

I Couldn't Agree More: The Role of Conversational Structure in Agreement and Disagreement Detection in Online Discussions

Sara Rosenthal
Kathleen McKeown
Columbia University



COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science



Motivation

- Detecting (dis)agreement is useful for understanding how conflicts arise and are resolved and the role of participants in a conversation
- It is also useful for other tasks such as detecting subgroups, stance, power, and interactions

Related Work

- **Agreement Detection in Speech**
 - Galley et al 2004; Hillard et al 2003; Hahn et al 2006
 - ICSI, AMI meeting corpora
 - Detecting Adjacency Pairs
 - Supervised System Features: sentiment, n-grams, (dis)agreement terms
 - motivate our approach
- **Agreement Detection in Online Discussions**
 - Yin et al. 2012; Abbott et al. 2011; Misra and Walker 2013; Mukherjee and Liu 2012
 - two-way agreement detection
 - IAC, US message board, Political Forum, AAWD
 - Largest dataset (IAC) is 2,800 posts
 - Supervised System Features: lexical, lexical-style, thread structure, polarity

Definition

Quote-Response (Q-R) Posts

Agreement occurs between two posts where one is an immediate response to the other

Quote: That's a good idea.

Response: I agree!

Definition

Quote-Response (Q-R) Posts

Agreement occurs between two posts where one is an immediate response to the other

Quote: That's a good idea.

Response: I agree!

Agreement!

Outline

- Data
 - Large self-labeled dataset
- Method
 - Supervised Approach
 - Rich suite of features: structural, lexical, and style
- Experiments
- Conclusion

Datasets

The screenshot shows a 4forums debate titled "Are high taxes theft?". The debate is split into two sides: "Yes they are theft" with a Side Score of 93, and "No they are not theft" with a Side Score of 72. The "VS" logo is in the center. Below the title, there are two columns of posts. The left column has a post by JoeCavalry (21537) with 4 points, discussing a parable about watering a tree and taking money from others. The right column has a post by ThewayItIs (2290) with 4 points, arguing that high taxes are not theft because they benefit society. Below these are several replies, including one by GuitaristDog (1092) who disputes the right side's argument about a bridge in Alaska, and another by ThewayItIs (2290) who is disputed for claiming a bridge was never built.

ABCD: Create Debate

The screenshot shows a Wikipedia Talk Page titled "Ahmadinejad lying about having served in Iran-Iraq war". The page contains a discussion about a claim that Ahmadinejad served in the Iran-Iraq war. A user named Kurdo777 (talk) states that the claim is false and based on a weak source. Another user, Stephan Schulz (talk), agrees and suggests that the claim should not be reported in the article without multiple reliable sources. A third user, Cs32en (Talk to me), adds that the claim is BLP sensitive and needs multiple reliable sources. The page also includes a signature block for Cs32en.

AWTP: Wikipedia Talk Pages

The screenshot shows a 4forums thread titled "When does life begin?". The thread starts with a question: "When does life begin? Is it at fertilization, or is it at the point when the requirements for considered life are met?" The thread includes a quote from a user named "sinjin" who asks "You first." and "A brain dead patient in a hospital bed is alive." The thread also includes a quote from a user named "Sin Simian" who asks "Both sperm and ova are composed of living cells. Try again." The thread ends with a quote: "They asked if I had found Jesus and I didn't even know He was missing."

IAC: 4forums

ABCD
Discussion
about
investigating
torture
claims
against
President
Bush

Agreement

Disagreement



Libertarian1 While im sure liberals would love for that to happen, it simply will do no good.you'd have to put on trial every military(or otherwise) organization that either took part in such a crime being committed. And we all know the government doesn't rat itself out.



chatturgha While he's at it, he should investigate the possible tens of thousands of innocent Iraqi civilians that were murdered during the second Iraq war, all on Bush's hands. Honestly, I believe in torture... but only in torture of the deserving. Since the tortured people were likely innocent, this should also be investigated. Not whether torture happened, but whether the people were horrendous, murdering and/or molesting monsters.



garry77777 "he should investigate the possible tens of thousands of innocent Iraqi civilians that were murdered during the second Iraq war, all on Bush's hands." I must disagree with your numbers, as most americans are unaware that best estimates put the actual number of dead in Iraq since the start of the invasion in 2003 at 1.2 million people



chatturgha Okay then, he killed MORE people then just tens of thousands. And you're disagreeing with me... why?



