Sentiment Flow

A General Model of Web Review Argumentation

Henning Wachsmuth, Johannes Kiesel, Benno Stein
henning.wachsmuth@uni-weimar.de

Bauhaus-Universität Weimar
www.webis.de

Web reviews across domains

★★★☆☆

“This book was different. I liked the first part. I could relate with Pi on his views about God and religion. He put into words my feelings when he said, “I just want to love God“ to the three religious leaders (Catholic, Muslim, Hindu) when they asked him why he practiced all three religions. I puzzled over the middle while he was lost at sea with the tiger. I didn't get the island at all. But in the end it all came together.“

●●●○○

“We stayed overnight at the Castle Inn in San Francisco in November. It was a fairly convenient to Alcatraz Island and California Academy of Science in Golden Gate Park. We were looking for a reasonably priced convenient location in SF that we did not have to pay for parking. Very basic motel with comfortable beds, mini refrig and basic continental breakfast. It was within walking distance to quite a few restaurants (Miller's East Coast Deli-yummy!)

I did find that the clerk at the desk was rather unfriendly, though helpful. The free parking spaces were extremely tight for our mini van. The noise was not too bad, being only 1 block from Van Ness Ave. If you are looking for a no frills, comfortable place to stay, Castle Inn was a good choice.“

Research questions

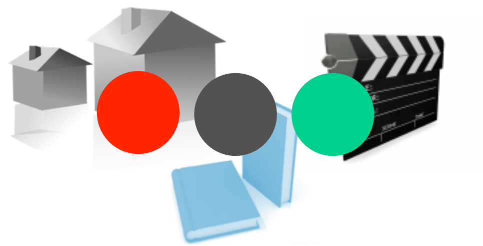
- **Web reviews** vary in several respects across domains

...

- **Sentiment analysis** of web reviews tends to be domain-dependent

1. Is there a
general way how people argue in web reviews?

2. Can we exploit that for
domain robustness in sentiment analysis?

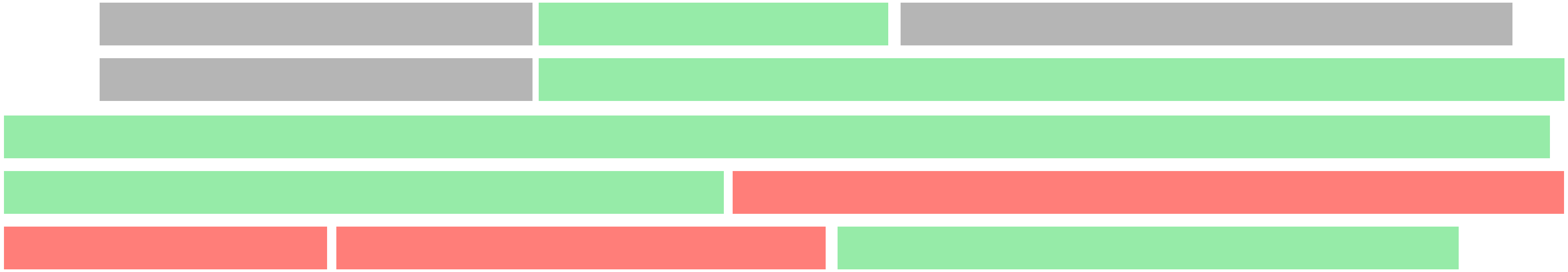


Web review argumentation

“ This book was different. I liked the first part. I could relate with Pi on his views about God and religion. He put into words my feelings when he said, “I just want to love God“ to the three religious leaders (Catholic, Muslim, Hindu) when they asked him why he practiced all three religions. I puzzled over the middle while he was lost at sea with the tiger. I didn't get the island at all. But in the end it all came together. “

“ We stayed overnight at the Castle Inn in San Francisco in November. It was a fairly convenient to Alcatraz Island and California Academy of Science in Golden Gate Park. We were looking for a reasonably priced convenient location in SF that we did not have to pay for parking. Very basic motel with comfortable beds, mini refrig and basic continental breakfast. It was within walking distance to quite a few restaurants (Miller's East Coast Deli-yummy!)
I did find that the clerk at the desk was rather unfriendly, though helpful. The free parking spaces were extremely tight for our mini van. The noise was not too bad, being only 1 block from Van Ness Ave. If you are looking for a no frills, comfortable place to stay, Castle Inn was a good choice. “

