favor

against

neither

Detecting Stance in Tweets And Analyzing its Interaction with Sentiment

Parinaz Sobhani[1], Saif M. Mohammad[2], and Svetlana Kiritchenko[2]

[1]University of Ottawa, [2]National Research Council Canada

Stance Detection

Automatically determining from text whether the author is in favor of, against, or neutral towards a proposition or target.

- The target may be:
  - a person (say, Donald Trump)
  - an organization (say, American Association of Candy Technologists)
  - an issue (say, Legalization of Abortion)
  - or any entity

For example, can a system infer from Barack Obama's speeches that he is in favor of stricter gun laws in the US?

Applications of automatic stance detection:
information retrieval, text summarization, textual entailment, social media analytics.

The Task

favor against neither

Given a tweet text and a target determine whether:

- the tweeter is in favor of the given target
- the tweeter is against the given target
- neither inference is likely

Example 1:

Target: Jeb Bush
Tweet: Jeb Bush is the only sane candidate in this republican lineup.

Systems have to deduce that the tweeter is likely in favor of the target.

Example 2:

Target: pro-life movement
Tweet: The pregnant are more than walking incubators, and have rights!

Systems have to deduce that the tweeter is likely against the target.

Subtleties of Stance Detection:

Stance vs. Sentiment

- positive language $\nRightarrow$ favor; negative language $\nRightarrow$ against
- the target can be expressed in different ways
  - impacts whether the instance is labeled favor or against
- the target of interest may not be mentioned in the text
  - especially for issue targets: legalization of abortion
- the target of interest may not be the target of opinion in the text

Example:

Target: Donald Trump

Tweet: Hillary Clinton is the only sane candidate in this election #rightchoice

The target of opinion in the tweet is Hillary Clinton.
Nonetheless, we can infer that the tweeter is likely unfavorable towards Donald Trump.

Subtleties of Stance Detection:
Neutral Stance

- lack of evidence for ‘favor’ or ‘against’
  - does not imply neutral stance
  - implies that one cannot deduce stance
- the number of tweets from which we can infer neutral stance is expected to be small

  Example:

  Target: Hillary Clinton
  Tweet: Hillary Clinton has some strengths and some weaknesses.

Thus, we merge all classes other than ‘favor’ and ‘against’ into one ‘neither’ class.

Dataset Creation

Selecting Tweet-Target Pairs

Selected as targets a small subset of entities that were:
(a) routinely discussed on Twitter by US residents at the time of data collection and (b) were controversial:

- Atheism
- Climate Change is a Real Concern
- Donald Trump
- Feminist Movement
- Hillary Clinton
- Legalization of Abortion

Selecting Tweet-Target Pairs (continued)

- created a small list of hashtags that people use when tweeting about the targets: query hashtags.
- polled the Twitter API to collect close to 2 million tweets containing these hashtags
- discarded tweets with URLs
- kept only those tweets where the query hashtags appeared at the end
- removed the query hashtags from the tweets to exclude obvious cues for the classification task
    - can sometimes result in tweets that do not explicitly mention the target