VenusEve Having been raised by Republicans I can say they are paranoid, anal-retentive @ssholes. By all means investigate. Republicans can gripe all they want to about Obama but at least Obama is a good father! I am with the Democrats now. Yes, the Bush torture claims should be investigated. It's only right.



CupioMinimus Of course he should, yes. But he won't. No one gets into power in the west unless the real PTB have got leverage on them. That's why none of our leaders do anything to rock the boat. Stray from the path but a little and it's character assassination. Not always with 'character' either ;]



ThePyg While I disagree with many aspects of the war, waterboarding, to me, shouldn't be something that's "investigated" as "torture". Our military and CIA have done what they can to protect the US citizens. Sure, I don't think they did it right, but to punish them for all they've done for OUR protection is... disturbing.



Phreekshow I do not look at it as a mark against the military who were doing what they were ordered to do by the Commander in Chief. Who is the final word when it comes to the military. maybe if Americans were able to experience waterboarding they would change their minds on whether it is torture.

ABCD
Discussion
about
investigating
torture
claims
against
President
Bush

Agreement

Disagreement



Libertarian1 While im sure liberals would love for that to happen, it simply will do no good.you'd have to put on trial every military(or otherwise) organization that either took part in such a crime being committed. And we all know the government doesn't rat itself out.



chatturgha While he's at it, he should investigate the possible tens of thousands of innocent Iraqi civilians that were murdered during the second Iraq war, all on Bush's hands. Honestly, I believe in torture... but only in torture of the deserving. Since the tortured people were likely innocent, this should also be investigated. Not whether torture happened, but whether the people were horrendous, murdering and/or molesting monsters.



garry77777 "he should investigate the possible tens of thousands of innocent Iraqi civilians that were murdered during the second Iraq war, all on Bush's hands." I must disagree with your numbers, as m
" I must disagree with your numbers, as m
Iraq since the number of dead in



chatturgha Okay then, he killed MORE people then just tens of thousands. And you're disagreeing with me... why?



VenusEve Having been raised by Republicans I can say they are paranoid, anal-retentive @ssholes. By all means investigate. Republicans can gripe all they want to about Obama but at least Obama is a good father! I am with the Democrats now. Yes, the Bush torture claims should be investigated. It's only right.



CupioMinimus Of course he should, yes he gets into power in the west unless the real PTB have got le
ers do anything to rock the boat. Stray from the path but a little and it's character assassination. Not always with 'character' either ;]



ThePyg While I disagree with many aspects of the war, waterboarding, to me, shouldn't be something that's "investigated" as "torture". Our military and CIA have done what they can to protect the US citizens. Sure, I don't think they did it right, but to punish them for all they've done for OUR protection is... disturbing.



Phreekshow I do not look at it as a mark against the military who were doing what they were ordered to do by the Commander in Chief. Who is the final word when it comes to the military. maybe if Americans were able to experience waterboarding they would change their minds on whether it is torture.

Data

Create Debate

- Website where people can start debates
 - Open-ended: no side
 - **For-or-against: two sided**
 - Multiple sides: three or more sides

Each post is labeled with the “for” or “against” side

Data

Create Debate

- **Agreement:** Quote and Response have same side
- **Disagreement:** Quote and Response have different side
- **None:**
 - Quote is Root
 - Quote and Response have same author

Agreement by Create Debaters (ABCD)

Data

Create Debate (ABCD)

Diets are nasty. Coke is the only soda in the world I will pretty much tolerate. **Side: *Regular***

Why are diet sodas nasty? They contain artificial sweeteners which actually start tasting good after you drink them for a couple of weeks. The upside is that you aren't consuming a can full of sugar (i.e. empty calories)! **Side: *Diet Coke***

ABCD Disagreement Example

Data

Create Debate (ABCD)

while diet coke is more likely to kill you and cause cancer and stuff, but, it does taste better. death tastes yummy. **Side: *Diet Coke***

