Human Language Technology Research Institute

Why are You Taking This Stance? Identifying and Classifying Reasons in Ideological Debates

Kazi Saidul Hasan and **Vincent Ng**
Human Language Technology Research Institute
University of Texas at Dallas

Debate Stance Classification

- Given a debate post written for a two-sided debate topic, determine the author's stance (*For* or *Against*)

Debate Stance Classification

- Given a debate post written for a two-sided debate topic, determine the author's stance (*For* or *Against*)

Should abortion be allowed?

Debate Stance Classification

- Given a debate post written for a two-sided debate topic, determine the author's stance (*For* or *Against*)

Should abortion be allowed?

Women should have the ability to choose what they do with their bodies.

Debate Stance Classification

- Given a debate post written for a two-sided debate topic, determine the author's stance (*For* or *Against*)

Should abortion be allowed?

Women should have the ability to choose what they do with their bodies.

For

Debate Stance Classification

- Given a debate post written for a two-sided debate topic, determine the author's stance (*For* or *Against*)

Should abortion be allowed?

Women should have the ability to choose what they do with their bodies.	Technically abortion is murder. They are killing the baby without a justified motive.
For	

Debate Stance Classification

- Given a debate post written for a two-sided debate topic, determine the author's stance (*For* or *Against*)

Should abortion be allowed?

Women should have the ability to choose what they do with their bodies.	Technically abortion is murder. They are killing the baby without a justified motive.
For	**Against**

Beyond Stance Classification

- Oftentimes, it's important to determine not only the author's stance, but also the reasons why she supports or opposes the issue under debate

Beyond Stance Classification

- Oftentimes, it's important to determine not only the author's stance, but also the reasons why she supports or opposes the issue under debate

- **Post-level reason classification**
  - Given a set of reasons associated with each stance in an online debate, identify those reason(s) that an author used to back up her stance in her debate post

Beyond Stance Classification

- Oftentimes, it's important to determine not only the author's stance, but also the reasons why she supports or opposes the issue under debate

- **Post-level reason classification**
  - Given a set of reasons associated with each stance in an online debate, identify those reason(s) that an author used to back up her stance in her debate post

- **Sentence-level reason classification**
  - Identify not only the reason(s) an author used in her post, but also the sentence(s) in the post that she used to describe each of her reasons

Reason Classification: An Example

I feel that abortion should remain legal, or rather, parents should have the power to make the decision themselves. Let us take a look from the social perspective. If parents cannot afford to provide for the child, it is understandable that abortion can remain as one of the options.

Reason Classification: An Example

> I feel that abortion should remain legal, or rather, parents should have the power to make the decision themselves. Let us take a look from the social perspective. If parents cannot afford to provide for the child, it is understandable that abortion can remain as one of the options.

- **Stance**: Pro-abortion

Reason Classification: An Example

> I feel that abortion should remain legal, or rather, parents should have the power to make the decision themselves. Let us take a look from the social perspective. If parents cannot afford to provide for the child, it is understandable that abortion can remain as one of the options.

- **Stance**: Pro-abortion
- **Reasons**: Woman's right to abort
  Unwanted babies are threat to their parents' future

Reason Classification: An Example

> I feel that abortion should remain legal, or rather, parents should have the power to make the decision themselves. Let us take a look from the social perspective. If parents cannot afford to provide for the child, it is understandable that abortion can remain as one of the options.

- **Stance**: Pro-abortion
- **Reasons**: Woman's right to abort
  Unwanted babies are threat to their parents' future

Goal

- Examine post- and sentence-level reason classification (RC) in ideological debates
  - RC in ideological debates is potentially more challenging than that in other debate settings such as congressional debates
    - Use of sarcasm and insults (Walker et al., 2012)

Goal

- Examine post- and sentence-level reason classification (RC) in ideological debates
  - RC in ideological debates is potentially more challenging than that in other debate settings such as congressional debates
    - Use of sarcasm and insults (Walker et al., 2012)

- **Specific goal**
  - Examine how automatically computed stance information can be profitably exploited for RC

Goal

- Examine post- and sentence-level reason classification (RC) in ideological debates
  - RC in ideological debates is potentially more challenging than that in other debate settings such as congressional debates
    - Use of sarcasm and insults (Walker et al., 2012)

- **Specific goal**
  - Examine how automatically computed stance information can be profitably exploited for RC
    - **Hypothesis**: the effectiveness of such information would depend in part on the way it is exploited in RC systems
    - Examine a set of stance-supported RC models that differ in terms of modeling sophistication

Plan for the Talk

- Corpus and annotation
- Baseline RC system
- Stance-supported RC systems
- Evaluation

Plan for the Talk

- Corpus and annotation
- Baseline RC system
- Stance-supported RC systems
- Evaluation

Datasets

- 4 datasets
  - collected from http://www.createdebate.com
  - contain debate posts collected from four debate topics

Topic	Posts	"for" %	Average Sequence Length
Support **Abortion**?	1741	54.9	4.1
Support **Gay Rights**?	1376	63.4	4.0
Support **Obama**?	985	53.9	2.6
Legalize **Marijuana**?	626	69.5	2.5

Datasets

- 4 datasets
  - collected from http://www.createdebate.com
  - contain debate posts collected from four debate topics

Topic	Posts	"for" %	Average Sequence Length
Support **Abortion**?	1741	54.9	4.1
Support **Gay Rights**?	1376	63.4	4.0
Support **Obama**?	985	53.9	2.6
Legalize **Marijuana**?	626	69.5	2.5

Datasets

- 4 datasets
  - collected from http://www.createdebate.com
  - contain debate posts collected from four debate topics

Topic	Posts	"for" %	Average Sequence Length
Support **Abortion**?	1741	54.9	4.1
Support **Gay Rights**?	1376	63.4	4.0
Support **Obama**?	985	53.9	2.6
Legalize **Marijuana**?	626	69.5	2.5

Datasets

- 4 datasets
  - collected from http://www.createdebate.com
  - contain debate posts collected from four debate topics

Topic	Posts	"for" %	Average Sequence Length
Support **Abortion**?	1741	54.9	4.1
Support **Gay Rights**?	1376	63.4	4.0
Support **Obama**?	985	53.9	2.6
Legalize **Marijuana**?	626	69.5	2.5

Datasets

- 4 datasets
  - collected from http://www.createdebate.com
  - contain debate posts collected from four debate topics

Topic	Posts	"for" %	Average Sequence Length
Support **Abortion**?	1741	54.9	4.1
Support **Gay Rights**?	1376	63.4	4.0
Support **Obama**?	985	53.9	2.6
Legalize **Marijuana**?	626	69.5	2.5

- Posts are not annotated with reasons

Reason Annotation: 5 Steps

- For each debate topic, the two human annotators
  1. independently examined each post and identified the reasons authors used to support their stances

Reason Annotation: 5 Steps

- For each debate topic, the two human annotators
  1. independently examined each post and identified the reasons authors used to support their stances
  2. discussed and agreed on the reason classes identified

Reason Annotation: 5 Steps

- For each debate topic, the two human annotators
  1. independently examined each post and identified the reasons authors used to support their stances
  2. discussed and agreed on the reason classes identified
  3. independently annotated each post with these reason classes

Reason Annotation: 5 Steps

- For each debate topic, the two human annotators
  1. independently examined each post and identified the reasons authors used to support their stances
  2. discussed and agreed on the reason classes identified
  3. independently annotated each post with these reason classes
     - Labeled each sentence with the set of reasons the author expressed in that sentence
     - Any sentence that does not belong to any reason class was assigned the NONE class

Reason Annotation: 5 Steps

- For each debate topic, the two human annotators
  1. independently examined each post and identified the reasons authors used to support their stances
  2. discussed and agreed on the reason classes identified
  3. independently annotated each post with these reason classes
     - Labeled each sentence with the set of reasons the author expressed in that sentence
     - Any sentence that does not belong to any reason class was assigned the NONE class
  4. collapsed rare reason classes (those that occur in < 2% of the sentences) into the OTHER class