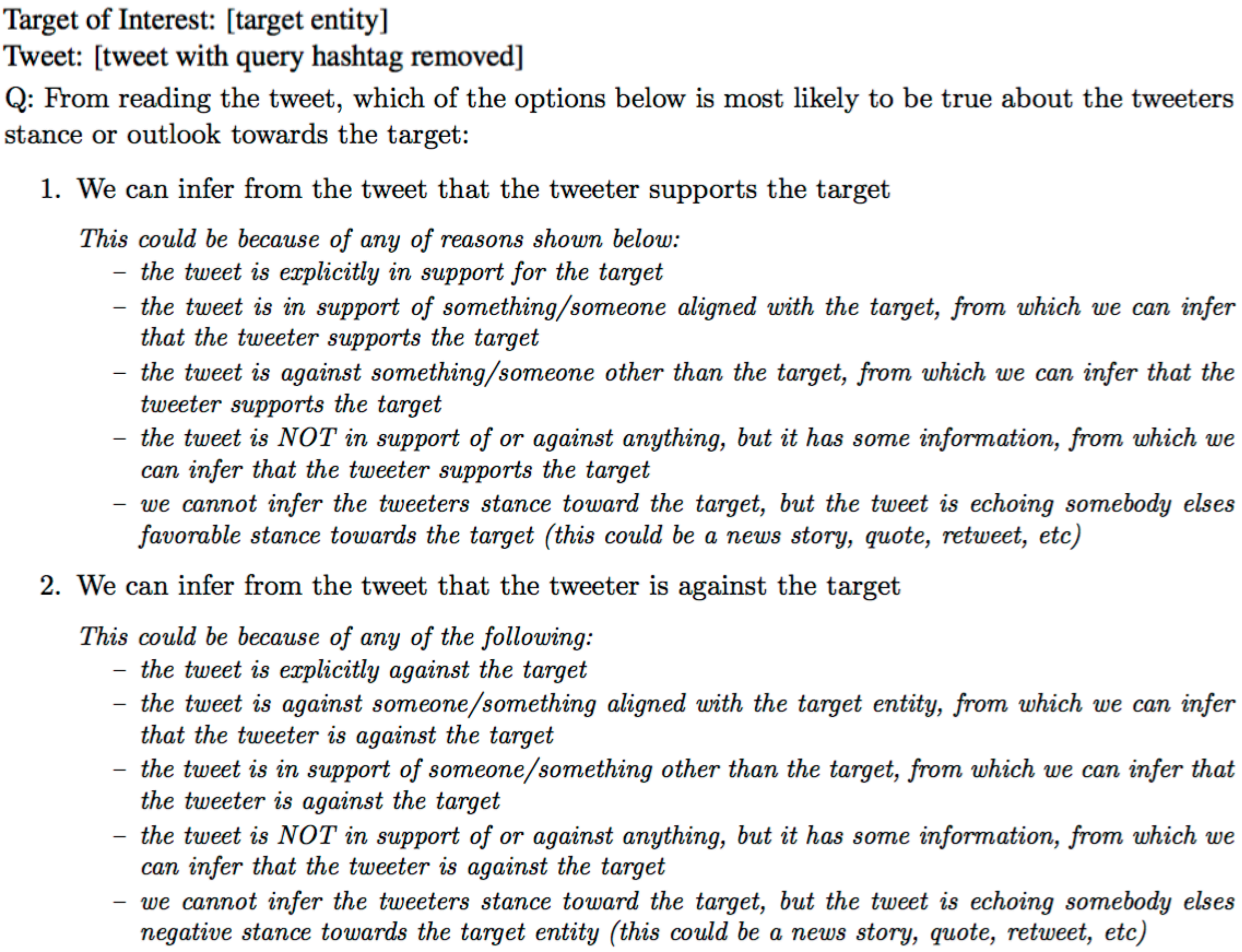
      Target: Hillary Clinton

      Tweet: Benghazi questions need to be answered #Jeb2016 #HillNo

      Removal of #HillNo leaves no mention of Hillary Clinton.

Data Annotation

Crowdsourced



Target of Interest: [target entity]
Tweet: [tweet with query hashtag removed]

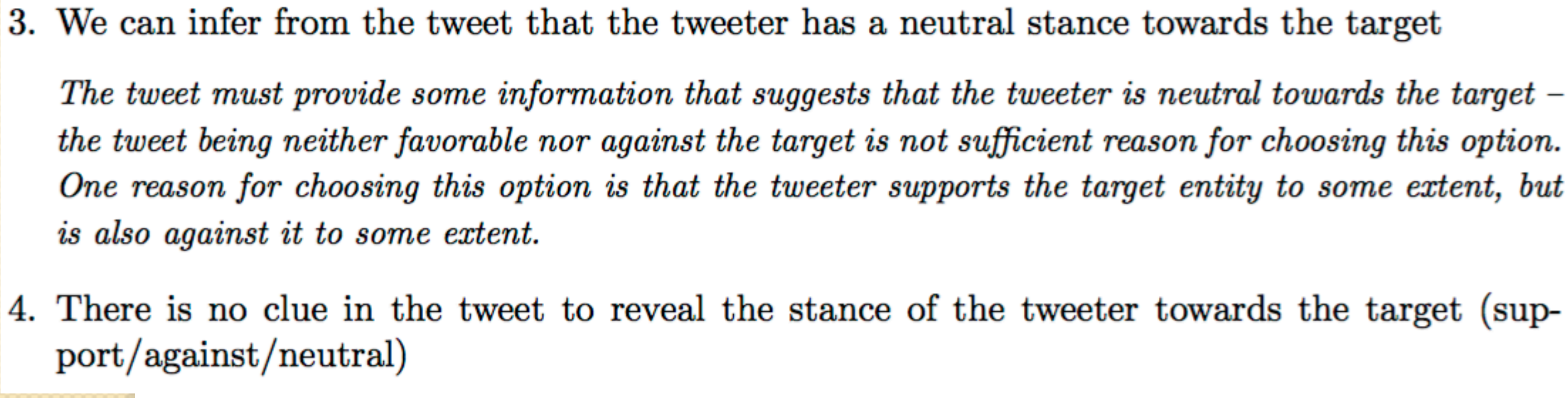
Q: From reading the tweet, which of the options below is most likely to be true about the tweeters stance or outlook towards the target:

1. We can infer from the tweet that the tweeter supports the target

   This could be because of any of reasons shown below:
   - *the tweet is explicitly in support for the target*
   - *the tweet is in support of something/someone aligned with the target, from which we can infer that the tweeter supports the target*
   - *the tweet is against something/someone other than the target, from which we can infer that the tweeter supports the target*
   - *the tweet is NOT in support of or against anything, but it has some information, from which we can infer that the tweeter supports the target*
   - *we cannot infer the tweeters stance toward the target, but the tweet is echoing somebody elses favorable stance towards the target (this could be a news story, quote, retweet, etc)*

2. We can infer from the tweet that the tweeter is against the target

   This could be because of any of the following:
   - *the tweet is explicitly against the target*
   - *the tweet is against someone/something aligned with the target entity, from which we can infer that the tweeter is against the target*
   - *the tweet is in support of someone/something other than the target, from which we can infer that the tweeter is against the target*
   - *the tweet is NOT in support of or against anything, but it has some information, from which we can infer that the tweeter is against the target*
   - *we cannot infer the tweeters stance toward the target, but the tweet is echoing somebody elses negative stance towards the target entity (this could be a news story, quote, retweet, etc)*



3. We can infer from the tweet that the tweeter has a neutral stance towards the target

   The tweet must provide some information that suggests that the tweeter is neutral towards the target – the tweet being neither favorable nor against the target is not sufficient reason for choosing this option. One reason for choosing this option is that the tweeter supports the target entity to some extent, but is also against it to some extent.

4. There is no clue in the tweet to reveal the stance of the tweeter towards the target (support/against/neutral)

- uploaded ~5000 instances on CrowdFlower
  - remaining instances form the Domain Corpus
- each instance on CrowdFlower was annotated by at least eight respondents
- quality control
  - 5% of the data annotated internally
- similarly, annotated the data for:
  - target of opinion: same as target of interest, or other
  - sentiment: positive, negative, or neutral language

Stance Data: Test and Training

- Less than 1% of instances that were marked as neutral stance
  - merged ‘neutral’ and ‘no clue’ into ‘neither’ (neither favor nor against)
- Selected instances with agreement equal to or greater than 60%
  - about 20% of the instances discarded
- Ordered tweets by timestamp
  - the first 70% formed the training set
  - the last 30% formed the test set