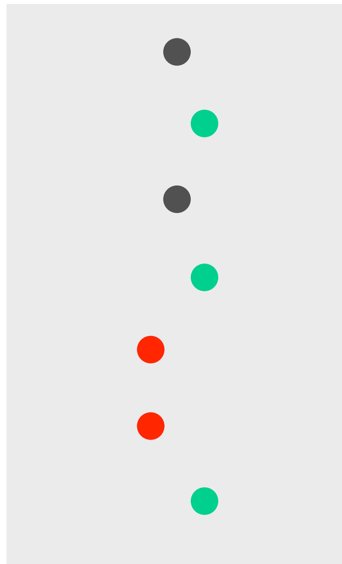
Sentiment flow



- **Model:** Overall argumentation of a web review as a sequence of local sentiments
  - Called “sentiment flow“ (Mao & Lebanon, NIPS’07)

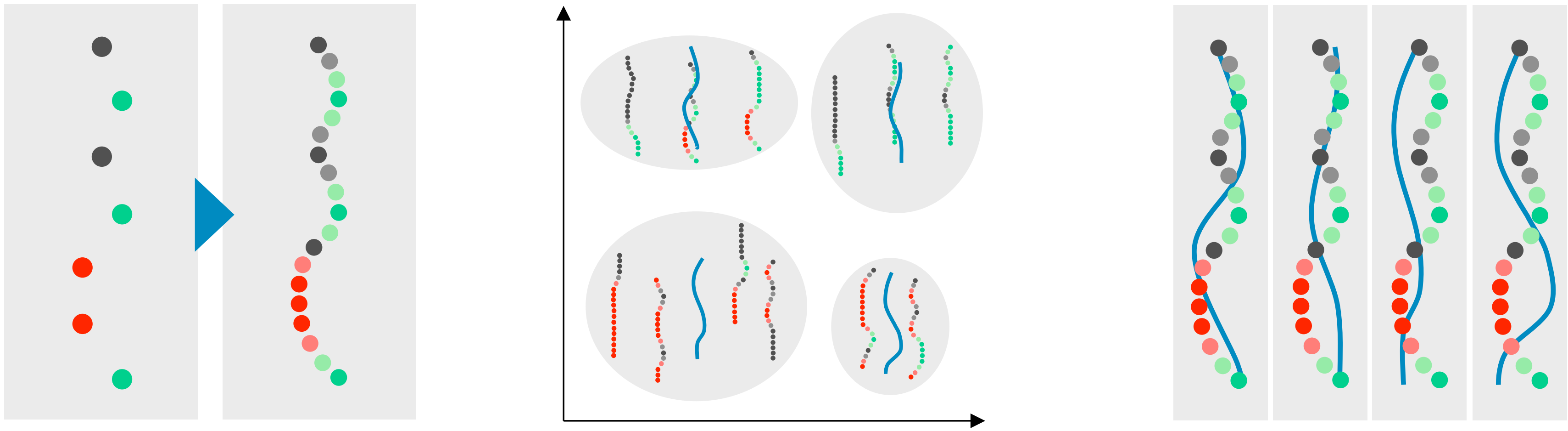
- **Hypothesis:** Similar sentiment flows express similar global sentiment across domains

- **Problem:** Original sentiment flow will not generalize well



Previous work

- **(Wachsmuth et. al., COLING'14):** Learn to predict global sentiment based on common "sentiment flow patterns"