Reason Annotation: 5 Steps

- For each debate topic, the two human annotators
  1. independently examined each post and identified the reasons authors used to support their stances
  2. discussed and agreed on the reason classes identified
  3. independently annotated each post with these reason classes
     - Labeled each sentence with the set of reasons the author expressed in that sentence
     - Any sentence that does not belong to any reason class was assigned the NONE class
  4. collapsed rare reason classes (those that occur in < 2% of the sentences) into the OTHER class
  5. picked and retained only the reason that was highlighted the most for each multi-labeled sentence (< 3% of the sentences)

Reason Annotation: Statistics

Topic	Reason-labeled posts	% Non-NONE sentences	Kappa (sentence)	Kappa (post)
Support **Abortion**?	463	20.4	0.66	0.82
Support **Gay Rights**?	561	29.8	0.63	0.80
Support **Obama**?	447	34.4	0.61	0.78
Legalize **Marijuana**?	432	43.7	0.67	0.83

Reason Annotation: Statistics

Topic	Reason-labeled posts	% Non-NONE sentences	Kappa (sentence)	Kappa (post)
Support **Abortion**?	463	20.4	0.66	0.82
Support **Gay Rights**?	561	29.8	0.63	0.80
Support **Obama**?	447	34.4	0.61	0.78
Legalize **Marijuana**?	432	43.7	0.67	0.83

Reason Annotation: Statistics

Topic	Reason-labeled posts	% Non-NONE sentences	Kappa (sentence)	Kappa (post)
Support **Abortion**?	463	20.4	0.66	0.82
Support **Gay Rights**?	561	29.8	0.63	0.80
Support **Obama**?	447	34.4	0.61	0.78
Legalize **Marijuana**?	432	43.7	0.67	0.83

Reason Annotation: Statistics

Topic	Reason-labeled posts	% Non-NONE sentences	Kappa (sentence)	Kappa (post)
Support **Abortion**?	463	20.4	0.66	0.82
Support **Gay Rights**?	561	29.8	0.63	0.80
Support **Obama**?	447	34.4	0.61	0.78
Legalize **Marijuana**?	432	43.7	0.67	0.83

Reason Annotation: Statistics

Topic	Reason-labeled posts	% Non-NONE sentences	Kappa (sentence)	Kappa (post)
Support **Abortion**?	463	20.4	0.66	0.82
Support **Gay Rights**?	561	29.8	0.63	0.80
Support **Obama**?	447	34.4	0.61	0.78
Legalize **Marijuana**?	432	43.7	0.67	0.83

Reason Annotation: Statistics

Topic	Reason-labeled posts	% Non-NONE sentences	Kappa (sentence)	Kappa (post)
Support **Abortion**?	463	20.4	0.66	0.82
Support **Gay Rights**?	561	29.8	0.63	0.80
Support **Obama**?	447	34.4	0.61	0.78
Legalize **Marijuana**?	432	43.7	0.67	0.83

- substantial post-level agreement; high sentence-level agreement

Reason Annotation: Statistics

Topic	Reason-labeled posts	% Non-NONE sentences	Kappa (sentence)	Kappa (post)
Support **Abortion**?	463	20.4	0.66	0.82
Support **Gay Rights**?	561	29.8	0.63	0.80
Support **Obama**?	447	34.4	0.61	0.78
Legalize **Marijuana**?	432	43.7	0.67	0.83

- Major source of disagreement
  - Annotators, while agreeing on the reason class, differ on how long the text span for a reason should be
  - This hurts sentence-level but not post-level agreement

Plan for the Talk

- Corpus and annotation
- Baseline RC system
- Stance-supported RC systems
- Evaluation

Baseline RC System

- **Sentence-level** reason classifier
  - determines whether a reason is expressed in a sentence
  - if so, assign to the sentence its reason class
- **Training instance creation**
  - Create one for each sentence in each training post
  - Class label: its human-annotated reason label (or NONE if it does not contain a reason)
- **Learning algorithm**
  - Maximum entropy

Baseline RC System: 5 Types of Features

- **N-grams**
  - Unigrams and bigrams
- **Dependency-based features**
  - Argument pairs as features, optionally generalized using POS tags and polarity labels
- **Quotation features**
  - Is the sentence a quote? Does it follow a quote?
- **Positional features**
  - Encode which of the 4 parts of a post the sentence appears in
- **Frame-semantic features**
  - Encode the semantic representation of the sentence's concepts using FrameNet frames

Baseline RC System

- **Sentence-level** reason classifier

Baseline RC System

- **Sentence-level** reason classifier

How can we produce **post-level reason labels** for a post?

Baseline RC System

- **Sentence-level** reason classifier

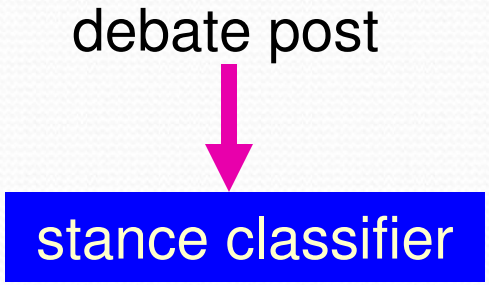
How can we produce **post-level reason labels** for a post?

Take the union of the set of reason labels assigned by the classifier to each of its sentences

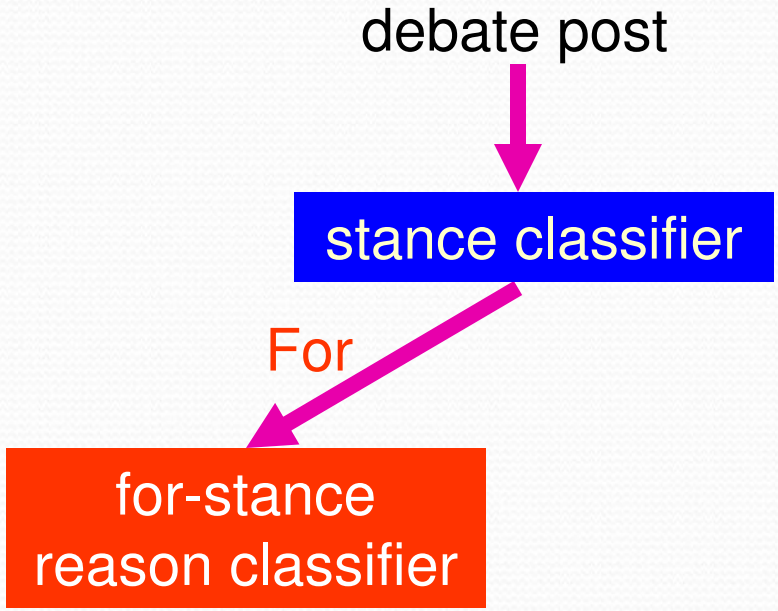
Plan for the Talk

- Corpus and annotation
- Baseline RC system
- Stance-supported RC systems
  - 7 systems with varying levels of modeling sophistication
- Evaluation

