

Textual sentiment summarization

Chris Potts

Linguist 287 / CS 424P: Extracting Social Meaning and Sentiment, Fall 2010

Oct 5

1 Overview

- Visualize summarization is covered in the slideshow ‘Visual sentiment summarization’.
- This handout looks briefly at two approaches to textual summarization:
 - The supervised approach of Beineke et al. (2003).
 - The largely unsupervised approach of Blair-Goldensohn et al. (2008) (whose lexicon construction method is covered in the ‘Sentiment lexicons’ handout).

2 Identifying summary sentences

Beineke et al. (2003) seek to extract summary sentences from reviews, using movie reviews from RottenTomatoes.com. The situation they face is very much like the (daunting) one in tab. 2.

2.1 Their approach

- Sentence-level classification:** The model predicts, for each sentence S in each review T , whether or not S is the summary sentence for T .
- Multinomial word features:** A number of words are over-represented in summary texts. These tend to be words having to do with emotion and with the general domain (movies). This suggests that word-level features will be useful. Beineke et al. (2003) retain the top 1,000 most frequent words for features.
- Binary location features:** Position of the sentence, position of the containing paragraph, position within the containing paragraph.

2.2 Results

Method	Features	Pct. Correct	Std. Error
Random	none	6.3%	–
Logist. Reg.	loc.	14.5%	0.3%
Naive Bayes	loc.; type	23.1%	0.5%
Logist. Reg.	loc.; type	25.8%	0.6%

Figure 4: Prediction match rate

Table 1: Beineke et al.’s (2003) results. It should be kept in mind that the chosen summary sentence could easily be just one of many that would serve as an adequate summary. This is bound to be a source of lowered effectiveness scores unless we hand-annotate the data with the full set of intuitively good summary sentences.

Helpful: 6 of 7 people found the following review helpful

Rating: 5.0 out of 5 stars

Summary: If you haven't already, introduce your child to Max.

Review: With "Where The Wild Things Are," Sendak set a new era of Children's story books - those that appeal to adults and children alike. At the time of its first publishing, "Where The Wild Things Are" brought groundbreaking graphics to children's publishing. More importantly, his story spoke to everyone - it evoked squeals of delight from children and heart-felt memories of childhood frustrations in adults.

 I owned a number of books from the 1960s list, which included, in order of preference, Dr. Suess' classic "Green Eggs and Ham," Richard Scarry's "Best Word Book Ever" and of course Sendak's "Where the Wild Things Are."

 Children simply can't help identifying with Sendak's hero Max, the type of kid who amuses himself by dressing in a wolf suit and chasing his hapless dog with a giant fork. Max's frazzled mother calls him a wild thing and sends him to bed without supper. "Wild thing?" As Max broods over the appellation, his imagination takes over. He soon finds himself on a boat to a strange land where he's named king of a herd of wild things all much larger, hairier, and meaner-looking than he ever pretended to be. The whole experience is fun, sure, who wouldn't like all the attention, all the swinging from trees and royal treatment? Still, the boat awaits, and eventually Max takes it back home, to where his mother (now calm) has thoughtfully kept his supper hot.

 I've read some recent children's books that stand up to the classics, but Pocahontas and Lion King don't quite make the grade. If you haven't already, introduce your child to Max. It will seem like they've known each other forever. If you do not have kids, but love artwork and the beauty of human spirit, then Where The Wild Things Are has a place on your shelf.
Where The Wild Things Are *is* a classic.

Helpful: 67 of 82 people found the following review helpful

Rating: 5.0 out of 5 stars

Summary: Thank you Dr. Fuhrman!

Review: I cannot thank Dr. Fuhrman enough! A year ago (at 27 years old)I was in bad shape. I have not only lost 50 pounds this last year, going from a size 16 to a 8, but have brought my high cholestrol/ high blood pressure down way down. I no longer have chest pains, acne, severe menstrual cycles, headaches, fatigue.. the list goes on. I had tried so many of the diets out there, only to fail or not lose any weight. I cannot believe the transformation in me since a year ago, it is simply amazing!..This will work for you if you commit to it! Dr. Fuhrman is very kind, I have consulted with him through email and on the telephone, he has gone over and above the call of duty in helping me to get well again..Thank you Dr. Fuhrman!

Movie Id: tt0122690

Rating: 8 of 10 stars

Username: ur1237981

Summary: This is how car chases SHOULD be done.

Review: Some movie car chases make you go 'Ooh' and 'Aah'. In this film they make you go 'Ouch' and cringe. Cars pelt down tiny back streets, skid through pedestrian-choked markets, and the many, MANY crashes are handled with bone-shattering realism. This is how car chases SHOULD be done.

 Alongside the gob-smacking chase scenes, the plot is smart and handled with style, disarmingly simple while requiring you to keep track of a large number of allies and villains. The ensemble cast are all excellent (dodgy Irish accents notwithstanding), and DeNiro and Reno underplay their roles to perfection.