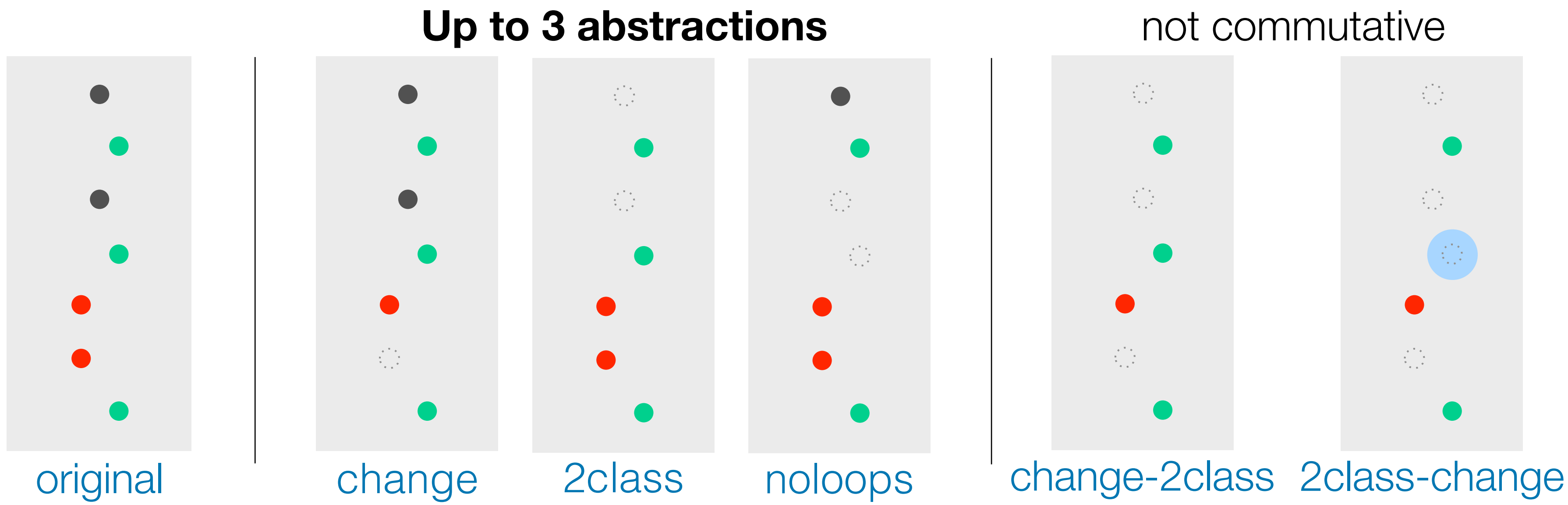


1. Normalize length of all sentiment flows
2. Cluster training flows to find sentiment flow patterns
3. Compare unknown flow to all patterns

- **Normalization** can maintain all flow information
  - Flows similar only if changes at similar positions (used Manhattan distance)
- **Clustering** can optimally group similar flows
  - Full of parameters and time-intensive

This work

- **Goal:** Obtain a sentiment flow model that generalizes across domains



- **Abstraction** aims to reduce domain differences
    - Length, subjectivity, “sub-reviews“
- **Resulting models** cover fewer but more common flows
    - No need for normalization and clustering
    - Favors measures like minimum edit distance (details in paper)

Ground-truth data

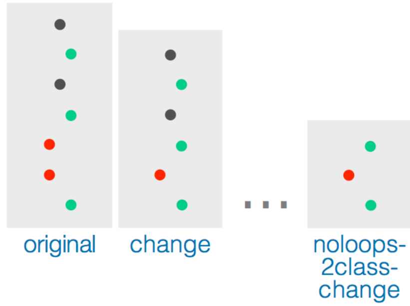
	 **Amazon product reviews** (Täckström et. al., ECIR'11)	 **TripAdvisor hotel reviews** (Wachsmuth et. al., CICLing'14)	 **Rotten tomatoes movie reviews** (Mao & Lebanon, NIPS'07)
texts	294 in total (from 5 categories)	2100 in total (from 7 locations)	450 in total (from 2 authors)
	175 for training	900 for training	201 for training
sentences	14.0 per text	11.5 per text	28.8 per text
local sentiment	● 34% negative ● 42% neutral ● 24% positive	● 42% negative ● 20% neutral ● 38% positive	● 21% negative ● 61% neutral ● 18% positive

- Mapped review overall ratings to three global sentiments: ● ● ●

Experiment on the generality

- **Hypothesis:** Similar sentiment flows are used generally across domains

- **Comparison** of 16 model variants (only 4 here)
  - Original sentiment flow
  - Each combination of 1 to 3 abstractions



- **Experiments** for all 9 domain combinations
  1. Compute known flows and their majority global sentiment on training set of one domain
  2. Compare with flows on test set for each domain

- **Measures** to assess generality (1 more in the paper)
  - Recall: % of test reviews with a known flow
  - Precision: % of known test flows whose global sentiment matches the majority

change

recall 44
precision 83

Selected generality results

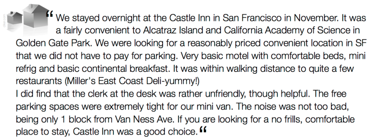
Sentiment flow model variant	Training domain	Recall in %				Precision in %			
original		1	4	0	3	100	95	0	39
		1	17	0		75	79	0	
		0	0	0		0	0	0	
2class		45	37	29	30	83	85	80	**86**
		54	33	30		83	[illegible]	78	
		22	11	8		91	97	95	
change-2class-noloops		86	86	77	84	78	74	73	74
		93	92	85		76	73	73	
		86	73	81		76	73	69	
2class-change-noloops		100	100	100	**99**	76	70	62	68
		100	100	100		71	69	62	
		97	95	100		72	70	59	