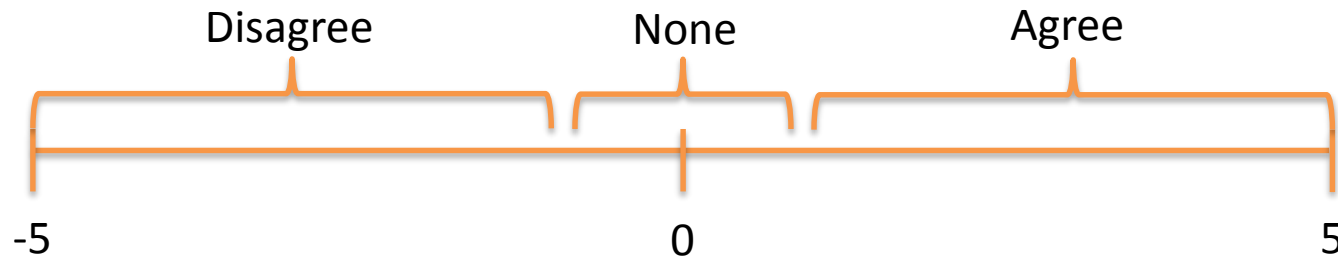
Death does taste yummy. **Side: *Diet Coke***

ABCD Agreement Example

Data

Internet Argument Corpus (IAC)

- Mechanical Turk
- Labeled on scale of $\{-5, 5\}$



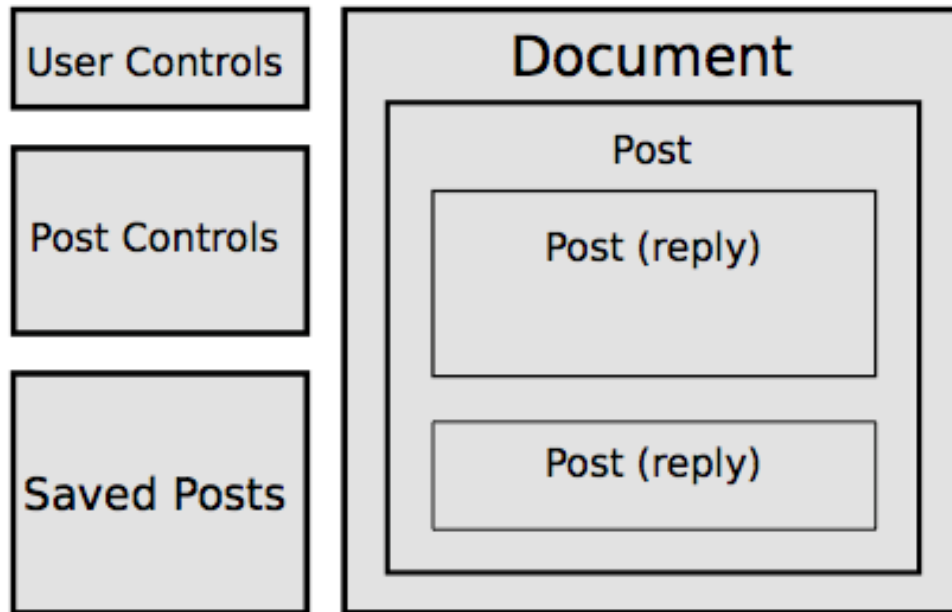
- Not all Q-R pairs in a thread were annotated

Converted to Post level annotations using majority pair level annotation

Data

Agreement in Wikipedia Talk Pages (AWTP)

Annotated using Annotation Tool



- Sentence Level Annotations
- 3 Annotators
- Inter-Annotator Agreement (IAA) computed on 30 sentence pairs
- Cohen's $\kappa = .90$ & $.70$

Converted to Post level annotations using majority sentence level annotation

Data Statistics

Dataset	Discussion Count	Post Count	Agreement	Disagreement	None
Create Debate (ABCD)	12553	207188	42689	68044	96455
Internet Argument Corpus (IAC)	1223	5940	428	1236	4276
Wikipedia Talk Pages (AWTP)	50	822	38	148	636


Data Statistics

Dataset	Discussion Count	Post Count	Agreement	Disagreement	None
Create Debate (ABCD)	12553	207188	42689	68044	96455
Internet Argument Corpus (IAC)	1223	5940	428	1236	4276
Wikipedia Talk Pages (AWTP)	50	822	38	148	636

30 Times Larger!

Data Statistics

Dataset	Discussion Count	Post Count	Agreement	Disagreement	None
Create Debate (ABCD)	12553	207188	42689	68044	96455
Internet Argument Corpus (IAC)	1223	5940	428	1236	4276
Wikipedia Talk Pages (AWTP)	50	822	38	148	636


Argumentative

Data Statistics

Dataset	Discussion Count	Post Count	Agreement	Disagreement	None
Create Debate (ABCD)	12553	207188	42689	68044	96455
Internet Argument Corpus (IAC)	1223	5940	428	1236	4276
Wikipedia Talk Pages (AWTP)	50	822	38	148	636