 Perhaps the ending, deliberately low key in comparison to the film 's action sequences, is a touch too understated and reflective, but overall this is a great film. There really should be more like this.

Table 2: Two Amazon reviews (top two) and one IMDB review (bottom) that contain their own summary lines.

2.3 New data

2,260 user-supplied reviews from IMDB.com, in the following format (see tab. 3 for an example):

`https://www.stanford.edu/class/linguist287/restricted/data/imdb-summaries.zip`

Ln. 1: Movie Id

Ln. 2: Author's rating

Ln. 3: Author Id

Ln. 4: Summary line

Lns. 5-: Review, one sentence per line

Notes

- Exactly one of the review lines is guaranteed to match line 4, up to (i) case, and (ii) occurrences of `<p>` or `</p>`. Thus, matching should be done on downcased versions of the lines from which the p-tags have been removed.
- Paragraphs span multiple lines and are delimited by the p-tags. The boundaries were inferred from authors' use of doubled newlines (`
` tags).
- The sentence-chunking and tokenization were done with the NLTK tokenizers. One post-tokenization change is that ellipsis dots have been fixed up and treated as end-of-sentence punctuation.
- It seems that when reviewers do not include a summary line, the site uses the first sentence of the review. Since it's generally infeasible to determine whether a summary was supplied legitimately or automatically, no reviews in which the first line is the summary are included.

```
tt0229119
10
ur1954541
Crocs Rule !
<p>I just heard this morning that Steve Irwin died .
Crikey !
I remember watching The Crocodile Hunter on Animal Planet a few years back .
Good show .
There was a controversy two years ago when Steve was feeding a crocodile with one hand while holding his son
with the other .
He apologized for it and said that if he had to do it over again he 'd go surfing .
He died doing what he loved .
When he died , he got too close to a sting ray and was stung in the heart .
Crikey !
It 's hard to believe he 's gone .
Steve was working on a documentary about the Great Barrier Reef when he died .
His wife and children were off hiking in Tasmania when the news came and they caught a plane back to Queens-
land .
RIP , Steve .
CROCS RULE !</p>
```

Table 3: Example from the new data set. The summary sentence is the last one. It matches the summary line up to case and p tags.

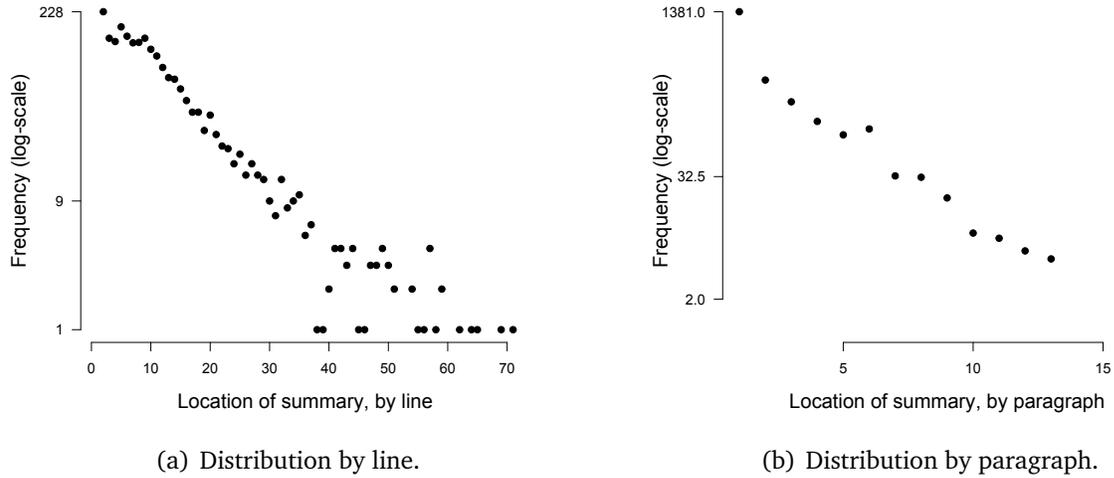


Figure 1: Distribution of summaries by position.

3 Natural language summarization

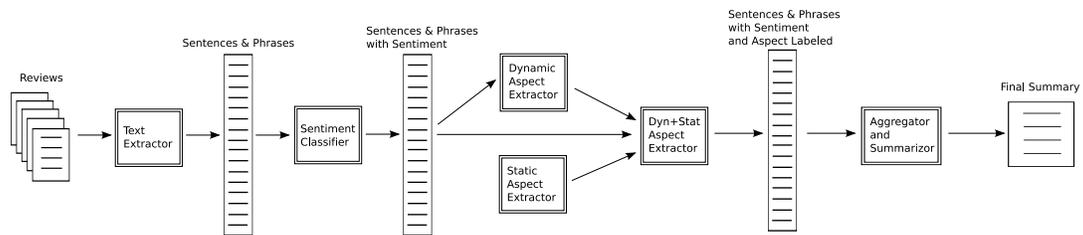


Figure 2: System Overview. Double boxed items are system components and single boxed items are text files (possibly marked-up with sentiment/aspect information).

