Exploring the characteristics of opinion expressions for political opinion classification

Bei Yu
Stefan Kaufmann
Daniel Diermeier

Northwestern University

Agenda

- General-purpose political opinion classification
- Low classification accuracy - why is it hard?
- Characteristics of political opinion expression
  - Sentiment strength
  - Part-of-speech contributions
  - Interrelated categories: Opinion – Tone – Vote
- Conclusions

Political opinion classification

- Applications in e-Rulemaking
- Classification across styles of political language

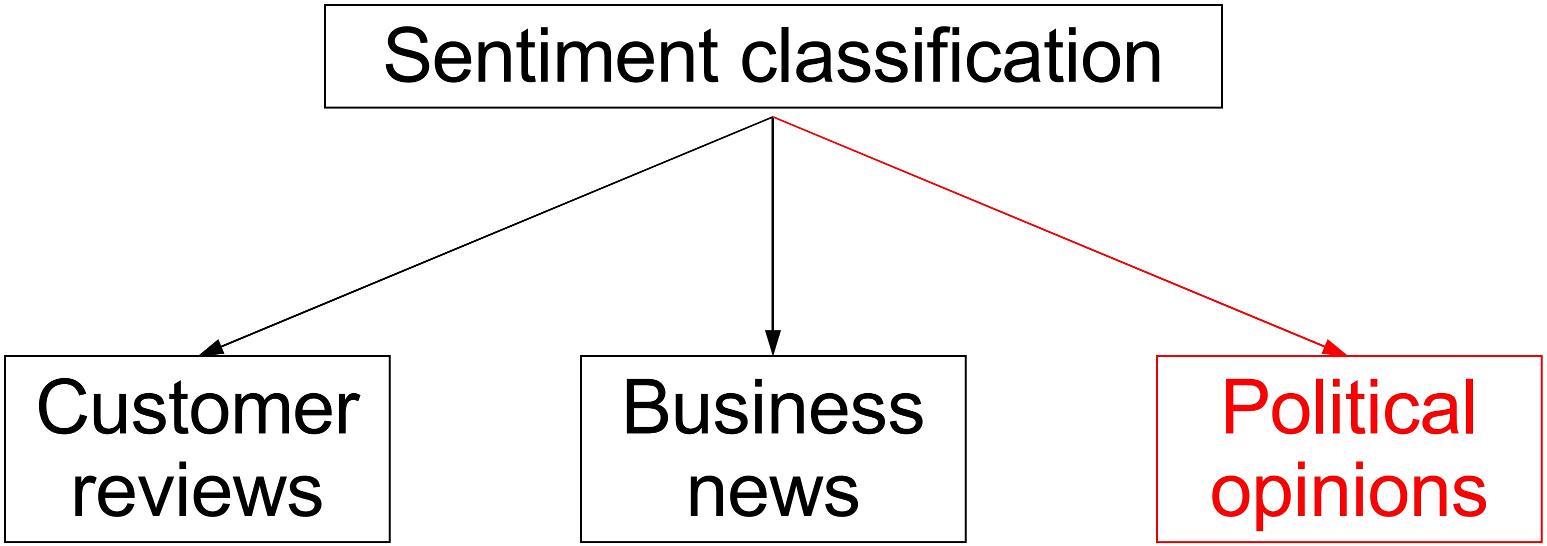
Online newsgroup discussions (Agrawal et al., 2003)

Congressional debates (Thomas et al.,2006)

Public email feedback on government policies (Kwon et al.,2006)

Common view

Political opinion classification
=
Sentiment classification in the political domain



Sentiment classification

Knowledge based

- General Inquirer (GI)
- Linguistic Inquiry and Word Count (LIWC)
- Dictionary of Affect in Language (DAL)

Supervised Machine Learning

- Support Vector Machines (SVM)
- Naïve Bayes (NB)
- Maximum Entropy (ME)

Classification across domains

Movie reviews

- Accuracy > *80%* (Pang and Lee, 2002)

Political opinion

- Newsgroups discussions: *worse than majority vote*
- Congressional debates: *~70%*
- EPA public email: *slightly better than baseline*

Question:

Why is political opinion classification hard?