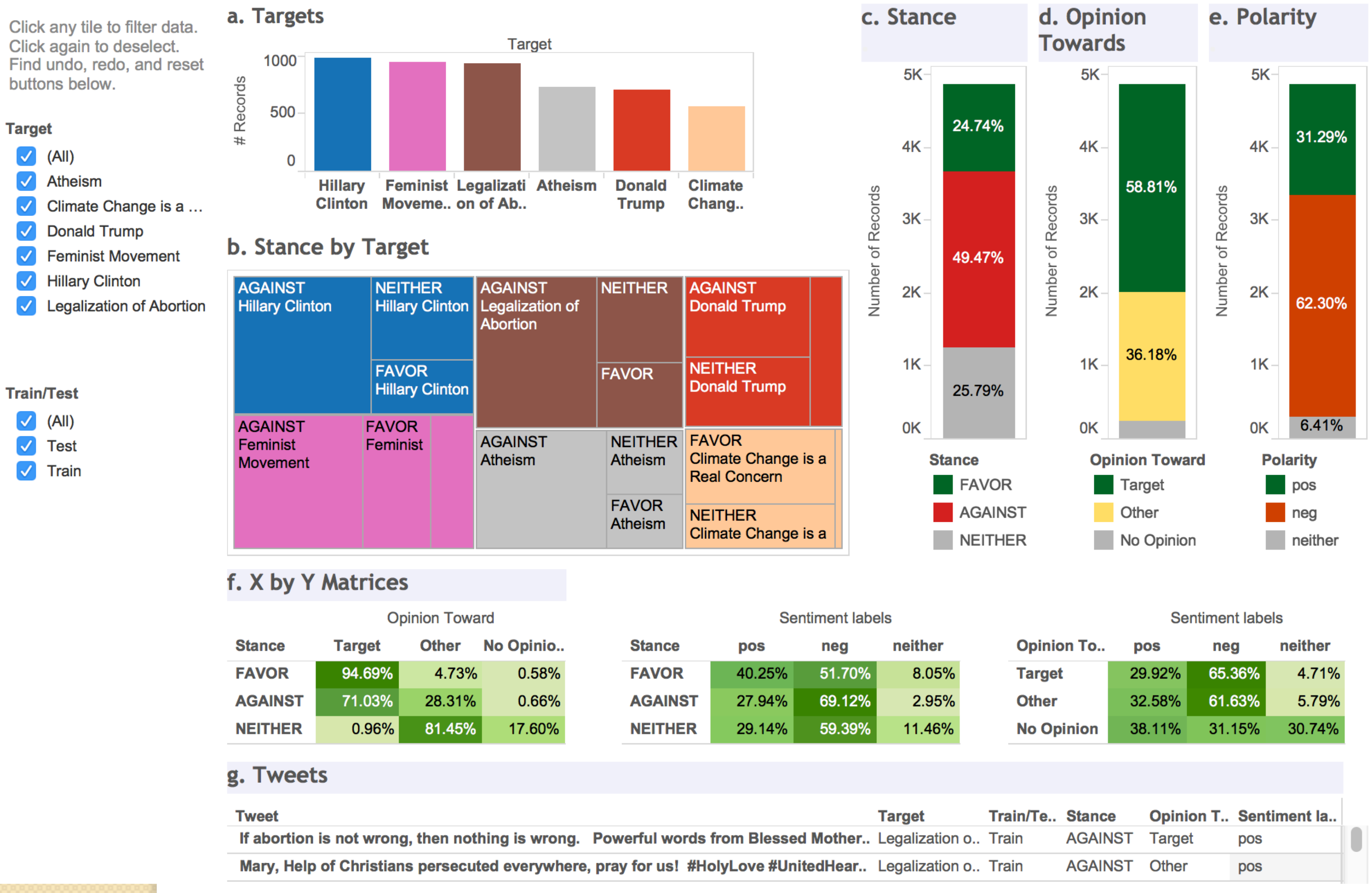
Stance Data: Analysis (continued)

- Often the target is not directly mentioned, and yet stance towards the target was determined by the annotators
  - about 30% of the ‘Hillary Clinton’ instances
    - did not mention ‘Hillary’ or ‘Clinton’
    - and yet stance is inferable
  - about 65% of the ‘Legalization of Abortion’ instances
    - did not mention ‘abortion’, ‘pro-life’, and ‘pro-choice’
    - and yet stance is inferable

Visualizing the Stance Dataset

An Interactive Visualization of the SemEval-2016 Stance Dataset:

A dataset of tweets manually annotated for stance towards given target, target of opinion (opinion towards), and sentiment (polarity).



Click any tile to filter data. Click again to deselect. Find undo, redo, and reset buttons below.

Target

- ☑ (All)
- ☑ Atheism
- ☑ Climate Change is a …
- ☑ Donald Trump
- ☑ Feminist Movement
- ☑ Hillary Clinton
- ☑ Legalization of Abortion

Train/Test

- ☑ (All)
- ☑ Test
- ☑ Train

f. X by Y Matrices

	Opinion Toward		
Stance	Target	Other	No Opinio..
FAVOR	94.69%	4.73%	0.58%
AGAINST	71.03%	28.31%	0.66%
NEITHER	0.96%	81.45%	17.60%

	Sentiment labels		
Stance	pos	neg	neither
FAVOR	40.25%	51.70%	8.05%
AGAINST	27.94%	69.12%	2.95%
NEITHER	29.14%	59.39%	11.46%

	Sentiment labels		
Opinion To..	pos	neg	neither
Target	29.92%	65.36%	4.71%
Other	32.58%	61.63%	5.79%
No Opinion	38.11%	31.15%	30.74%

g. Tweets

Tweet	Target	Train/Te..	Stance	Opinion T..	Sentiment la..
If abortion is not wrong, then nothing is wrong. Powerful words from Blessed Mother..	Legalization o..	Train	AGAINST	Target	pos
Mary, Help of Christians persecuted everywhere, pray for us! #HolyLove #UnitedHear..	Legalization o..	Train	AGAINST	Other	pos

Parinaz Sobhani, Saif M. Mohammad, and Svetlana Kiritchenko.