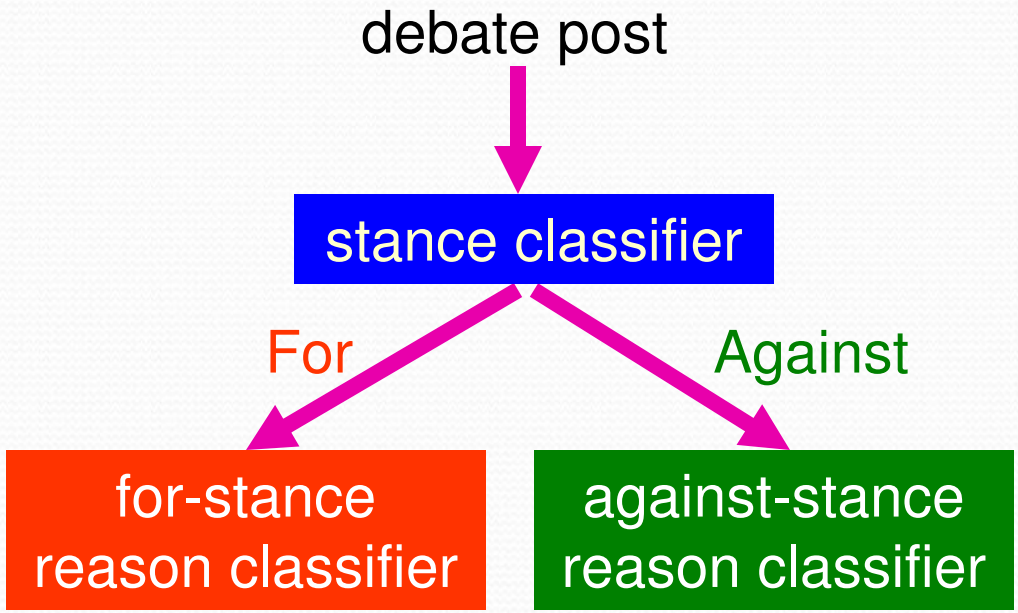
Pipeline Systems: Architecture



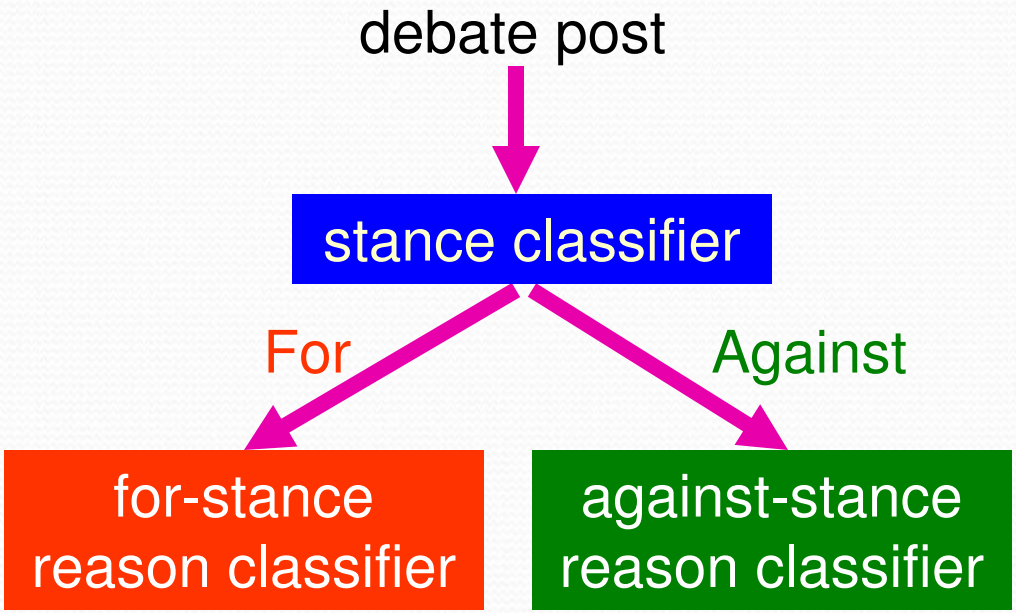
Pipeline Systems: Architecture



Pipeline Systems: Architecture

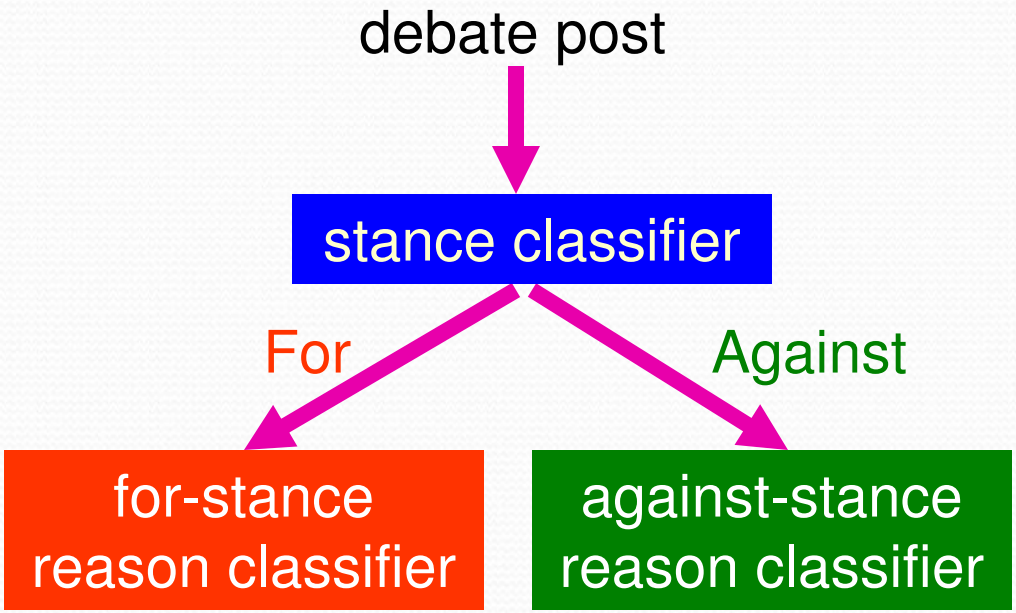


Pipeline Systems: Architecture



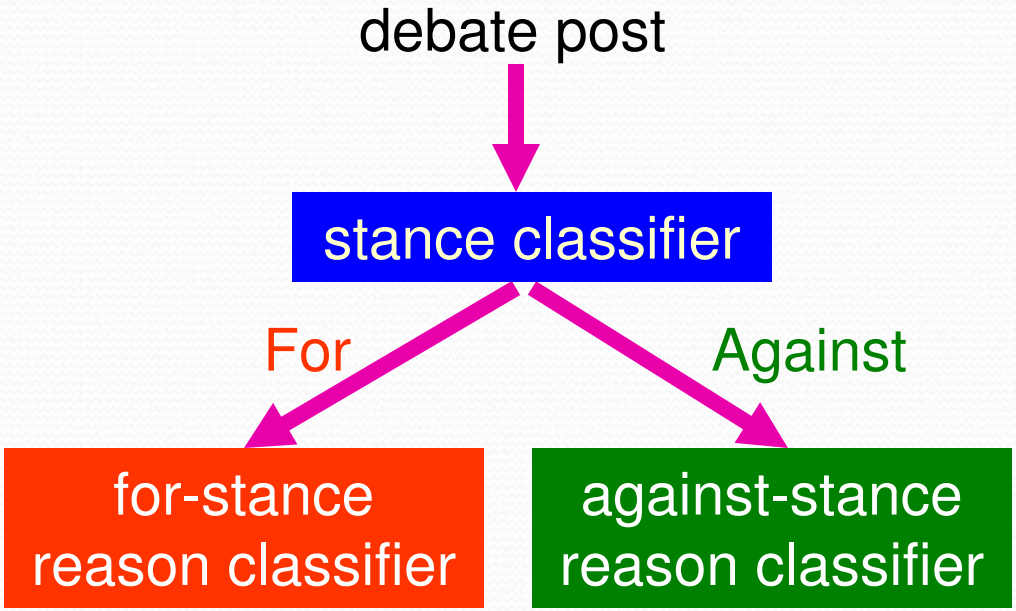
- Each stance-dependent reason classifier trained in the same way as the baseline (stance-independent) reason classifier except:

Pipeline Systems: Architecture



- Each stance-dependent reason classifier trained in the same way as the baseline (stance-independent) reason classifier except:
  - For-stance classifier: trained only on for-posts
  - Against-stance classifier: trained only on the against-posts