Two experiments:

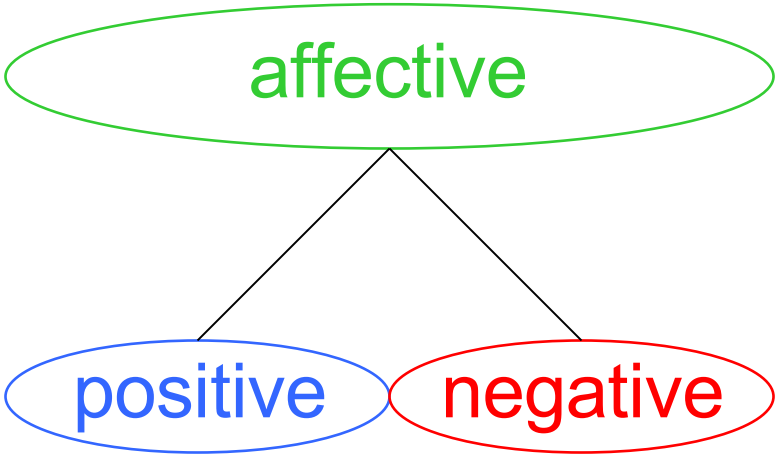
- Sentiment strength
- Part-of-speech contributions

Sentiment strength

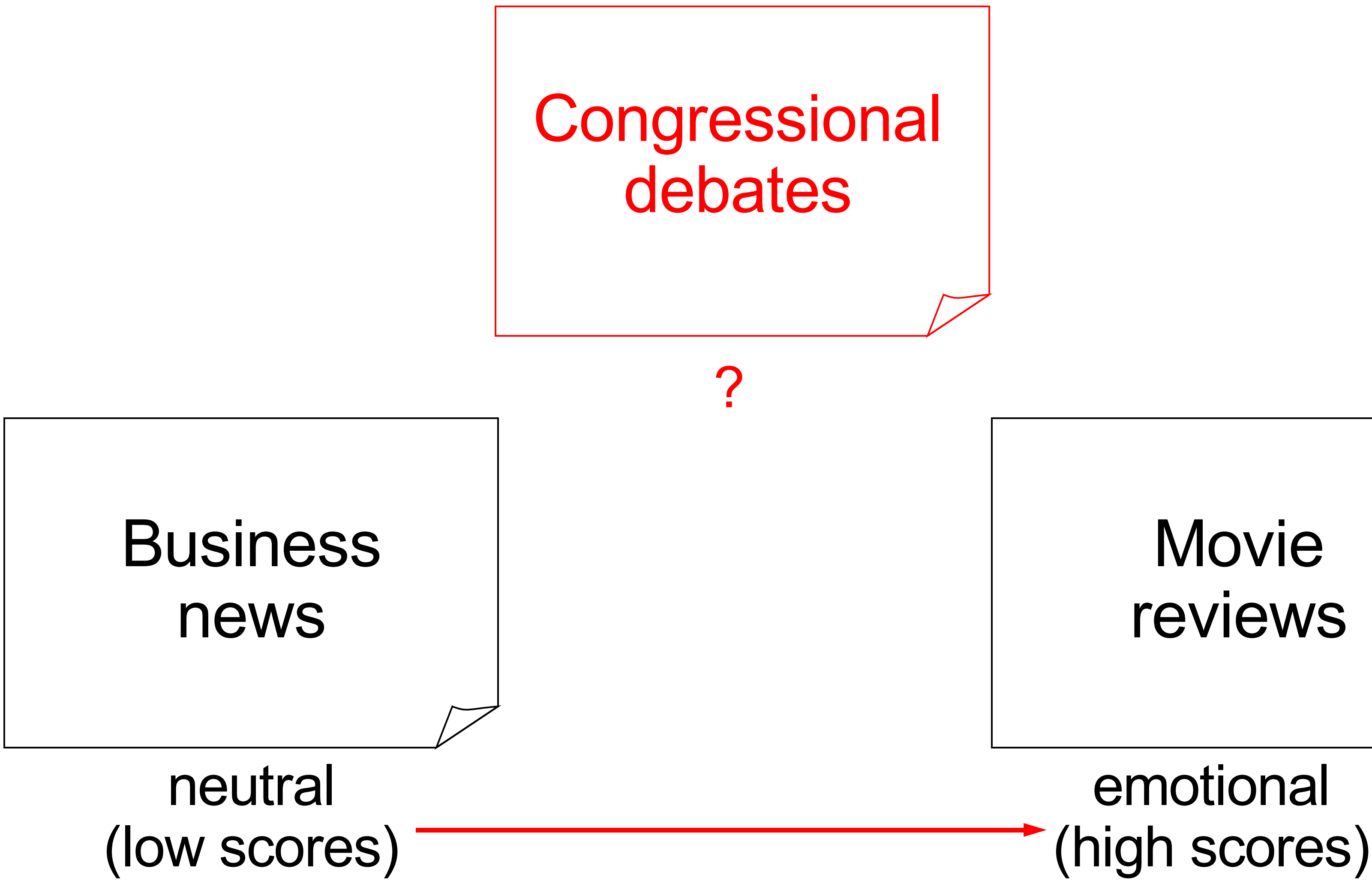
Measured by LIWC scores:

relative frequency of words classified in the LIWC dictionary as

affective, positive, and negative



Sentiment strength



Sentiment strength

Comparing the three domains

- <u>News articles</u>: 0.13M on Wal-Mart in 2006
  - 12 subsets, one for each month
- <u>Movie reviews</u>: 2000 (Pang et al., 2002)
  - 10 subsets, randomly partitioned
- <u>Senatorial speeches</u>: 1989-2006
  - 18 subsets, one for each year

Sentiment strength

	affective	positive	negative
News articles	2.78 (0.07)	1.91(0.06)	0.84 (0.07)
Senate debates	3.43 (0.13)	2.35 (0.09)	1.05 (0.06)
Movie reviews	4.98 (0.12)	3.04 (0.09)	1.94 (0.06)

Mean (std. dev.) of scores across subsets of each corpus

- Low standard deviation → consistent within each domain
- Increasing scores (significant differences)
  - News articles < Senate debates < Movie reviews

Parts of speech

Movie reviews

Congressional debates

Adjectives most important

?

Parts of speech

SVM- based feature selection:

1. Use content words as features for SVM classification
3. Select the top-ranked 100 features
4. Count the proportion of different parts of speech

Two domains:

- Movie reviews
- Congressional debate

Parts of speech

Accuracy

Before FS: 85.7%

After FS: 85.5%

Adjectives most important in movie reviews

Movie reviews

Nouns	Verbs	Adj.	Adv.
28	16	35	21
nothing	have	bad	unfortunately
mess	work	worst	only
script	supposed	boring	maybe
others	show	awful	perfectly
plot	tries	hilarious	also
performances	falls	ridiculous	especially
flaws	‘ve	memorable	sometimes
life	makes	terrific	wonderfully
reason	fails	poor	then
fun	looks	excellent	definitely
		lame	most
			very

Parts of speech

Accuracy

Before FS: 65.5%

After FS: 64.9%

Nouns

most important

in congressional speech

House debates

Nouns	Verbs	Adj.	Adv.
48	27	17	8
i	oppose	republican	n’t
demand	appreciate	parliamentary	very
leader	is	present	not
vote	recorded	small	here
majority	support	important	instead
support	vote	corporate	forward
cuts	reclaiming	safe	finally
inquiry	give	right	already
nays		many	
yeas		open	
work			
opposition			

Parts of speech

Accuracy

Before FS: 80.3%

After FS: 92.2%

Nouns

most important

in senatorial speech

Senatorial speeches (Partial-Birth Ban Act)

Nouns	Verbs	Adj.	Adv.
45	27	17	11
boxer	been	unconstitutional	much
women	say	partial-birth	obviously
health	done	medical	forward
abortion	have	right	then
people	has	legislative	simply
legislation	asking	pregnant	unfortunately
doctor	told	terrible	necessarily
v	means	multiple	roughly
opportunity	think	better	also
	address		really
	pass		

Parts of speech

Movie Reviews:

- adjectives are most indicative
- general, evaluative words (*bad, hilarious)*

Political speech:

- nouns are most indicative
- topic dependent, ordinarily neutral words (*women, choice)*

Summary:

Why is political opinion classification hard?