A Common Text Classification Framework for Stance and Sentiment

How useful are the sentiment classification features for stance detection?

- which features are less useful?
- which features are more useful?

Classification System

Preprocessing

- tweets tokenized and part-of-speech tagged - CMU Twitter NLP tool (Gimpel et al., 2011)

Machine Learning

- linear-kernel Support Vector Machine (SVM) classifier
  - trained on the Stance training set

Features

- n-grams:
  - contiguous sequences of 1, 2, and 3 tokens
  - contiguous sequences of 2, 3, 4, and 5 characters

- word embeddings: the average of the word vectors for words appearing in a given tweet.
  - 100-dimensional vectors using Word2Vec Skip-gram model trained over the Domain Corpus

- sentiment features: features drawn from sentiment lexicons as suggested in (Mohammad et al., 2013; Kiritchenko et al., 2014b)
  - NRC Emotion Lexicon (Mohammad and Turney, 2010)
  - Hu and Liu Lexicon (Hu and Liu, 2004)
  - MPQA Subjectivity Lexicon (Wilson et al., 2005)
  - NRC Hashtag Sentiment and Emoticon Lexicons (Kiritchenko et al., 2014b)

Baselines

- random:
  - a classifier that randomly assigns stance to each instance
- majority:
  - a classifier that simply labels every instance with the majority class per target
- oracle sentiment:
  - for each target,
    - select a sentiment-to-stance assignment that maximizes the F-score

    i.e., maps all positive instances to 'favor' and all negatives to 'against'

    or

    maps all positive instances to 'against' and all negatives to 'favor'

Evaluation Metric

- Macro-average of the F1-score for ‘favor’ and the F1-score for ‘against’

$$F_{avg} = \frac{F_{favor} + F_{against}}{2}$$

  - F1-score for ‘favor’ and the F1-score for ‘against’ are each taken across all target (micro across targets)

Stance Classification Results

	F_{avg}
Benchmarks:	
i. random	34.6
ii. majority	65.2
iii. oracle sentiment	57.2
Our Classifiers:	
i. n-grams	69.0
ii. n-grams, embeddings	70.3
iii. n-grams, sentiment lexicons	66.8
iv. n-grams, embeddings, sent. lexicons	69.8

Stance Classification Results: On subsets where opinion is expressed towards the target and where it is not

	towards target	towards other
Benchmarks:		
i. random	37.4	21.6
ii. majority	71.2	41.3
iii. oracle sentiment	65.3	33.3
Our Classifiers:		
i. n-grams, embeddings, sent. lexicons	75.3	44.2

Example:

Target: Donald Trump

Tweet: Hillary Clinton is the only sane candidate in this election #rightchoice

Sentiment Classification Results

	F_{avg}
Benchmarks:	
i. random	35.7
ii. majority	38.8
Our Classifiers:	
i. n-grams	73.3
ii. n-grams, embeddings	76.4
iii. n-grams, sentiment lexicons	78.9
iv. n-grams, embeddings, sent. lexicons	78.6

Sentiment Classification Results: On subsets where opinion is expressed towards the target and where it is not

	towards target	towards other
Benchmarks:		
i. random	29.2	34.6
ii. majority	40.0	36.9
Our Classifiers:		
i. n-grams, embeddings, sent. lexicons	79.6	77.8

Example:

~~Target: Donald Trump~~

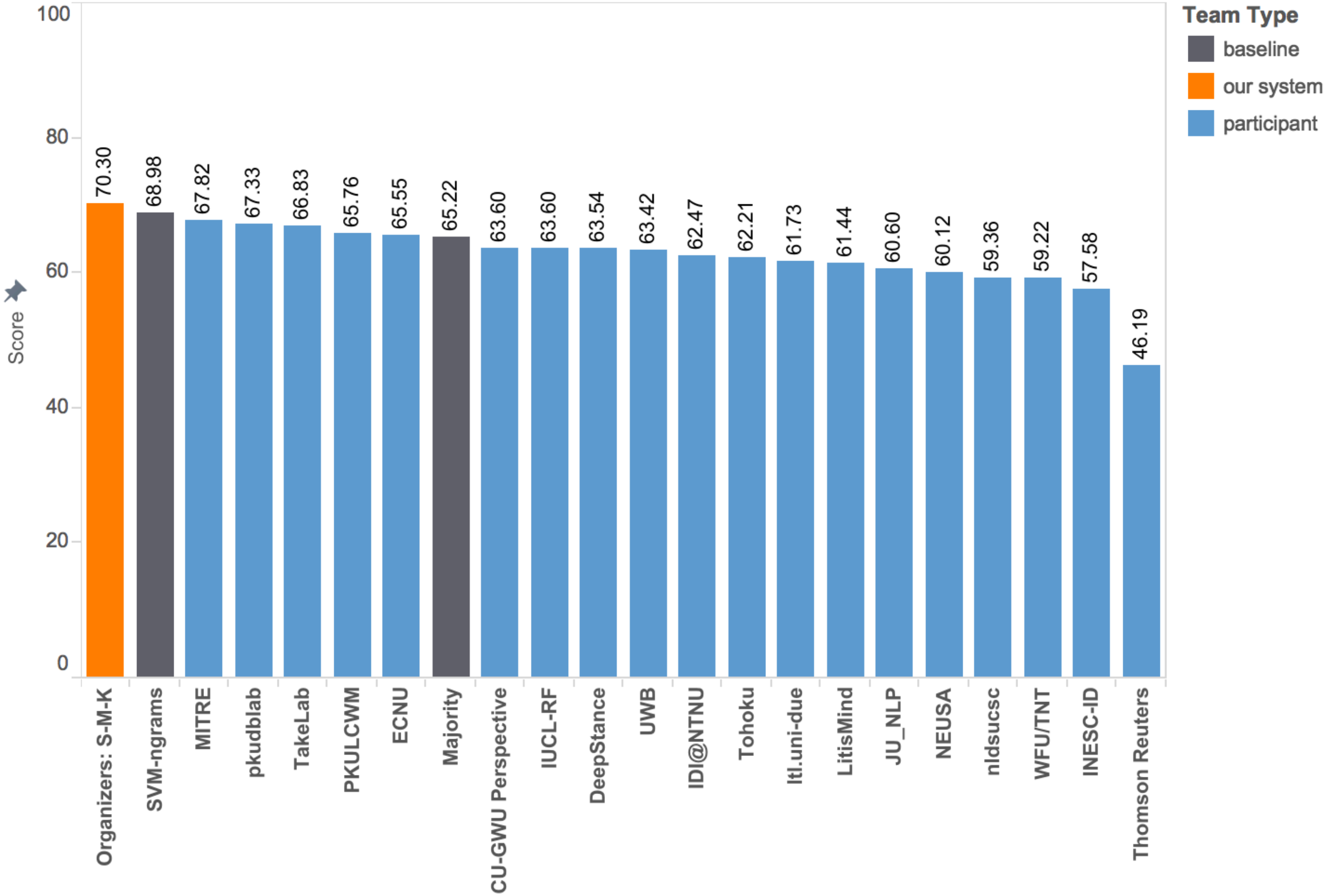
Tweet: Hillary Clinton is the only sane candidate in this election #rightchoice

SemEval-2016 Task#6: Detecting Stance in Tweets

- Task A: Supervised Framework
  - training data: 2,914 labeled instances for five targets
  - test data: 1,249 instances for the same five targets

- Task B: Weakly Supervised Framework
  - training data: none
  - test data: 707 tweets for one target ‘Donald Trump’
  - unlabeled data: 78,000 tweets associated with ‘Donald Trump’ to various degrees – the *domain corpus*
    - tweets that include hashtags associated with Donald Trump