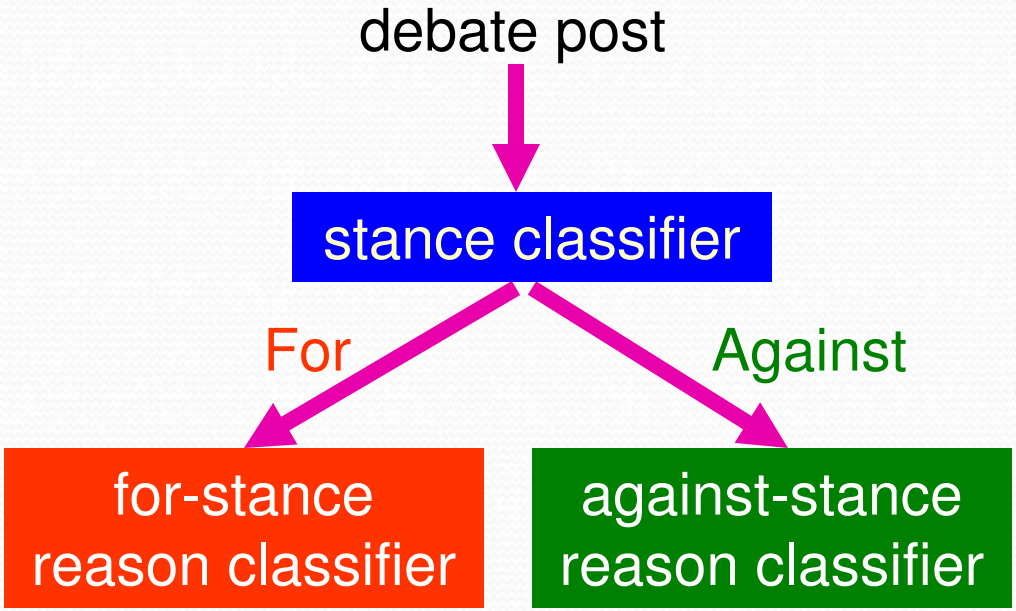
Pipeline Systems: Architecture



- Each stance-dependent reason classifier estimates

$$P(r \mid s, t)$$

Pipeline Systems: Architecture

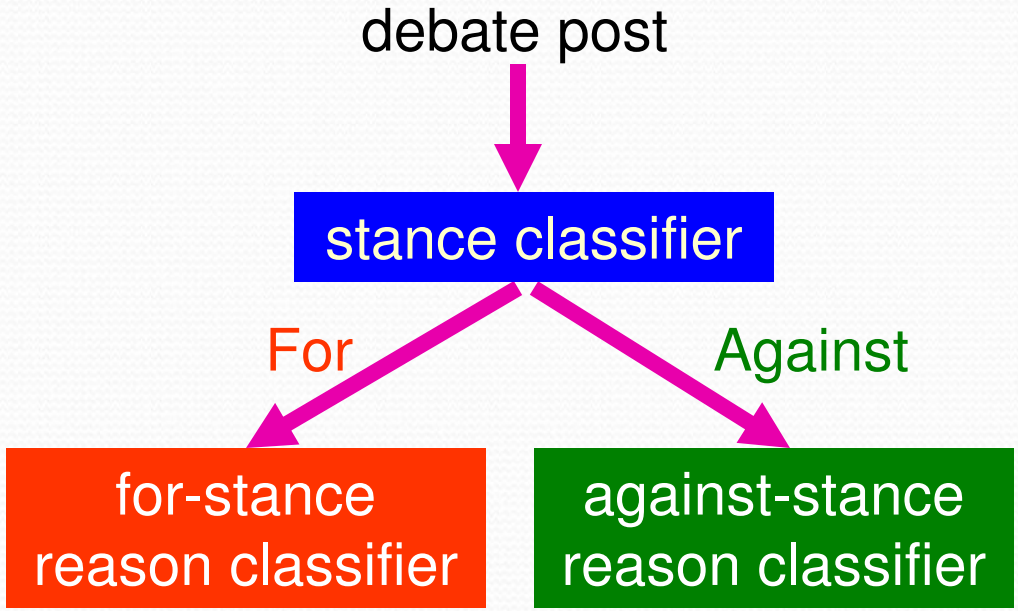


- Each stance-dependent reason classifier estimates

$$P(r \mid s, t)$$

reason

Pipeline Systems: Architecture

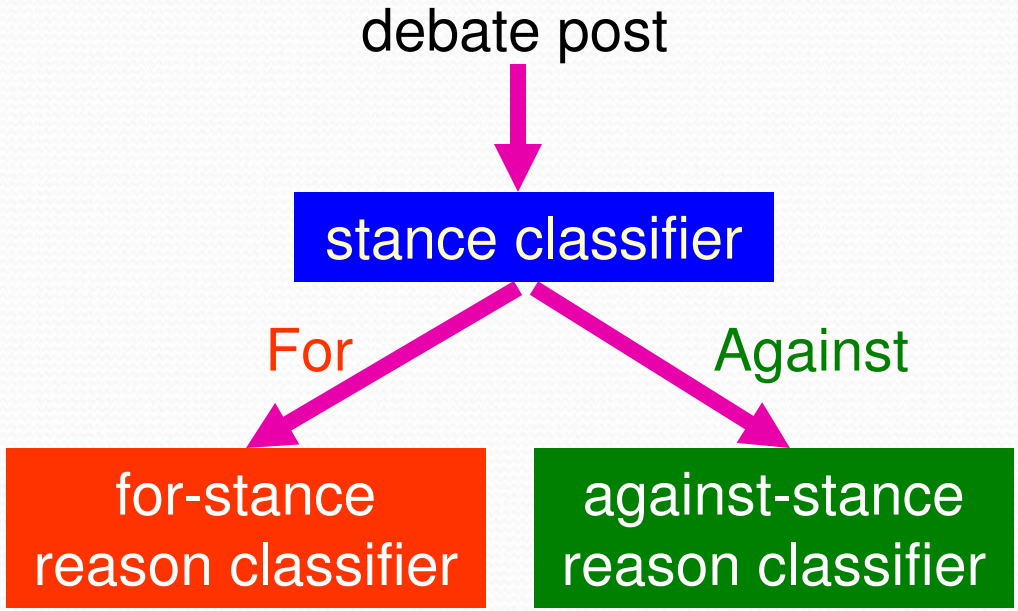


- Each stance-dependent reason classifier estimates

$$P(r \mid s, t)$$

reason stance

Pipeline Systems: Architecture

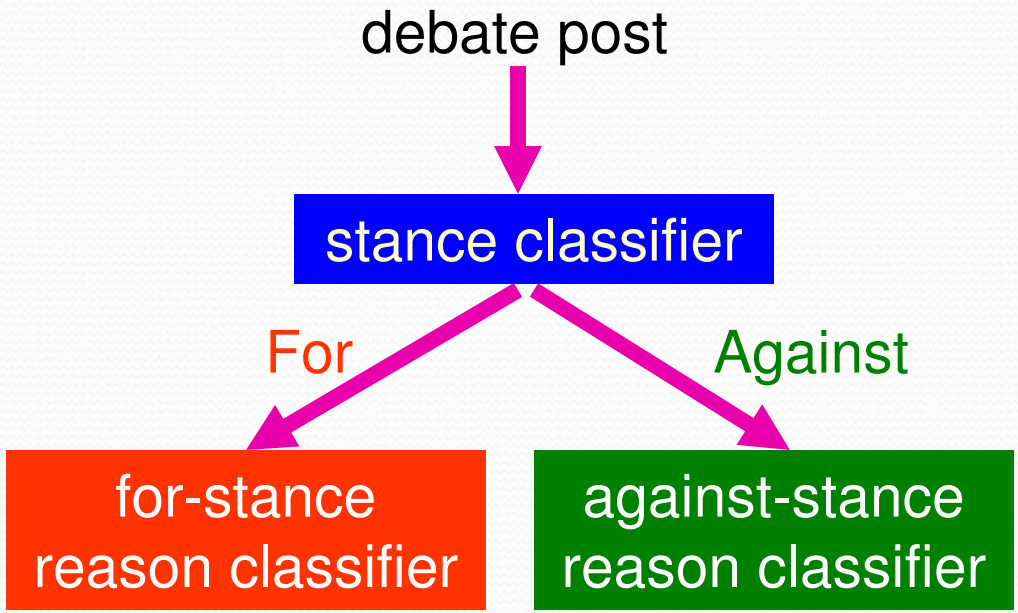


- Each stance-dependent reason classifier estimates

$$P(r \mid s, t)$$

reason stance sentence

Pipeline Systems: Architecture



How to train the stance classifier?

