

Note: N-gram examples are from the **Corpus of Contemporary American English** (http://www.ngrams.info/intro.asp).

Use of N-Grams

- Why use n-grams?
  - Word sense disambiguation
  - Depends on the corpus
- Why not?
  - Dramatically increases the feature space!
- Strategies to limit the feature vector size:
  - Set a threshold for the frequency of n-grams
  - Incorporate n-grams that fit a certain pattern like adjective-noun or adverb verb (e.g. Chaovalit & Zhou, 2005)

Source: Boiy, Hens, Deschacht, & Moens (2007).

Other Features

- Hapax legomena (once-occurring words)
- Topic-tagged features
- Dependency-tree features
- http://nlp.stanford.edu:8080/parser/index.jsp

Approaches

- Supervised classification
- Domain adaptation
  - Pivot features
  - Non-pivot features
- Cross-lingual adaptation
- Topic + sentiment classification
- Subjectivity + sentiment classification
- Attribute + sentiment classification

Thumbs up? Sentiment Classification Using Machine Learning Techniques

- Corpus: 700 positive and 700 negative movies reviews
- Three-fold cross-validation
- Pre-processing:
  - Stemming? NO
  - Stopwords? NO
  - Negation? Added "NOT_" tag.
  - Frequency threshold? For unigrams, freq>=4

Data source: http://www.cs.cornell.edu/people/pabo/movie-review-data/

Thumbs up? – Results

	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**

What is an opinion?

- An opinion (or regular opinion) is a quintuple, **(E, A, O, H, T)**, where **E** is the **name of an entity**, **A** is an **attribute** of **E**, **O** is the **orientation of the opinion** about attribute **A** of entity **E**, **H** is the **opinion holder**, and **T** is the **time** when the opinion is expressed by **H**. The opinion orientation **O** can be positive, negative or neutral, or be expressed with different strength/intensity levels.

Source: Liu & Zhang (2012)

- "(1) I bought an iPhone a few days ago. (2) It was such a nice phone. (3) The touch screen was really cool. (4) The voice quality was clear too. (5) However, my mother was mad with me as I did not tell her before I bought it. (6) She also thought the phone was too expensive, and wanted me to return it to the shop . . . "

Source: Liu & Zhang (2012)

Attribute Sentiment Classification

- An object has different attributes:
  - Camera
    - Shutter speed
    - Lens
    - Picture quality
  - Notebook
    - CPU
    - RAM
    - Battery life

Attribute Discovery

- Find frequent nouns and phrases by using a part-of-speech (POS) tagger
- Find infrequent attributes by exploiting the relationships between attributes and opinion words

Source: Liu & Zhang (2012)

Group Activity

Questions to Answer

- What types of features would you use?
- What approach would you take?
- What are some of the attributes mentioned in the domain?
- What are some of the topics mentioned in the domain?

HINT: Take a look at concrete examples of reviews on the websites to decide which features might work!

Restaurants

Came here on recommendation from a local friend… Plenty of parking, but in a slightly sketchy part of town. Service was excellent, they were really attentive to our water and constantly asked us if wanted more …! There was a huge selection of …plates, too. The food was not amazing … but it wasn't completely terrible, either. A very solid restaurant, but kind of more expensive than I am used to paying ….

Nam Kang is one of the weirdest restaurants in Baltimore - and I mean that in the best way! The food is amazing, they stay open late, and somehow it feels like not that many people know about it - or else it would be packed all the time.

If you haven't been there before, they bring out about six complimentary sample bowls which makes for the perfect appetizer. I recently went there with a group of 6 people and they brought out double servings of the samples, which was really awesome…

The only thing about Nam Kang is that it is one of those places where you can ring up a huge bill really easily, so watch out for that.

Overall - it's one of my most favorite restaurants in Baltimore - highly recommended.

Hotels

tripadvisor®

I have stayed at numerous 5-star hotels. Royal Sonesta lives up to its 4-star rating.

Rooms are a tad bit dated. In-room furniture wasn't as modern to my liking and could have used some sprucing up. Grand stairway isn't as grand as pictures would lead you to believe, but very nice anyways. They gave the lobby lots of decor attention and it shows - very nice.