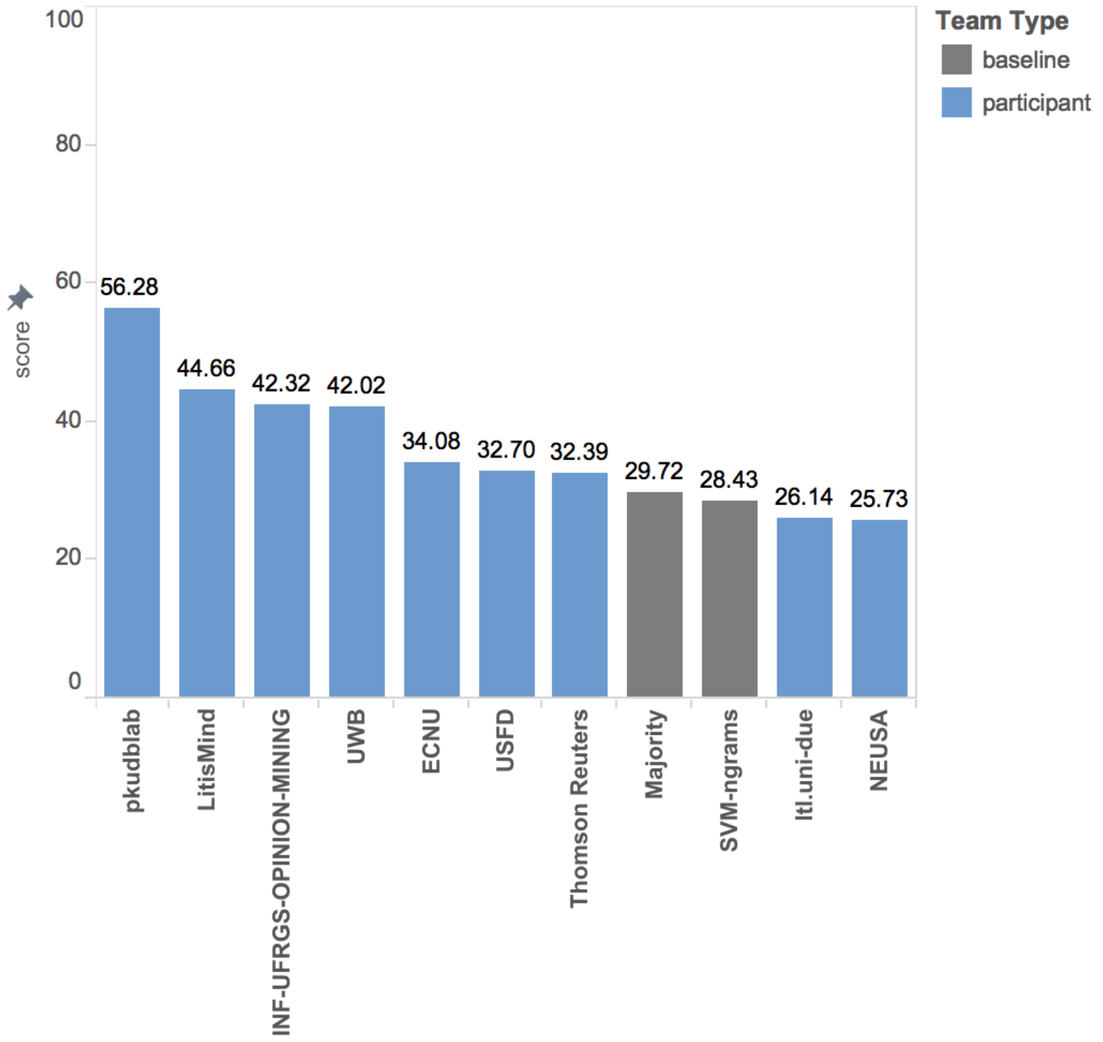
Results: Task A (19 teams participated)



Automatic Systems to Detect Stance

- Nineteen teams competed in Task A (supervised stance detection)
- Best results by a participating system (MITRE): F-score of 67.82
  - two recurrent neural network (RNN) classifiers
  - used a large unlabeled Twitter corpus
- Our baseline (SVM-ngrams): F-score of 68.98
  - word n-grams (1-, 2-, and 3-gram) features
  - character n-grams (2-, 3-, 4-, and 5-gram) features
- S-M-K (SVM-ngrams-embeddings): F-score of 70.30

Results: Task B (9 teams participated)



Summary

- Created a dataset for detecting stance towards pre-chosen targets from tweets
- Annotated the same dataset for target of opinion and sentiment
- Created an interactive visualization to explore the data
- Conducted classification experiments for stance and sentiment
    - stance results better than 19 participating teams at SemEval-2016
    - showed that sentiment features much less useful for determining stance than for sentiment
    - performance is much lower when the target of opinion is an entity other than the target of interest
- Unsupervised form of stance detection attractive as it does not require new labeled data

Stance Project Homepage

http://www.saifmohammad.com/WebPages/StanceDataset.htm

- Complete Stance Dataset with annotation for both stance and sentiment
- Interactive visualization for the Stance Dataset