Training: 80% of discussions
Test + Dev: 20% of discussions

Method

- Supervised Approach
- Features
 - Structural
 - Response related
 - Lexical, lexical style, LIWC, opinion
 - Q-R related
 - Sentence Similarity, Accommodation

Method

Thread Structure

Q is root

Q and R have same author

Distance of R from root

The number of sentences in R

X Post 1:

Y Post 2:

X Post 3:

Z Post 4:

Y Post 5:

Y Post 6:

X Post 7:

X Post 8:

Z Post 9:

Z Post 10:

Method

Thread Structure

Q is root

Q and R have same author

Distance of R from root

The number of sentences in R

X Post 1:

Y Post 2:

X Post 3:

Z Post 4:

Y Post 5:

Y Post 6:

X Post 7:

X Post 8:

Z Post 9:

Z Post 10:

Method

Thread Structure

Q is root

Q and R have same author

Distance of R from root

The number of sentences in R

X Post 1:

Y Post 2:

X Post 3:

Z Post 4:

Y Post 5:

Y Post 6:

X Post 7:

X Post 8:

Z Post 9:

Z Post 10:

Method

Thread Structure

Q is root

Q and R have same author

Distance of R from root

The number of sentences in R

X Post 1:

Y Post 2:

X Post 3:

Z Post 4:

Y Post 5:

Y Post 6:

X Post 7:

X Post 8:

D=2

Z Post 9:

Z Post 10:

Method

Thread Structure

Q is root

Q and R have same author

Distance of R from root

The number of sentences in R

X Post 1:

Y Post 2:

X Post 3:

Z Post 4:

Y Post 5:

Y Post 6:

X Post 7:

X Post 8:

Z Post 9:

Z Post 10:

Method

Lexical Features in *R*

- n-grams
- Part-of-Speech tags
- Terms:
 - Negation (11): not, nothing
 - Disagreement (14): disagree, differ
 - Agreement (16): agree, concur
- Did the response ask a question

RESPONSE: Do you think it is the best scholarly material published in the past 2000 years?

RESPONSE: Do you claim that Israel cannot exist without an occupying regime?

Method

Lexical-Stylistic Features in *R*

Feature	Example	Feature	Example
All Caps Words	WHAT	Punctuation Count	5
Out of Vocabulary	dunno	Exclamation Points	!
Emoticons	:)	Repeated Exclamations	!!!!
Acronyms	LOL	Question Marks	?
Punctuation	.	Repeated Questions	???
Repeated Punctuation	#\$@.	Ellipses	...
Link/Image	url.com	Word Lengthening	sweeeet
Capital Words	Hello	Avg. Word Length	4

Method

Linguistic Inquiry Word Count (LIWC)

Linguistic Processes	Psychological Processes	Personal Concerns	Spoken Categories
Negation	Family	Work	Assent
Pronouns	Positive Emotion	Money	Nonfluencies
Past Tense	Certainty	Home	Fillers
Swear Words	Health	Religion	

Include all categories that are used in *R* by looking at each word in the response and its associated categories

YR Tausczik and JW Pennebaker. 2010. **The psychological meaning of words: LIWC and computerized text analysis methods.**

Method

Opinion Detection

■ positive ■ negative ■ subjective ■ objective

[while diet coke] **[is more likely to kill you] [and cause cancer and stuff]**, [but,] **[it does taste better.] [death tastes yummy.]** Side: *Diet Coke*

- Features
 - R has subjective/polarity
 - Normalized count of subjective/polarity in R
 - n-grams of polarity words in R

Rosenthal et al. SemEval 2014. **Columbia NLP: Sentiment Detection of Sentences and Subjective Phrases in Social Media.**

Method

Sentence Similarity

while diet coke is more likely to kill you and cause cancer and stuff, but, it does taste better. death tastes yummy. Side: *Diet Coke*

Death does taste yummy. Side: *Diet Coke*

Does Q and R have similar sentences based on a given threshold (.66)

Method

Phrase Similarity + Sentiment

■ positive ■ negative ■ subjective ■ objective

while diet coke is more likely to kill you and cause cancer and stuff, but, it does taste better. death tastes yummy. Side: *Diet Coke*

Death does taste yummy. Side: *Diet Coke*

Features

- Has similar phrase(s)
- Similar phrases and polarity type
- Unique words from similar phrase(s)