picomall.com clker.com canstockphoto.com

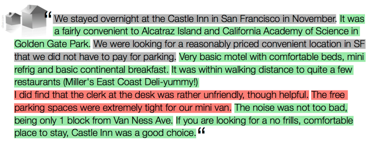
Experiment on the robustness

- **Hypothesis:** Sentiment flows allow for domain-robust sentiment analysis

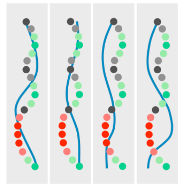
- **Comparison** of 4 feature types (more detailed in the paper)



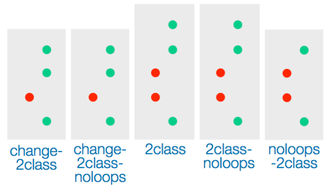
Bag-of-words



Local sentiment distribution

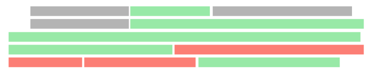


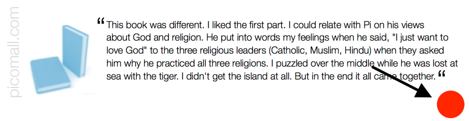
Sentiment flow patterns (Wachsmuth et. al., COLING‘14)

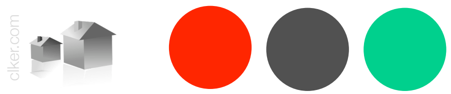


5 sentiment flow model variants

- **Experiments** for all 9 domain combinations
  1. Determine all local sentiment with Stanford CoreNLP (Socher et. al., EMNLP‘13)
  2. Learn default random forest classifier on training set of one domain
  3. Classify global sentiment on test set for each domain without any domain adaptation







Selected robustness results

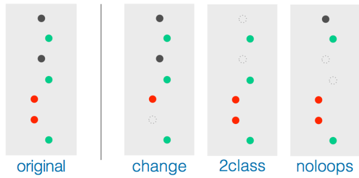
Feature type	Accuracy in %			ø Within domain	ø Out-of domain
Bag-of-words	49	46	32	65	39
	38	80	40		
	35	41	65		
Local sentiment distribution	52	50	39	58	46
	51	64	51		
	43	44	59		
Sentiment flow patterns	47	58	48	63	48
	51	74	51		
	42	40	67		
5 sentiment flow model variants	51	51	42	60	49
	53	69	55		
	45	50	61		

68 53

All 4 feature types

Conclusion

- **Sentiment flow as an argumentation model** for web reviews
  - Sequence of local sentiment only
  - Represents argumentation regarding global sentiment



- **Generalizes across domains** when abstracted adequately
  - The same flows are frequent across domains
  - Flows imply similar global sentiment across domains

Sentiment flow model variant	Training domain	Recall in %			Precision in %			Recall in %	Precision in %
original		1	4	0	100	95	0	3	39
		1	17	0	75	[illegible]	0		
		0	0	0	0	0	0		
2class		45	37	29	83	85	80	30	86
		54	[illegible]	30	83	[illegible]	78		
		22	11	8	91	97	95		
change-2class-noloops		86	86	77	78	74	73	84	74
		93	[illegible]	85	76	[illegible]	73		
		86	73	81	76	73	69		
2class-change-noloops		100	100	100	76	70	62	99	68
		100	[illegible]	100	71	[illegible]	62		
		97	95	100	72	70	59		

- **Benefits domain robustness** in sentiment analysis
  - Flow features best out-of-domain
  - Accuracy still improvable

Feature type		Accuracy in %			ø Within domain	ø Out-of domain
Bag-of-words		49	46	32	64	39
		38	**80**	40		
		35	41	65		
Local sentiment distribution		**52**	50	39	58	47
		51	64	51		
		43	44	59		
Sentiment flow patterns		47	**58**	**48**	63	48
		51	74	51		
		42	40	**67**		
5 sentiment flow model variants		51	51	42	60	49
		53	69	**55**		
		45	**50**	61		

68 **53**
All 4 feature types

- **Promising for domain adaptation** and shallow text analyses
  - Pivot features in domain adaptation
  - Candidate retrieval in argumentation mining