Analyzing Appraisal Automatically

Maite Taboada and Jack Grieve

1. Classifying sentiment

- Classification of texts based on subjective content (=sentiment)
- Not impossible in principle: humans tend to agree on subjective content
- Test of agreement:
  - 3 judges; 30 texts; 5 categories (1: very negative; 5: very positive)
  - Measure agreement using kappa (Carletta 1996, Krippendorf 1980)
  - Complete agreement kappa: 0.529 (Judges A, B, C all assign 5 to same text)
  - Neighboring agreement kappa: 0.928 (Judge A says 5; Judge B says 4)
- Conclusion: we can try to build a system that agrees with human judges as much as they agree with each other

2. Method

Semantic orientation

- Words have semantic content or orientation (SO)
- SO for adjectives extracted automatically (Hatzivassiloglou and McKeown 1997)
- SO for adjectives or other words can be extracted using different methods:
  - Machine learning (Pang et al. 2002)
  - Pointwise Mutual Information (Turney 2002)

Appraisal

- Linguistic theory of subjectivity (Martin 2000, 2003; White 2003)
- Three subsytems: emotional (Affect), moral (Judgement), and aesthetic (Appreciation) opinions
- Also systems of Engagement (commitment of the author), and Amplitude (intensification or weakening of the opinion)

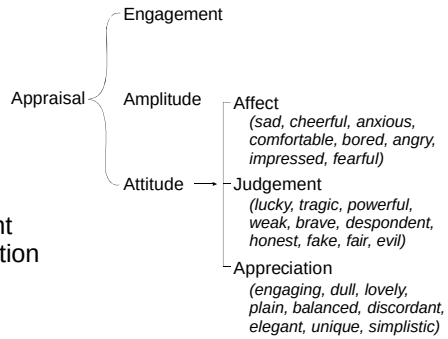

Fig. 1: Appraisal system

3. Texts

- 400 reviews from epinions.com
- 200 classified as 'recommended', 200 as 'not recommended'
- 8 subcategories:
  - movies
  - books
  - cars
  - cookware
  - phones
  - hotels
  - music
  - computers
- Output: SO value + Attitude values
  - SO: 2.54; Affect: 0.3; Judgement: 0.6; Appreciation: 0.1

4. Improving SO classification

- Texts divided in parts
- Prominence schema for text (Fig. 2)
- Adjectives weighed according to position in text
- Weighed SO values for a text averaged
- Split between negative and positive raised to 0.228

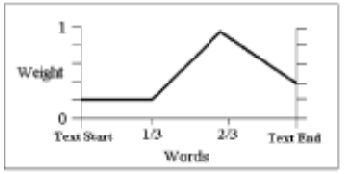

Fig. 2: Prominence schema

	Positive	Negative	Overall
Books	28%	88%	58%
Computers	52%	8%	66%
Hotels	92%	52%	72%
Music	48%	8%	64%
Phones	68%	68%	68%
Movies	32%	88%	6%
Cars	8%	6%	7%
Cookware	96%	28%	62%
All	62%	68%	65%

Table 1: SO accuracy

- Results (Table 1):
  - Compared to authors' recommendations
  - More accurate on **positive** reviews for: books, movies, music
  - More accurate on **negative** reviews for: phones, cars, cookware
  - SO questionable: *good* is used more often in negative than in positive reviews

5. Analyzing Appraisal

- Adjectives express Affect, Judgement, Appreciation depending on context
- Need to determine an adjective's *evaluative potential*: probability that it will be used to express one Appraisal (Attitude) type
- First: manually assigned values for 50 adjectives
- Then: values extracted using mutual information, based on collocation: *I was ADJ, he was ADJ, it was ADJ*
- Mutual information close to researchers' intuitions
- Appraisal calculated for 400 reviews (Table 2)
- Different Appraisal types according to review type

	Affect	Judgement	Appreciation
Books	23	27	50
Computers	20	24	56
Hotels	21	26	53
Music	22	28	50
Phones	17	22	61
Movies	23	26	51
Cars	20	23	57
Cookware	19	24	57

Table 2: Appraisal values per review type

6. Conclusions

- Adaptation of an existing SO method:
  - Classification based on adjectives
  - Position (prominence schema)
  - Extraction of Appraisal values
- Related project: Literary reputation
  - 5 authors, very popular in their time; not popular now
  - 5 authors, unknown in their time; popular now
  - Extract information about their reception then and now
- Future work:
  - Verbs, adverbs, noun+adjective
  - Negation
  - Rhetorical relations (Mann & Thompson 1988)
  - Other collocations for Appraisal

References

Carletta, J. 1996. Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics* 22 (2): 249-154.
Hatzivassiloglou, V., and McKeown, K. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of 35th ACL*., 174-181.
Krippendorf, K. 1980. *Content Analysis: An Introduction to Its Methology.* Beverly Hills, CA: Sage.
Martin, J.R. 2000. Beyond Exchange: Appraisal Systems in English. In Hunston, S., and Thompson, G., eds., *Evaluation in Text*. Oxford: OUP. 142-175.
Martin, J. 2003. Introduction, special issue on Appraisal. *Text* 23 (2):171–181.
Mann, W., and Thompson, S. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text* 8 (3): 243–281.
Pang, B., Lee, L., and Vaithyanathan, S. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proc. Conf. on Empirical Methods in NLP*. 79-86.
Turney, P. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proc. 40th ACL*., 417–424.
White, P.R.R. 2003. An Introductory Course in Appraisal Analysis. Appraisal website. (http://www.grammatics.com/appraisal/).

Maite Taboada and Jack Grieve
mtaboada@sfu.ca, jwgrieve@sfu.ca
www.sfu.ca/~mtaboada

Department of Linguistics - Simon Fraser University
Burnaby, B.C., V5A 1S6, Canada