Method

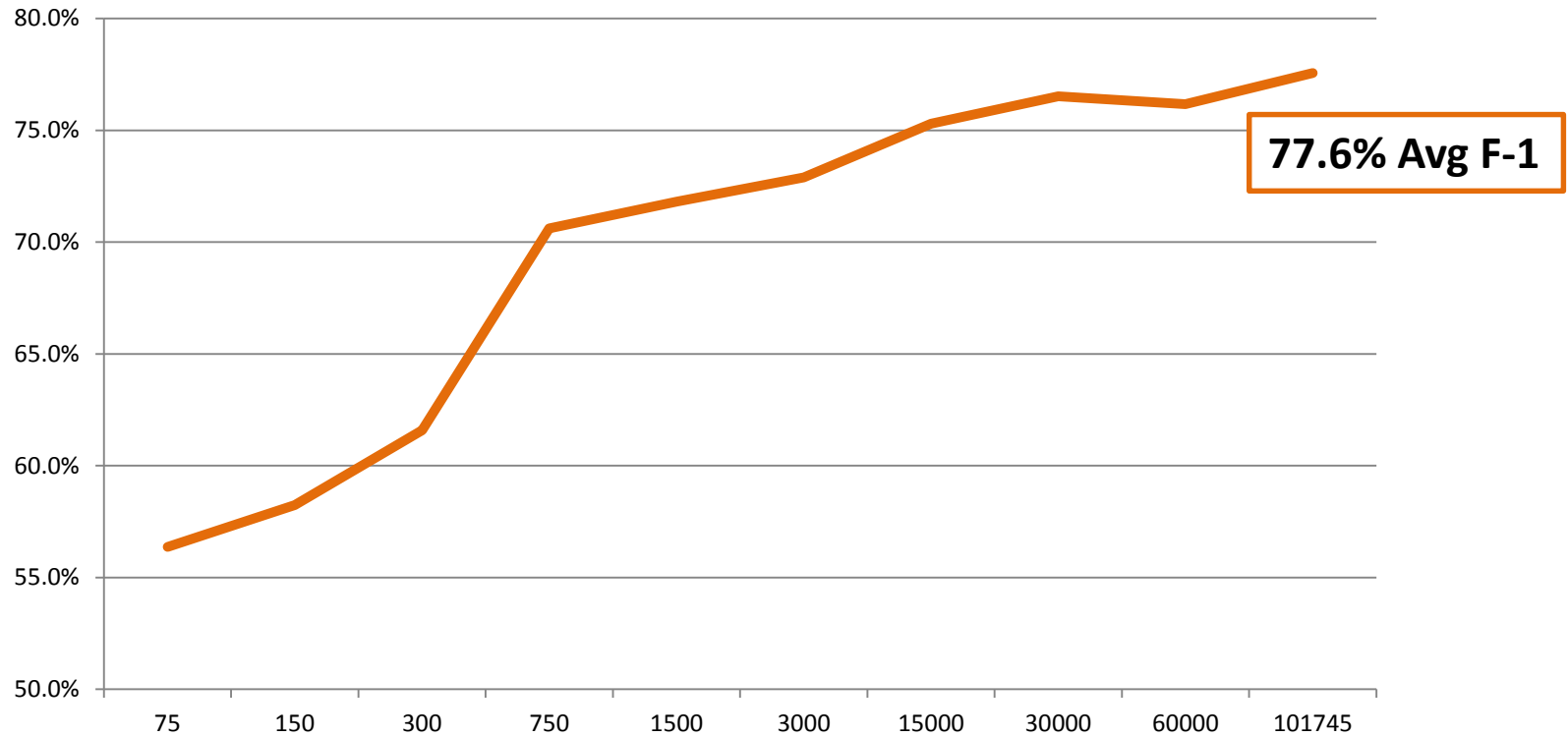
Accommodation

- Shared POS
 - e.g. Quote and Response have **DT JJ NN**
- Shared Lexical Style
 - e.g. Quote and Response have **emoticons**
- Share LIWC
 - e.g. Quote and Response have words regarding **family**

Experiments

- Logistic Regression
- 3-Way: Agreement / Disagreement / None
- Balanced training set
- Results in Average F-Score because (dis)agreement is rare