Training the Stance Classifier: Method 1

- Binary classifier that assigns a stance label (for/against) to each debate post p independently of other posts

Training the Stance Classifier: Method 1

- Binary classifier that assigns a stance label (for/against) to each debate post p independently of other posts
  - Each training instance corresponds to a debate post
  - **Features**: as those used for reason classification
  - **Learning algorithm**: maximum entropy
    - estimates $P(s \mid p)$

Training the Stance Classifier: Method 2

Training the Stance Classifier: Method 2

- Same as Method 1, except we recast stance classification (SC) as a sequence learning problem
  - Input: a post sequence $P_S = (p_1, p_2, ..., p_n)$
  - Output: a stance sequence $S = (s_1, s_2, ..., s_n)$
- A sequence model estimates $P(S \mid P_S)$ and returns the most probable stance sequence

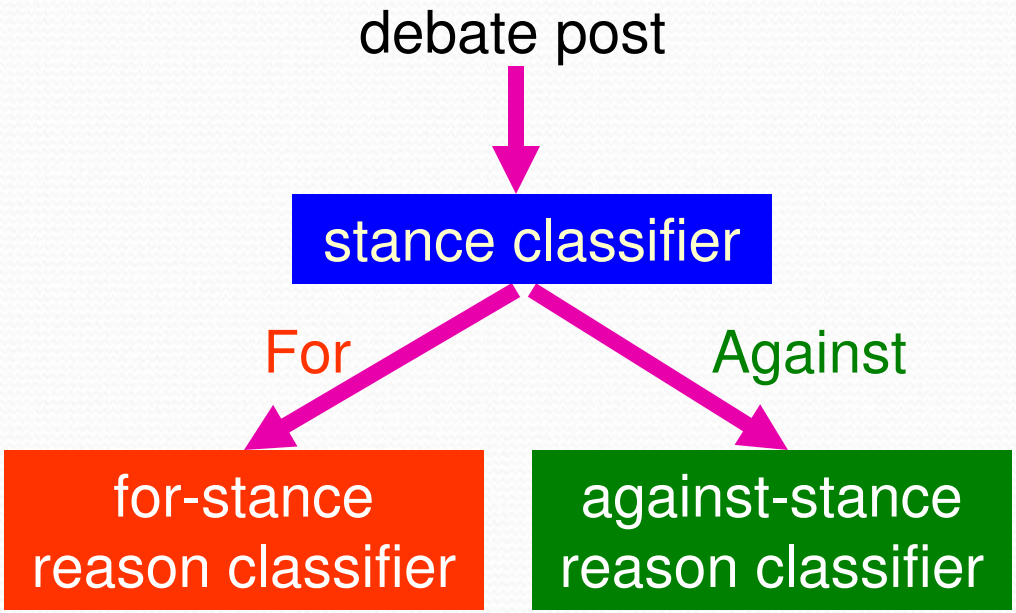
Training the Stance Classifier: Method 2

- Same as Method 1, except we recast stance classification (SC) as a sequence learning problem
  - Input: a post sequence $P_S = (p_1, p_2, ..., p_n)$
  - Output: a stance sequence $S = (s_1, s_2, ..., s_n)$
- A sequence model estimates $P(S \mid P_S)$ and returns the most probable stance sequence
- **Motivation**
  - Since a post in a post sequence is a reply to its parent post, its label should be determined in dependent relation to its parent's

Training the Stance Classifier: Method 2

- Same as Method 1, except we recast stance classification (SC) as a sequence learning problem
  - Input: a post sequence $P_S = (p_1, p_2, ..., p_n)$
  - Output: a stance sequence $S = (s_1, s_2, ..., s_n)$
- A sequence model estimates $P(S \mid P_S)$ and returns the most probable stance sequence
- **Motivation**
  - Since a post in a post sequence is a reply to its parent post, its label should be determined in dependent relation to its parent's
- To train sequence models, we employ MEMM

Pipeline Systems: Architecture



- 2 pipeline systems
  - **P1**: classifies the stance of each post independently
  - **P2**: uses sequence learning for stance classification

Joint Inference

- **Motivation:** in the pipeline systems, errors may propagate from the SC system to the RC system
  - incorrectly labeled stance → incorrectly labeled reasons

Joint Inference

- **Motivation:** in the pipeline systems, errors may propagate from the SC system to the RC system
  - incorrectly labeled stance → incorrectly labeled reasons

- Use ILP to perform joint inference over the outputs of the independently trained stance classifier and reason classifier

Joint Inference

- **Motivation:** in the pipeline systems, errors may propagate from the SC system to the RC system
  - incorrectly labeled stance → incorrectly labeled reasons

- Use ILP to perform joint inference over the outputs of the independently trained stance classifier and reason classifier

Use the stance classifier in P2

Joint Inference

- **Motivation:** in the pipeline systems, errors may propagate from the SC system to the RC system
  - incorrectly labeled stance → incorrectly labeled reasons

- Use ILP to perform joint inference over the outputs of the independently trained stance classifier and reason classifier

Use the stance classifier in P2

Use the baseline RC system

Joint Inference

- **Motivation:** in the pipeline systems, errors may propagate from the SC system to the RC system
  - incorrectly labeled stance → incorrectly labeled reasons

- Use ILP to perform joint inference over the outputs of the independently-trained stance classifier and reason classifier

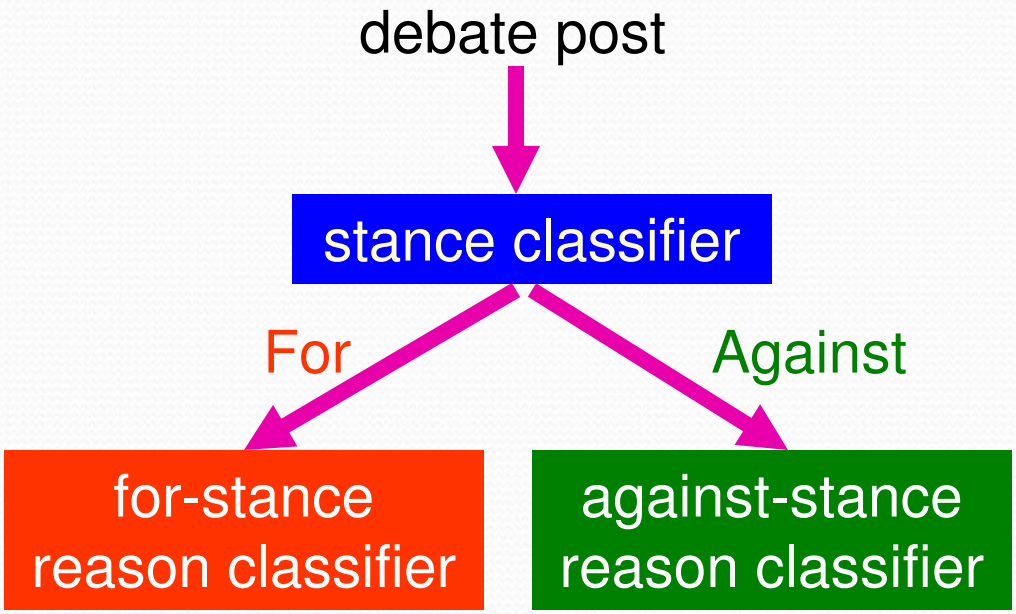
Use the stance classifier in P2

Use the baseline RC system

- ILP constraints
  - If a post contains a For reason, its stance label should be For
  - A post stance-labeled as For should contain a For reason
  - Similar constraints are defined for the Against label