Figure 2: Overview of the model of Blair-Goldensohn et al. (2008) (henceforth G08). The two major preliminary parts are the sentiment scoring (columns 1-5) and the aspect extractor (middle two columns). These come together to inform the summarizer.

3.1 WordNet-based sentiment lexicon

The continuous sentiment lexicon described in tab. 4, built using the method of G08, is available from the course’s code repository, in CSV format:

<http://code.google.com/p/linguist287/source/browse/#svn/trunk/lexicons>

For details on the construction, see the ‘Sentiment lexicons’ handout, section 5.2.

	Pos	Neg	Obj	Total
a (adj)	29	11	101	141
n (noun)	638	727	2,906	4,271
r (adv)	52	31	95	178
v (verb)	366	621	1,167	2,154
Total	1,085	1,390	4,269	6,744

(a) $\text{Inquirer}_{\text{WN}}$, the subset of Inquirer that is also in WordNet. For a fuller description, see section 3.5 of the ‘Sentiment lexicons’ handout.

	Pos	Neg	Obj	Total
a	166	217	21,155	21,538
n	2,710	2,380	113,944	119,034
r	116	161	4,204	4,481
v	1,904	1,512	8,115	11,531
Total:	4,896	4,270	147,418	156,584

(b) WordNet sentiment lexicon. Since Obj is basically a default, expanding to cover to whole of WordNet, the interesting gains are for Pos and Neg.

Table 4: Summarization lexicon consisting of a mapping of ⟨string, tag⟩ pairs to positive and negative reals, using the method of G08. This lexicon is bigger and more robust than the one constructed using the Micro-WNOp corpus as the seed-sets (section 5.2.3 of the ‘Sentiment lexicons’ handout).

3.2 Text classifier

Two scoring mechanisms provide the central features for the sentence-level classifier:

Definition 1 (Raw score). For a word-list x of length n and a sentiment dictionary s mapping words to their real-valued sentiment scores:

$$\text{raw-score}(x) \stackrel{\text{def}}{=} \sum_{w \in x} s(w)$$

For words immediately preceding a negation, $s(w) \leftarrow -s(w)$.

Definition 2 (Purity). For a word-list x of length n and a sentiment dictionary s mapping words to their real-valued sentiment scores:

$$\text{purity}(x) \stackrel{\text{def}}{=} \frac{\text{raw-score}(x)}{\sum_{w \in x} \text{abs}(s(w))}$$

These are used to score both sentences and whole reviews. The resulting scores are used as predictors in a MaxEnt model classifying sentences as positive, negative, or neutral. These basic predictors can be supplemented with user-supplied meta-data and perhaps with other features.

3.3 Aspect extraction

This module attempts to classify the sentences in reviews as being about certain properties (‘aspects’) of the item under discussion. The aspects might be known ahead of time, or they might be generated dynamically, say, by a user’s query. For aspects that are known ahead of time, the problem is another classification problem. For the dynamic aspects, G08 rely on properties of the texts themselves, including frequency and part-of-speech.

3.4 Summarization

If we set aside the aspects, focusing instead on the sentiment scores, then the summarizer picks the L -most non-neutral sentences from the review (for some value of L), with the ratio of positive-to-negative sentences given by the ratio of the review overall. The positive/negative measures are given by the model described in sec. 3.2 above. Including the aspects then means limiting attention to sentences relevant to that aspect when making selections. (One can imagine that it is also useful to bring the objective sentences into the picture to temper the overall summary appropriately.)

4 A few thoughts on extensions

- The OpenTable data described in the ‘Sentiment classification’ handout have star ratings for various aspects of dining experiences: food, service, ambiance, and noise. These could be used to inform both the sentiment and aspect modules of the G08 system, perhaps leading to better summary-sentence selection.
- Many Web corpora of product reviews include extensive information about the product itself. This could be used to inform aspect identification.
- The results discussed in sec. 2 suggest that certain locations in the text correlate with summary sentences. Such information could be brought into the G08 summarizer, allowing it to favor sentences in those positions. (One would want to think carefully about how this would interact with the raw-score and purity values.)

References

- Beineke, Philip; Trevor Hastie; Christopher D. Manning; and Shivakumar Vaithyanathan. 2003. Exploring sentiment summarization. In Qu Yan; James G. Shanahan; and Janyce Wiebe, eds., *Exploring Attitude and Affect in Text: Theories and Applications*, Papers from the AAI Spring Symposium, 4–7.
- Blair-Goldensohn, Sasha; Kerry Hannan; Ryan McDonald; Tyler Neylon; George A. Reis; and Jeff Reynar. 2008. Building a sentiment summarizer for local service reviews. In *WWW Workshop on NLP in the Information Explosion Era (NLPiX)*. Beijing, China.