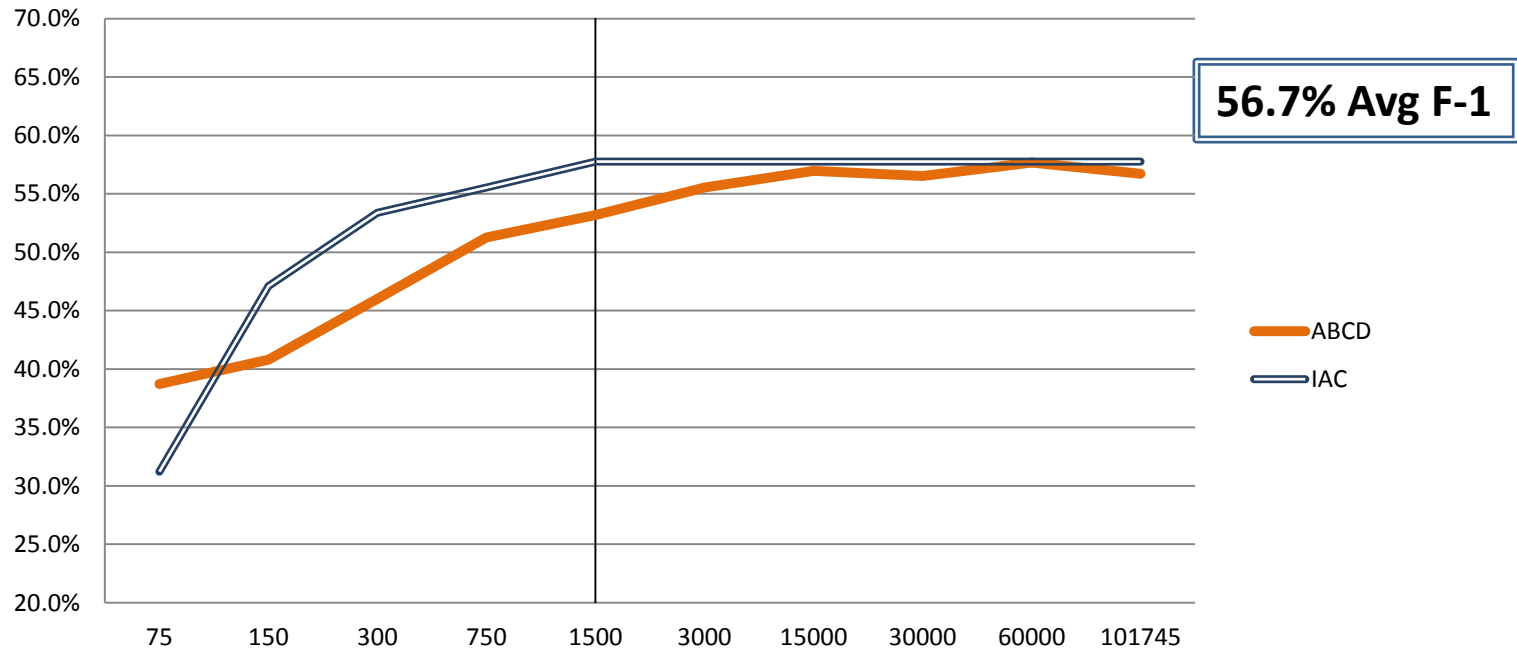
Agreement By Create Debaters



The Average F-score increases with the size of the training set

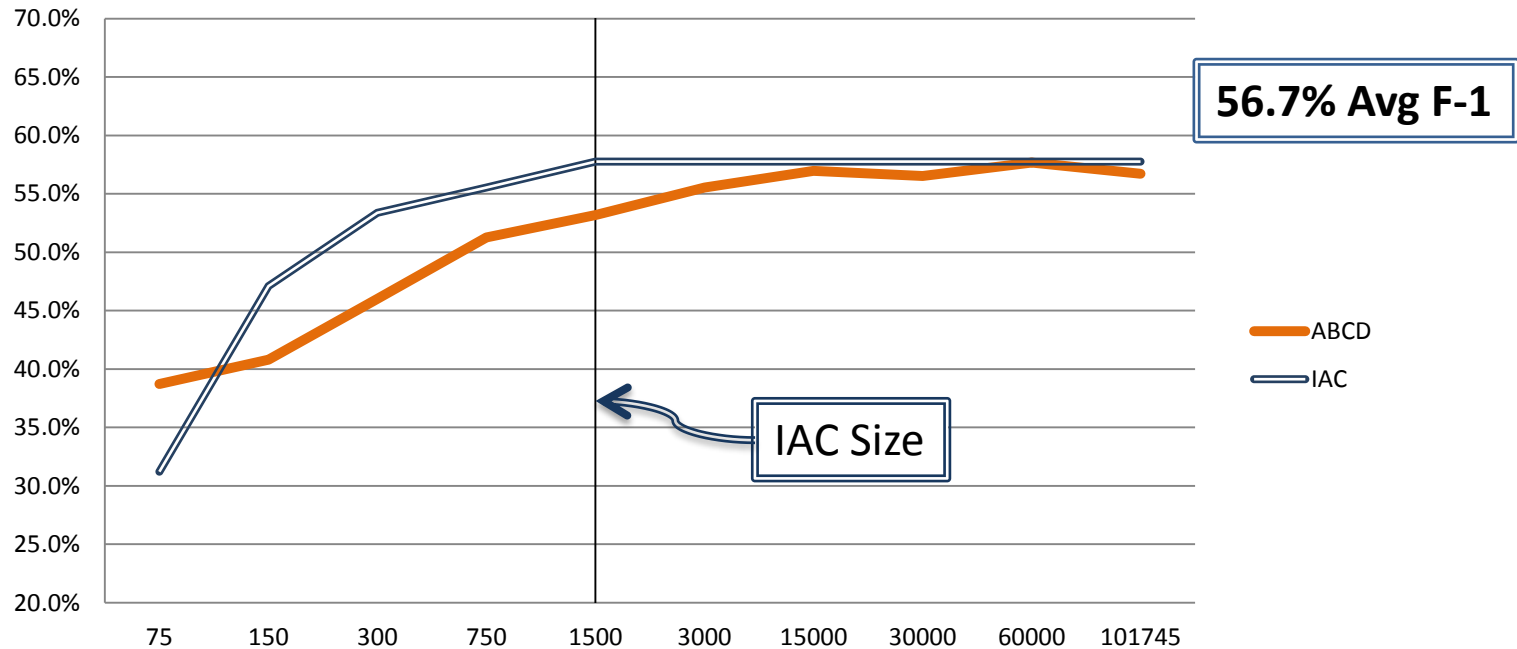
Can the ABCD corpus be used