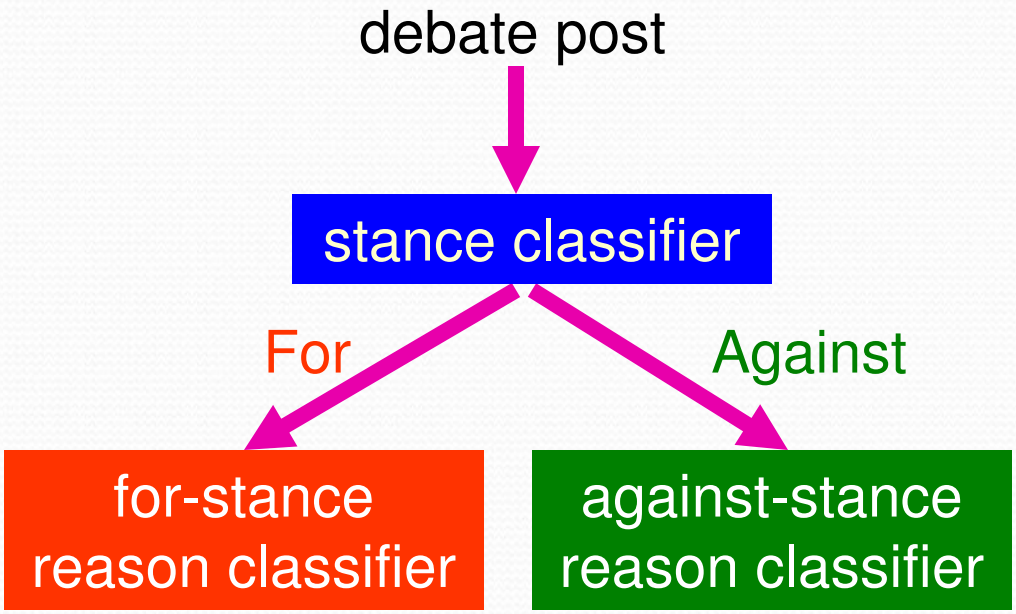
Joint Maximization

- Another way to address the error propagation problem inherent in the pipeline systems



Joint Maximization

- Another way to address the error propagation problem inherent in the pipeline systems

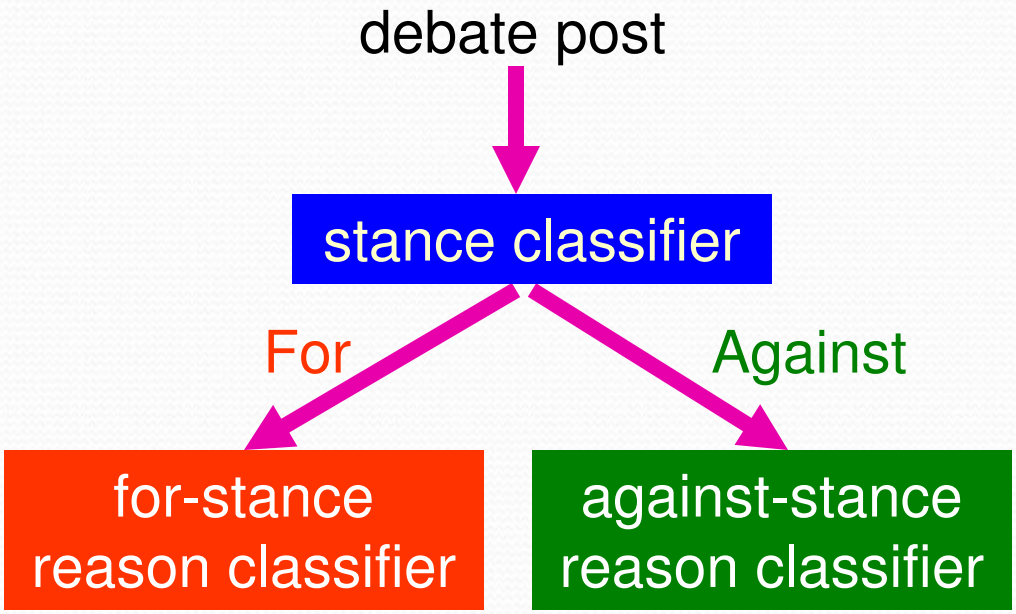


$P(s \mid p)$

$P(r \mid s,t)$

Joint Maximization

- Another way to address the error propagation problem inherent in the pipeline systems



$P(s \mid p)$

$P(r \mid s,t)$

- **Goal**: Determine the stance label using not only the stance classifier but also the stance-dependent reason classifiers
  - Let the reason classifiers influence the choice of the stance

Joint Maximization

- **Goal**: Find the stance s that maximizes

$$P(s \mid p) \prod_i P(r_i \mid s, t_i)$$

Joint Maximization

- **Goal**: Find the stance s that maximizes

$$P(s \mid p) \prod_i P(r_i \mid s, t_i)$$

stance classifier
used in P1

Joint Maximization

- **Goal**: Find the stance s that maximizes

$$P(s \mid p) \prod_i P(r_i \mid s, t_i)$$

stance classifier used in P1

stance-dependent reason classifier

Joint Maximization

- **Goal**: Find the stance s that maximizes

$$P(s \mid p) \prod_i P(r_i \mid s, t_i)$$

stance classifier used in P1

stance-dependent reason classifier

over all sentences i in the post

Joint Maximization Model J1

- **Goal**: Find the stance s that maximizes

$$P(s \mid p) \prod_i P(r_i \mid s, t_i)$$

stance classifier used in P1

over all sentences i in the post

stance-dependent reason classifier

Joint Maximization Model J2

Joint Maximization Model J2

- Same as J1, except that we recast SC as sequence labeling

Joint Maximization Model J2

- Same as J1, except that we recast SC as sequence labeing

- Given a post sequence $P_S = (p_1, p_2, ..., p_n)$, J2 finds the stance sequence $S = (s_1, s_2, ..., s_n)$ and reasons $R = (R_1, R_2, ..., R_n)$ that jointly maximize

$$P(R, S \mid P_s)$$

Joint Maximization Model J2

- Same as J1, except that we recast SC as sequence labeing

- Given a post sequence $P_S = (p_1, p_2, ..., p_n)$, J2 finds the stance sequence $S = (s_1, s_2, ..., s_n)$ and reasons $R = (R_1, R_2, ..., R_n)$ that jointly maximize

$$P(R, S \mid P_s)$$

  - R_i is the list of reasons assigned to the sentences in post *i*

Joint Maximization Model J2

- Same as J1, except that we recast SC as sequence labeing
- Given a post sequence $P_S = (p_1, p_2, ..., p_n)$, J2 finds the stance sequence $S = (s_1, s_2, ..., s_n)$ and reasons $R = (R_1, R_2, ..., R_n)$ that jointly maximizes

$$P(R, S \mid P_s)$$

  - R_i is the list of reasons assigned to the sentences in post *i*
- The R and S that jointly maximize $P(R, S \mid P_s)$ can be found efficiently using dynamic programming

Joint Maximization Model J3

- **Observation**
  - A post in a post sequence is a reply to its preceding post
  - In many cases, a reply is a rebuttal to the preceding post
    - Author argues against the points raised in the preceding post and provides her reasons for the opposing stance

Joint Maximization Model J3

- **Observation**
  - A post in a post sequence is a reply to its preceding post
  - In many cases, a reply is a rebuttal to the preceding post
    - Author argues against the points raised in the preceding post and provides her reasons for the opposing stance
- **Hypothesis**
  - The reasons mentioned in the preceding post could be useful for predicting the reasons in the current post