Two reasons:

- Relatively low sentiment strength
- Topic-dependent nouns are most indicative

Question:

Is political opinion classification
hard for people?

Study:

- Manual annotation

Manual annotation

Materials:

- Development set of Thomas et al. (2006)'s 2005 House debate corpus
- 5 debates, each on one bill
- 113 speeches (concatenating speeches from same speakers on same bills)

Annotators: 3 undergraduate students

Categories:

- **S**upport
- **O**ppose
- **N**eutral
- **I**rrelevant

Manual annotation

A1 vs A2
($\kappa = 0.80$)

	S	O	N	I
S	38	1	5	2
O	0	30	0	0
N	2	2	15	2
I	0	2	0	14

A1 vs A3
($\kappa = 0.61$)

	S	O	N	I
S	38	0	5	3
O	0	27	3	0
N	2	2	15	2
I	0	2	0	14

A2 vs A3
($\kappa = 0.64$)

	S	O	N	I
S	35	0	3	2
O	1	30	4	0
N	5	3	9	3
I	5	2	1	10

- Much agreement
- Very little disagreement on *Support* vs. *Oppose*
- Some disagreement on *Support/Oppose* vs. *Neutral/Irrelevant*
- Discrepancies resolved after discussion

Manual annotation

Recurring reasons for disagreement:

- Mixed opinions
- Implicit opinions
- Implicit bills in speeches
- Opinions on part vs. whole bills
- Procedural speeches

Summary:

Is political opinion classification
hard for people?

Much easier than for computers:

- Domain-specific vocabulary requires background knowledge (which humans have but computers don’t)
- Dispassionate, objective style contains few “general-purpose” sentiment words

Question:

What are we classifying, anyway?

If sentiment is not a good indicator, what is?

- Voting patterns?
- Tonality?

Opinion – Vote – Tone

Political Opinions	Votes	Tones
Support Oppose Neutral	Yea Nay	Positive Negative Neutral

Political opinion classification = Vote classification?

Political opinion classification = Tone classification?

Opinion = Vote?

Manual annotation of speeches vs. vote of speaker

	S	O	N	I
Yea (56)	38	0	9	9
Nay (57)	0	30	21	6

60%

40%

- Speakers vote the way humans predict...
- But humans make many non-predictions.
- Bad prospects for vote-based opinion classifiers?

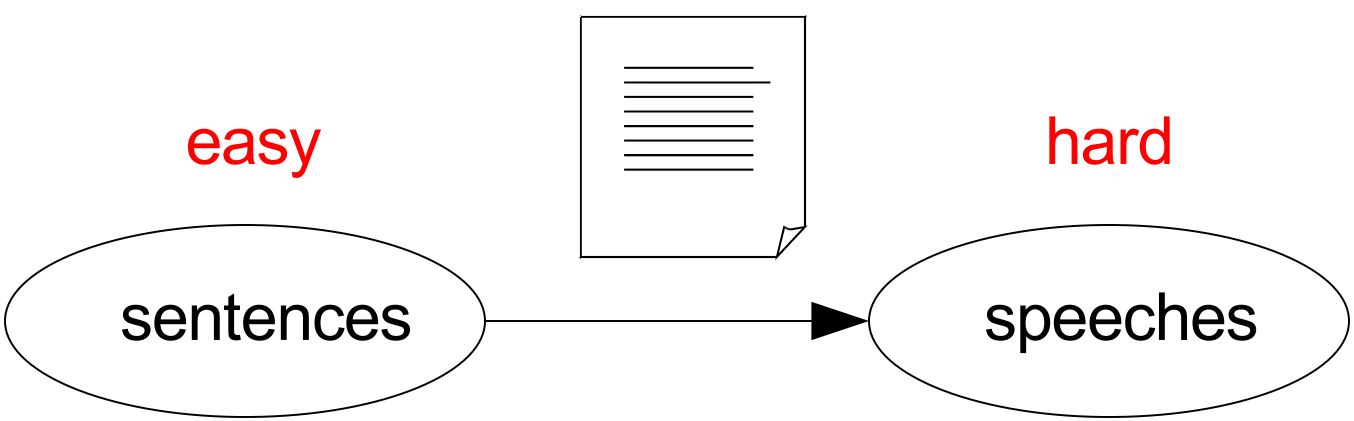
Opinion ≠ Vote

- Speakers are not required to defend their votes in speeches
- Votes are subject to stronger institutional pressures than speeches
- Speakers may list pros and cons of the bill without indicating their votes

Opinion = Tone?