to predict (dis)agreement in
other corpora?

Internet Argument Corpus



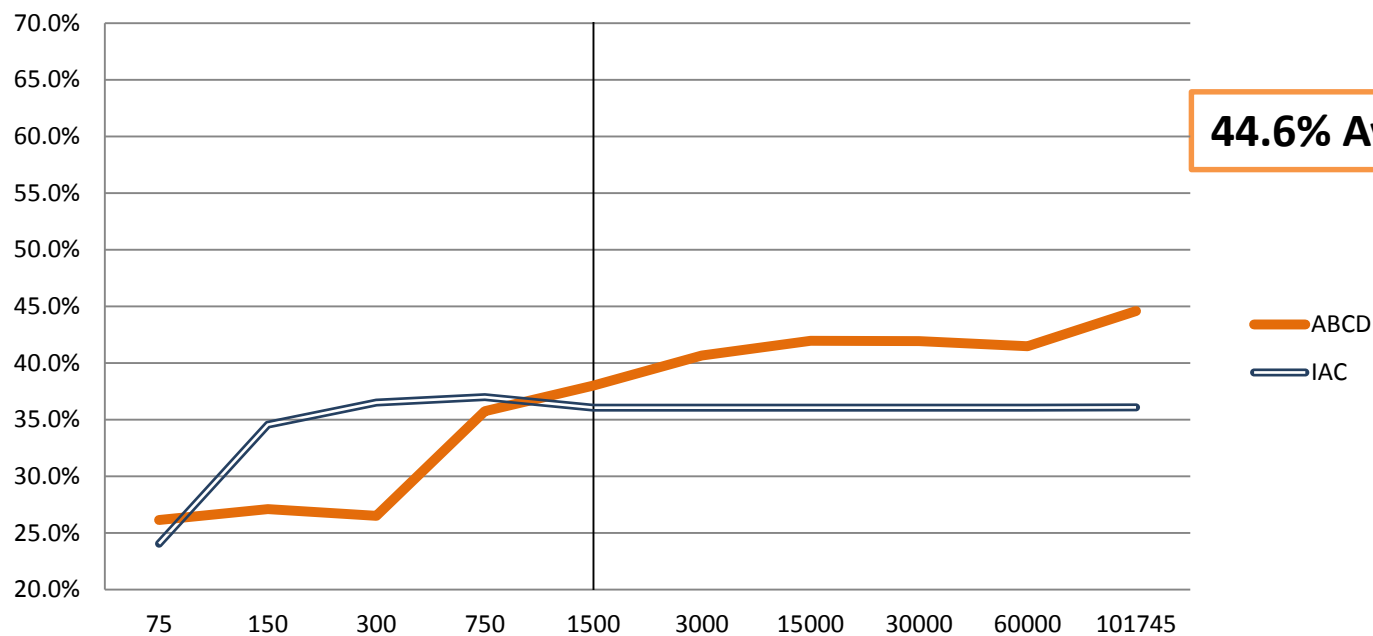
Using a large amount of naturally occurring ABCD labels does as well as a small set of in-domain gold labels