Joint Maximization Model J3

- **Observation**
  - A post in a post sequence is a reply to its preceding post
  - In many cases, a reply is a rebuttal to the preceding post
    - Author argues against the points raised in the preceding post and provides her reasons for the opposing stance
- **Hypothesis**
  - The reasons mentioned in the preceding post could be useful for predicting the reasons in the current post
- **Joint maximization model J3**
  - Build on top of J2: augment the feature set of the stance-dependent reason classifiers with a set of reason features
    - One binary feature for each reason class indicating presence/absence

Recap: 7 Stance-supported RC systems

Recap: 7 Stance-supported RC systems

- 2 pipeline systems (SC followed by RC)
  - **P1**: stance label of each post determined independently
  - **P2**: sequence labeling for SC

Recap: 7 Stance-supported RC systems

- 2 pipeline systems (SC followed by RC)
  - **P1**: stance label of each post determined independently
  - **P2**: sequence labeling for SC
- joint inference using ILP

Recap: 7 Stance-supported RC systems

- 2 pipeline systems (SC followed by RC)
  - **P1**: stance label of each post determined independently
  - **P2**: sequence labeling for SC
- joint inference using ILP
- 3 joint maximization systems
  - **J1**: joint version of P1
  - **J2**: joint version of P2
  - **J3**: J2 where the features for training the reason classifiers are augmented with the reason labels predicted for previous post

Plan for the Talk

- Corpus and annotation
- Baseline RC system
- Stance-supported RC systems
- Evaluation

Evaluation: Goal

- Evaluate the stance-supported RC systems

Evaluation: Goal

- Evaluate the stance-supported RC systems

Will RC performance improve as we employ more sophisticated methods for modeling stances and reasons?

Evaluation: Goal

- Evaluate the stance-supported RC systems

Will RC performance improve as we employ more sophisticated methods for modeling stances and reasons?

Can SC performance improve when SC is jointly modeled with RC?

Evaluation: Setup

- 5-fold cross validation for each debate topic

Evaluation: Setup

- 5-fold cross validation for each debate topic

- Evaluation metrics
  - SC: accuracy
  - RC: F-score micro-averaged over all reason classes except the NONE class
    - **Sentence-level RC** scores computed over sentences
    - **Post-level** RC scores computed over posts

Evaluation: Setup

- 5-fold cross validation for each debate topic
  - **No** cross-domain training/testing

- Evaluation metrics
  - SC: accuracy
  - RC: F-score micro-averaged over all reason classes except the NONE class
    - **Sentence-level RC** scores computed over sentences
    - **Post-level** RC scores computed over posts
      - Post-level reason labels are derived from the sentence-level reason labels

Evaluation: RC F-scores

System	Abortion		Gay Rights		Obama		Marijuana	
	Sent.	Post	Sent.	Post	Sent.	Post	Sent.	Post
Baseline	32.7	45.0	23.3	40.5	19.5	31.5	28.7	44.2

Evaluation: RC F-scores

System	Abortion		Gay Rights		Obama		Marijuana	
	Sent.	Post	Sent.	Post	Sent.	Post	Sent.	Post
Baseline	32.7	45.0	23.3	40.5	19.5	31.5	28.7	44.2

Evaluation: RC F-scores

System	Abortion		Gay Rights		Obama		Marijuana	
	Sent.	Post	Sent.	Post	Sent.	Post	Sent.	Post
Baseline	32.7	45.0	23.3	40.5	19.5	31.5	28.7	44.2

Evaluation: RC F-scores

System	Abortion		Gay Rights		Obama		Marijuana	
	Sent.	Post	Sent.	Post	Sent.	Post	Sent.	Post
Baseline	32.7	45.0	23.3	40.5	19.5	31.5	28.7	44.2

Evaluation: Baseline

System	Abortion		Gay Rights		Obama		Marijuana	
	Sent.	Post	Sent.	Post	Sent.	Post	Sent.	Post
Baseline	32.7	45.0	23.3	40.5	19.5	31.5	28.7	44.2

- Sentence-level RC scores are lower than post-level RC scores

Evaluation: Pipeline Systems

System	Abortion		Gay Rights		Obama		Marijuana	
	Sent.	Post	Sent.	Post	Sent.	Post	Sent.	Post
Baseline	32.7	45.0	23.3	40.5	19.5	31.5	28.7	44.2
P1	34.5	46.3	24.5	43.2	20.3	33.5	30.5	47.3
P2	36.1	47.7	26.6	45.5	21.1	34.4	32.9	48.8

Evaluation: Pipeline Systems

System	Abortion		Gay Rights		Obama		Marijuana	
	Sent.	Post	Sent.	Post	Sent.	Post	Sent.	Post
Baseline	32.7	45.0	23.3	40.5	19.5	31.5	28.7	44.2
P1	34.5	46.3	24.5	43.2	20.3	33.5	30.5	47.3
P2	36.1	47.7	26.6	45.5	21.1	34.4	32.9	48.8

- P1 and P2 significantly outperform Baseline
  - RC can be improved even when stance information is incorporated in a simple manner

Evaluation: Pipeline Systems

System	Abortion		Gay Rights		Obama		Marijuana	
	Sent.	Post	Sent.	Post	Sent.	Post	Sent.	Post
Baseline	32.7	45.0	23.3	40.5	19.5	31.5	28.7	44.2
P1	34.5	46.3	24.5	43.2	20.3	33.5	30.5	47.3
P2	36.1	47.7	26.6	45.5	21.1	34.4	32.9	48.8

- P2 outperforms P1
  - Better SC leads to better RC

Evaluation: Joint Inference via ILP

System	Abortion		Gay Rights		Obama		Marijuana	
	Sent.	Post	Sent.	Post	Sent.	Post	Sent.	Post
Baseline	32.7	45.0	23.3	40.5	19.5	31.5	28.7	44.2
P1	34.5	46.3	24.5	43.2	20.3	33.5	30.5	47.3
P2	36.1	47.7	26.6	45.5	21.1	34.4	32.9	48.8
ILP	36.5	48.4	28.0	46.7	22.8	35.0	33.1	48.9

Evaluation: Joint Inference via ILP

System	Abortion		Gay Rights		Obama		Marijuana	
	Sent.	Post	Sent.	Post	Sent.	Post	Sent.	Post
Baseline	32.7	45.0	23.3	40.5	19.5	31.5	28.7	44.2
P1	34.5	46.3	24.5	43.2	20.3	33.5	30.5	47.3
P2	36.1	47.7	26.6	45.5	21.1	34.4	32.9	48.8
ILP	36.5	48.4	28.0	46.7	22.8	35.0	33.1	48.9

- ILP beats P2 on ABO and GAY and achieves the same level of performance as P2 on OBA and MAR
  - Joint inference is no worse (and sometimes better) than pipeline learning when exploiting stance information for RC

Evaluation: Joint Maximization Systems

System	Abortion		Gay Rights		Obama		Marijuana	
	Sent.	Post	Sent.	Post	Sent.	Post	Sent.	Post
Baseline	32.7	45.0	23.3	40.5	19.5	31.5	28.7	44.2
P1	34.5	46.3	24.5	43.2	20.3	33.5	30.5	47.3
P2	36.1	47.7	26.6	45.5	21.1	34.4	32.9	48.8
ILP	36.5	48.4	28.0	46.7	22.8	35.0	33.1	48.9
J1	36.0	47.6	26.7	45.6	23.1	35.7	33.3	49.2
J2	37.9	50.6	29.6	48.5	24.5	37.1	34.5	50.5

Evaluation: Joint Maximization Systems

System	Abortion		Gay Rights		Obama		Marijuana	
	Sent.	Post	Sent.	Post	Sent.	Post	Sent.	Post
Baseline	32.7	45.0	23.3	40.5	19.5	31.5	28.7	44.2
P1	34.5	46.3	24.5	43.2	20.3	33.5	30.5	47.3
P2	36.1	47.7	26.6	45.5	21.1	34.4	32.9	48.8
ILP	36.5	48.4	28.0	46.7	22.8	35.0	33.1	48.9
J1	36.0	47.6	26.7	45.6	23.1	35.7	33.3	49.2
J2	37.9	50.6	29.6	48.5	24.5	37.1	34.5	50.5