Staff was extremely friendly. Goes above and beyond to make your stay pleasant.

Rooms are roomy and accommodates playful children. Sheets were a tad bit dated.

Spa, lounge, pool, gym, and racquetball court adds to the hotels 4-star rating. Very happy about that. Extremely well landscaped courtyard.

Very close to Inner Harbor. Makes a nice walk to-from.

Overall a very great stay. Will stay again if in the area.

Cars

For those poeple who complain on the power: Take it to the freeway or somewhere safe and PRESS on the pedal!!! then come back and talk to me! When utilized Turbo adds 1/3 of your listed HP which means that you can smoke a 2012 Porche Boxter with your GTI (first hand experience). I have 49k on the DSG transmission with no issues and the car flies. Change the oil on time. 5k for me (including transmission oil every 40k or less). You can fit anything and everything in the back. I brought home a Queen size bed (in pieces but still). Love the lights because when you hit the turbo people need to know you're coming! Not that the engine won't let them know!! The car is a perfect combination of fun and convinient. I am not planning on selling this baby ever. Hopefully will never have to.

Fun to drive and very reliable. I have 50,000 miles on my 2010 VW GTI with no problems. Average mileage in town about 27-28 and on the highway 31. Automobile magazine's Car of the Year in 2010, and now I know why.

References

Boiy, E. , Hens, P. , Deschacht, K. , & Moens, M. F. (2007). Automatic sentiment analysis of on-line text. In *Proceedings of the 11th International Conference on Electronic Publishing*. Vienna, Austria.

Chaovalit, P., & Zhou, L. (2005) Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches. *HICSS 2005*.

Liu, B., Zhang, L. (2012). A Survey of opinion mining and sentiment analysis. In C. C. Aggarwal, C. Zhai (Eds.), *Mining Text Data*, 415-463.

Pang, B.; & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, Vol. 2, p.1-135.

Pang, B.; Lee, L.; & Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. *Proceedings of EMNLP*.

Somasundaran, S. & Wiebe, J. (2009). Recognizing Stances in Online Debates. *ACL-IJCNLP 2009.*

Wiebe, J. M. (1994). Tracking point of view in narrative. *Computational Linguistics, 20* (2): 233-287.

Aspect Sentiment Classification: Lexicon-Based Approach

- Mark opinion words
  - Opinion lexicon: http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html
- Handle opinion shifters ("valence shifters"): words or phrases that can change or shift opinion orientations
  - Negation words
  - Sarcasm
- Handle but-clauses: "the opinion orientation before and after but are opposite to each other if one cannot be determined"
- Aggregate opinions

Source: Liu & Zhang (2012)

Syntactic Rules for Finding Targets

DIRECT OBJECT Rule: dobj(opinion, target)
In words: The target is the direct object of the opinion
Example: I love$_{opinion1}$ Firefox$_{target1}$ and defended$_{opinion2}$ it$_{target2}$

NOMINAL SUBJECT Rule: nsubj(opinion, target)
In words: The target is the subject of the opinion
Example: IE$_{target}$ breaks$_{opinion}$ with everything.

ADJECTIVAL MODIFIER Rule: amod(target, opinion)
In words: The opinion is an adjectival modifier of the target
Example: The annoying$_{opinion}$ popup$_{target}$

PREPOSITIONAL OBJECT Rule: if prep(target1, IN) $\Rightarrow$ pobj(IN, target2)
In words: The prepositional object of a known target is also a target of the same opinion
Example: The annoying$_{opinion}$ popup$_{target1}$ in IE$_{target2}$ ("popup" and "IE" are targets of "annoying")

RECURSIVE MODIFIERS Rule: if conj(adj2, opinion$_{adj1}$) $\Rightarrow$ amod(target, adj2)
In words: If the opinion is an adjective (adj1) and it is conjoined with another adjective (adj2), then the opinion is tied to what adj2 modifies
Example: It is a powerful$_{opinion(adj1)}$ and easy$_{opinion(adj2)}$ application$_{target}$
("powerful" is attached to the target "application" via the adjective "easy")

Source: Somasundaran & Wiebe (2009)