Internet Argument Corpus



Using a large amount of naturally occurring ABCD labels does as well as a small set of in-domain gold labels

Agreement in Wikipedia Talk Pages



44.6% Avg F-1

Using naturally occurring ABCD labels does significantly better than gold labels from an out of domain dataset (IAC)

Experiments and Results

Features	Training	ABCD	IAC	ABCD	IAC	ABCD
	Testing	ABCD	IAC		AWTP	
n-gram		40.9%	32.7%	30.3%	34.1%	26.7%
n-gram+LIWC+POS+Lexical-Style in Response		50.8%	31.9%	29.2%	33.0%	39.3%
Thread Structure		69.2%	54.2%	55.8%	31.4%	37.3%
Accommodation		59.4%	33.1%	33.6%	31.8%	36.1%
Thread Structure+Accommodation		75.2%	54.3%	56.9%	35.7%	43.9%
All		76.9%	54.2%	51.8%	38.7%	43.7%
Best		77.6%	57.8%	56.7%	36.1%	44.4%

Results in Average F-Score

Experiments and Results

Features	Training	ABCD	IAC	ABCD	IAC	ABCD
	Testing	ABCD	IAC		AWTP	
n-gram		40.9%	32.7%	30.3%	34.1%	26.7%
n-gram+LIWC+POS+Lexical-Style in Response		50.8%	31.9%	29.2%	33.0%	39.3%
Thread Structure		69.2%	54.2%	55.8%	31.4%	37.3%
Accommodation		59.4%	33.1%	33.6%	31.8%	36.1%
Thread Structure+Accommodation		75.2%	54.3%	56.9%	35.7%	43.9%
All		76.9%	54.2%	51.8%	38.7%	43.7%
Best		77.6%	57.8%	56.7%	36.1%	44.4%

Results in Average F-Score

Thread-Structure + Accommodation outperforms using thread structure and response only features

Experiments and Results

Features	Training	ABCD	IAC	ABCD	IAC	ABCD
	Testing	ABCD	IAC		AWTP	
n-gram		40.9%	32.7%	30.3%	34.1%	26.7%
n-gram+LIWC+POS+Lexical-Style in Response		50.8%	31.9%	29.2%	33.0%	39.3%
Thread Structure		69.2%	54.2%	55.8%	31.4%	37.3%
Accommodation		59.4%	33.1%	33.6%	31.8%	36.1%
Thread Structure+Accommodation		75.2%	54.3%	56.9%	35.7%	43.9%
All		76.9%	54.2%	51.8%	38.7%	43.7%
Best		77.6%	57.8%	56.7%	36.1%	44.4%

Results in Average F-Score

Using naturally occurring ABCD labels does as good, or better than smaller manually annotated datasets!

Discussion

	Quote	Response	Description
ABCD	The same thing people use all words for; to convey information.	to convey information. Give me an ex- ample of when you are fully capable of saying this without offending someone.	The first sentence sounds like agreement but the second sentence is argumentative
IAC	Nowhere does it say, that she kept a gun in the bathroom emoticon xkill	And nowhere does it say she went to her bedroom and retrieved a gun.	Agreement. It is an elaboration. Further context would help.

Detecting Agreement is Hard

Conclusion

- Conversational structure is important
 - thread-structure and accommodation
- Using naturally occurring labels does as good, or better than smaller manually annotated datasets
- Data Available at:
 - <http://www.cs.columbia.edu/~sara/data.php>

Future Work

- Use domain adaptation to combine the datasets
- Use system to correct mislabeling and retrain the model

Questions?