Support = Positive
Oppose = Negative
Neutral = Neutral

Tone identification



Opinion = Tone?

Defining tones at speech level:

- Fundamental tones: the majority of sentence tones
- Beginning tones: tones of beginning sentences
- Ending tones: tones of ending sentences

Annotator: One student

Agreement Test:

- Debate on Stem Cell Research Act in House and Senate
  - Controversial topic
  - High emotion level

Opinion and Fundamental Tone

House	**Positive**	**Negative**	**Neutral**	*Total*
Support	43	6	31	80
Oppose	4	14	19	37
Neutral	1	0	8	9
Total	48	20	58	126

Senate	**Positive**	**Negative**	**Neutral**	*Total*
Support	21	3	19	43
Oppose	5	0	11	16
Neutral	1	0	3	4
Total	27	3	33	63

➜ Strong opinions expressed in neutral tones

Opinion and Fundamental Tone

- Debate on Stem Cell Research Act
  - Agreement 51.6% in House debate
  - Agreement 38.1% in Senate debate
  - Some conflicts
    - Support in negative tones (criticizing opponents)
    - Opposition in positive tones (praising life)
  - <10% neutral opinions
  - ~50% neutral fundamental tones

~50% - accuracy upper bound for fundamental tone based opinion classifier

Opinion and Beginning Tone

House	**Positive**	**Negative**	**Neutral**	*Total*
Support	63	2	15	80
Oppose	13	7	17	37
Neutral	0	1	8	9
Total	76	10	40	126

Senate	**Positive**	**Negative**	**Neutral**	*Total*
Support	19	1	23	43
Oppose	3	1	12	16
Neutral	0	0	4	4
Total	22	2	39	63

Opinion and Ending Tone

House	Positive	Negative	Neutral	*Total*
Support	64	1	15	80
Oppose	2	20	15	37
Neutral	2	0	7	9
Total	68	21	37	126

Senate	Positive	Negative	Neutral	*Total*
Support	7	0	36	43
Oppose	0	1	15	16
Neutral	1	0	3	4
Total	8	1	54	63

Opinion ≠ Tone

Agreement between opinion and tone

	Fundamental	**Beginning**	**Ending**
House	51.6%	61.9%	72.2%
Senate	38.1%	36.5%	17.5%

Note: Overall higher agreement in House than Senate

Hypothesis: House debates are more ideologically divided.

Summary:

What are we classifying, anyway?

- Not voting or tone.
- We need a better understanding of the categories and the corresponding linguistic characteristics.

Conclusions

Characteristics of political opinion expressions that affect sentiment classification

- Medium level sentiment strength
  - Lacking convergent sentiment vocabulary
- Nouns, not adjectives, as opinion indicators
  - Introducing topic relevance
- High percentage of neutral tones in political opinion expressions
  - Tone recognition is not enough

What might help

- Domain-specific sentiment vocabulary
  – e.g. “unconstitutional”
- Speaker ideology identification
  – Ideology Governs opinions on various topics
- Speaker social network analysis
  – Agreement or disagreement

Anything else?

References

Agrawal, R., Rajagopalan, S., Srikant, R., & Xu, Y. (2003). Mining newsgroups using networks arising from social behavior. Proceedings of the 12th international conference on World Wide Web (WWW2003), 529-535

Kwon, N., Zhou, L., Hovy, E., & Shulman, S.W. (2006). Identifying and classifying subjective claims. Proceedings of the 8th Annual International Digital Government Research Conference, 76-81

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumps up?: Sentiment classification using machine learning techniques. Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP2002), 79-86