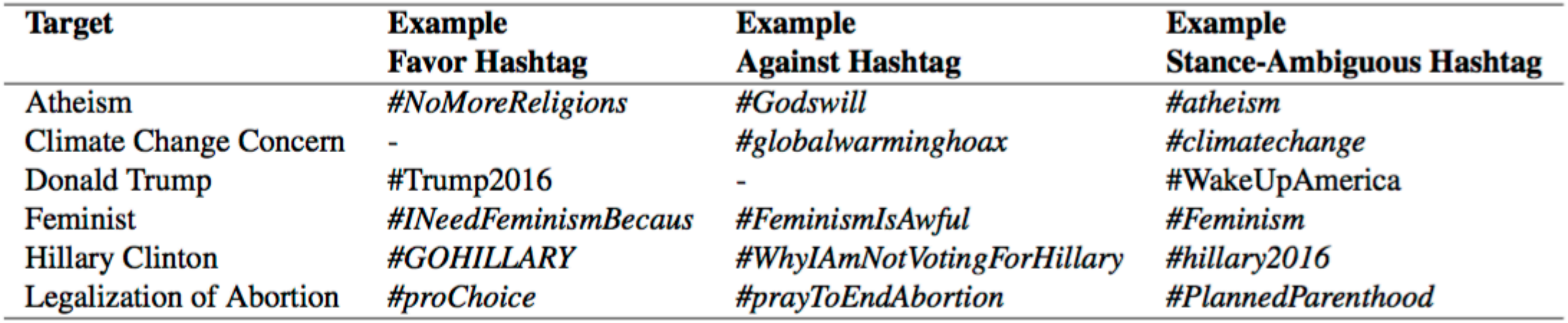
SemEval-2016 Task #6: Detecting Stance from Tweets

http://alt.qcri.org/semeval2016/task6/index.php?id=data-and-tools

- Training and test sets for Task A (only stance annotations)
- Test set and domain corpus for Task B (only stance annotations)
- Evaluation script and format checker
- Questionnaire to the annotators

Selecting Tweet-Target Pairs (continued)

Examples of the query hashtags (stance-indicative and stance-ambiguous)



Target	Example Favor Hashtag	Example Against Hashtag	Example Stance-Ambiguous Hashtag
Atheism	*#NoMoreReligions*	*#Godswill*	*#atheism*
Climate Change Concern	-	*#globalwarminghoax*	*#climatechange*
Donald Trump	#Trump2016	-	#WakeUpAmerica
Feminist	*#INeedFeminismBecaus*	*#FeminismIsAwful*	*#Feminism*
Hillary Clinton	*#GOHILLARY*	*#WhyIAmNotVotingForHillary*	*#hillary2016*
Legalization of Abortion	*#proChoice*	*#prayToEndAbortion*	*#PlannedParenthood*

Task A teams

- Used many standard text classification features
  - n-grams, word embedding vectors, sentiment lexicons, pos, hashtags

- Polled Twitter for additional unlabeled data and noisy labeled data (using hashtags)

- Used many standard machine learning algorithms
  - SVMs, recurrent neural networks

Properties of a Good Stance-Labeled Dataset

1. The tweets and targets are commonly understood
   - to avoid need for obscure world knowledge
   - to help annotators judge stance

2. It has significant amount data for each of the three classes: favor, against, none
   - avoid processes that lead to highly skewed distributions

3. It has significant amount of data where:
   - the target of interest is referred to by many different names
   - or, opinion is expressed without referring to target by name

Example mentions: Hillary Clinton, Hillary, Clinton, HillNo, Hillary2016
Example tweet: Benghazi questions need to be answered #Jeb2016

Properties of a Good Stance-Labeled Dataset (continued)

4. It has significant amount of data where the target of opinion is an entity other than the given target of interest
   - challenging for automatic systems
   - downstream applications often require stance towards particular pre-chosen targets

Example:

Target: Donald Trump
Tweet: Jeb Bush is the only sane candidate in this republican lineup.

The target of opinion in the tweet is Jeb Bush.
Nonetheless, we can infer that the tweeter is likely unfavorable towards Donald Trump.

Task B teams

- pkudblab
  - annotated the domain corpus with rules
  - trained a deep convolutional neural network
  - combined its output with rules to predict stance
- Polled Twitter for additional unlabeled data and noisy labeled data
  - using hashtags (ListisMind)
  - using keyword rules (pkudblab)
  - combination of rules and sentiment classifiers (INF-URGS)
- Generalized from labeled data for Task A

Areas of Future Work

- Stance and Opinion / Implicit Stance and Implicit Opinion
  - performance is much lower when the target of opinion is an entity other than the target of interest
- Stance and Relationships Extraction
  - knowing that entity X is an adversary of entity Y can be useful in detecting stance towards Y in tweets that mention X
- Stance and Textual inference (Textual Entailment)
  - to determine whether the favorability of the target is entailed by the tweet