- J2 outperforms J1
  - Better SC leads to better RC

Evaluation: Joint Maximization Systems

System	Abortion		Gay Rights		Obama		Marijuana	
	Sent.	Post	Sent.	Post	Sent.	Post	Sent.	Post
Baseline	32.7	45.0	23.3	40.5	19.5	31.5	28.7	44.2
P1	34.5	46.3	24.5	43.2	20.3	33.5	30.5	47.3
P2	36.1	47.7	26.6	45.5	21.1	34.4	32.9	48.8
ILP	36.5	48.4	28.0	46.7	22.8	35.0	33.1	48.9
J1	36.0	47.6	26.7	45.6	23.1	35.7	33.3	49.2
J2	37.9	50.6	29.6	48.5	24.5	37.1	34.5	50.5

- J1 and J2 outperform their pipeline counterparts, P1 and P2
  - Joint learning is a better way to incorporate stance information for RC than pipeline learning

Evaluation: Joint Maximization Systems

System	Abortion		Gay Rights		Obama		Marijuana	
	Sent.	Post	Sent.	Post	Sent.	Post	Sent.	Post
Baseline	32.7	45.0	23.3	40.5	19.5	31.5	28.7	44.2
P1	34.5	46.3	24.5	43.2	20.3	33.5	30.5	47.3
P2	36.1	47.7	26.6	45.5	21.1	34.4	32.9	48.8
ILP	36.5	48.4	28.0	46.7	22.8	35.0	33.1	48.9
J1	36.0	47.6	26.7	45.6	23.1	35.7	33.3	49.2
J2	37.9	50.6	29.6	48.5	24.5	37.1	34.5	50.5
J3	**39.5**	**52.3**	**31.4**	**49.8**	**25.1**	**38.0**	**35.1**	**51.1**

Evaluation: Joint Maximization Systems

System	Abortion		Gay Rights		Obama		Marijuana	
	Sent.	Post	Sent.	Post	Sent.	Post	Sent.	Post
Baseline	32.7	45.0	23.3	40.5	19.5	31.5	28.7	44.2
P1	34.5	46.3	24.5	43.2	20.3	33.5	30.5	47.3
P2	36.1	47.7	26.6	45.5	21.1	34.4	32.9	48.8
ILP	36.5	48.4	28.0	46.7	22.8	35.0	33.1	48.9
J1	36.0	47.6	26.7	45.6	23.1	35.7	33.3	49.2
J2	37.9	50.6	29.6	48.5	24.5	37.1	34.5	50.5
J3	**39.5**	**52.3**	**31.4**	**49.8**	**25.1**	**38.0**	**35.1**	**51.1**

- J3 exploits reason labels predicted for the previous post

Evaluation: Joint Maximization Systems

System	Abortion		Gay Rights		Obama		Marijuana	
	Sent.	Post	Sent.	Post	Sent.	Post	Sent.	Post
Baseline	32.7	45.0	23.3	40.5	19.5	31.5	28.7	44.2
P1	34.5	46.3	24.5	43.2	20.3	33.5	30.5	47.3
P2	36.1	47.7	26.6	45.5	21.1	34.4	32.9	48.8
ILP	36.5	48.4	28.0	46.7	22.8	35.0	33.1	48.9
J1	36.0	47.6	26.7	45.6	23.1	35.7	33.3	49.2
J2	37.9	50.6	29.6	48.5	24.5	37.1	34.5	50.5
J3	**39.5**	**52.3**	**31.4**	**49.8**	**25.1**	**38.0**	**35.1**	**51.1**

- J3 significantly beats J2 for ABO and GAY, and yields small, statistically insignificant gains for OBA and MAR
  - Reasons predicted for the previous post provide useful info

Evaluation: Joint Maximization Systems

System	Abortion		Gay Rights		Obama		Marijuana	
	Sent.	**Post**	**Sent.**	**Post**	**Sent.**	**Post**	**Sent.**	**Post**
Baseline	32.7	45.0	23.3	40.5	19.5	31.5	28.7	44.2
P1	34.5	46.3	24.5	43.2	20.3	33.5	30.5	47.3
P2	36.1	47.7	26.6	45.5	21.1	34.4	32.9	48.8
ILP	36.5	48.4	28.0	46.7	22.8	35.0	33.1	48.9
J1	36.0	47.6	26.7	45.6	23.1	35.7	33.3	49.2
J2	37.9	50.6	29.6	48.5	24.5	37.1	34.5	50.5
J3	**39.5**	**52.3**	**31.4**	**49.8**	**25.1**	**38.0**	**35.1**	**51.1**

- J3 significantly beats Baseline by an average of 6.7 and 7.5 points at the sentence and post levels respectively

Evaluation: Joint Maximization Systems

System	Abortion		Gay Rights		Obama		Marijuana	
	Sent.	Post	Sent.	Post	Sent.	Post	Sent.	Post
Baseline	32.7	45.0	23.3	40.5	19.5	31.5	28.7	44.2
P1	34.5	46.3	24.5	43.2	20.3	33.5	30.5	47.3
P2	36.1	47.7	26.6	45.5	21.1	34.4	32.9	48.8
ILP	36.5	48.4	28.0	46.7	22.8	35.0	33.1	48.9
J1	36.0	47.6	26.7	45.6	23.1	35.7	33.3	49.2
J2	37.9	50.6	29.6	48.5	24.5	37.1	34.5	50.5
J3	**39.5**	**52.3**	**31.4**	**49.8**	**25.1**	**38.0**	**35.1**	**51.1**

- Results suggest that the usefulness of automatically computed stance information depends in part on the way it is exploited

Evaluation: Goal

- Evaluate the stance-supported RC systems

Will RC performance improve as we employ more sophisticated methods for modeling stances and reasons?

Can SC performance improve when SC is jointly modeled with RC?

Evaluation: SC accuracies

System	Abortion	Gay Rights	Obama	Marijuana
Baseline	--	--	--	--
P1	62.8	63.4	61.0	67.2
P2	65.1	64.2	63.8	68.5
ILP	65.2	64.6	63.6	68.8
J1	62.5	64.0	61.2	67.8
J2	65.9	65.3	63.5	68.7
J3	**66.3**	**65.7**	**64.0**	**69.0**

Evaluation: SC accuracies

System	Abortion	Gay Rights	Obama	Marijuana
Baseline	--	--	--	--
P1	62.8	63.4	61.0	67.2
P2	65.1	64.2	63.8	68.5
ILP	65.2	64.6	63.6	68.8
J1	62.5	64.0	61.2	67.8
J2	65.9	65.3	63.5	68.7
J3	**66.3**	**65.7**	**64.0**	**69.0**

- J3 achieved the best SC accuracies for all four datasets

Evaluation: SC accuracies

System	Abortion	Gay Rights	Obama	Marijuana
Baseline	--	--	--	--
P1	62.8	63.4	61.0	67.2
P2	65.1	64.2	63.8	68.5
ILP	65.2	64.6	63.6	68.8
J1	62.5	64.0	61.2	67.8
J2	65.9	65.3	63.5	68.7
J3	**66.3**	**65.7**	**64.0**	**69.0**

- Comparing J3 (best joint model) and P2 (best pipeline model),
  - J3 is better, sometimes significantly so, than P2 on all datasets
  - Joint modeling of SC and RC has a positive impact on SC

Summary

- Examined the task of reason classification in ideological debates
- Demonstrated on our reason-annotated corpus that sophisticated models of stances and reasons can indeed yield more accurate reason and stance classification results than their simpler counterparts
- Reason classification remains a challenging task
  - Best post-level F-scores are in the low 50s.