favor

against

neither

SemEval-2016 Task #6: Detecting Stance in Tweets

Saif M. Mohammad[1], Svetlana Kiritchenko[1], Parinaz Sobhani[2],
Xiaodan Zhu[1], Colin Cherry[1]
[1]National Research Council Canada, [2]University of Ottawa

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: 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.

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.

Selecting Tweet-Target Pairs

Selecting Tweet-Target Pairs

- selected as targets a small subset of entities that were routinely discussed on Twitter at the time of data collection, and were controversial (Property 1):
  - 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 (Property 2)
- 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 (Properties 3 and 4)

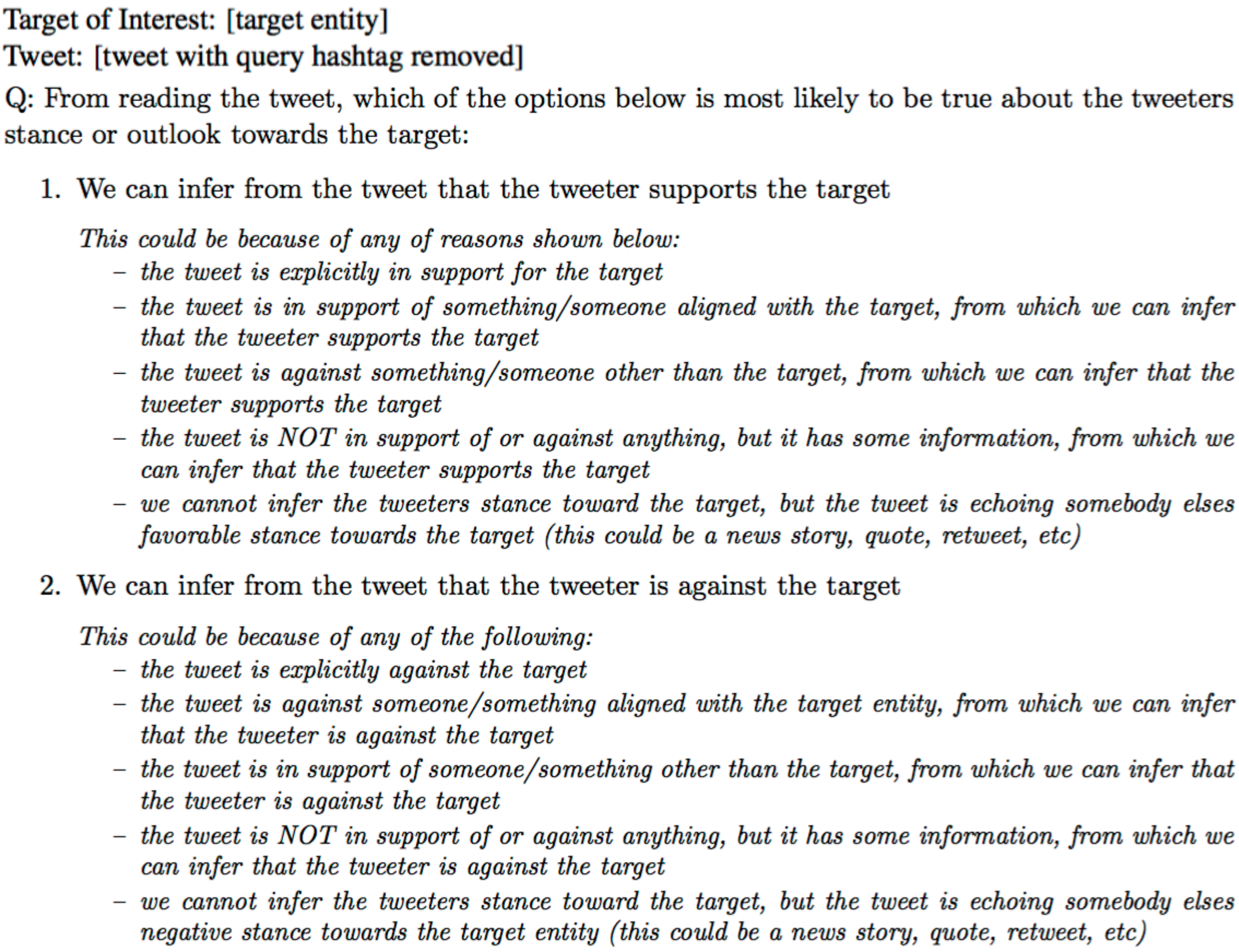
    Target: Hillary Clinton

    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]

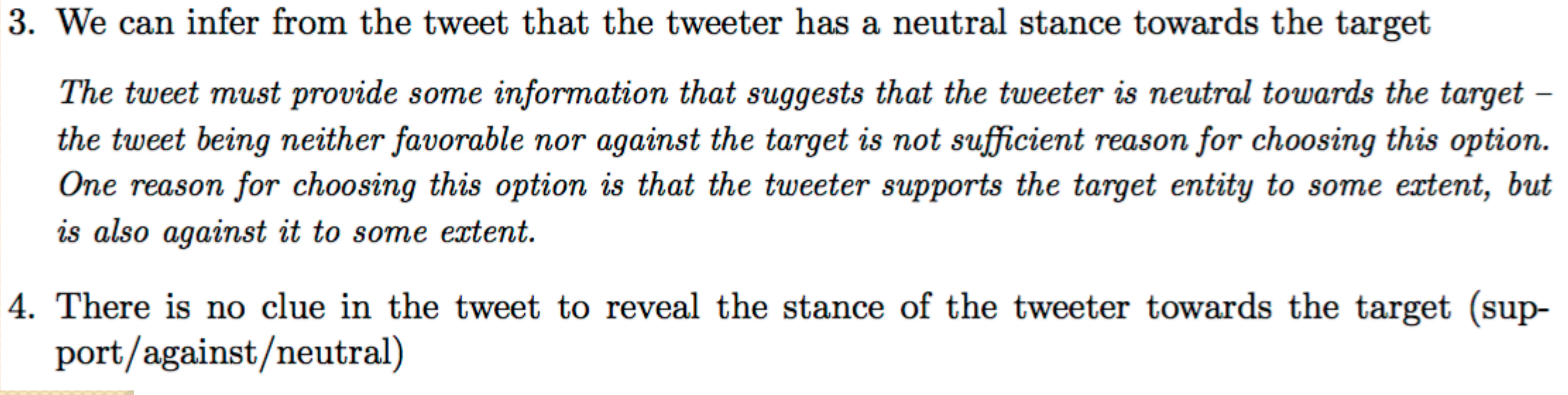
Q: From reading the tweet, which of the options below is most likely to be true about the tweeters stance or outlook towards the target:

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 on CrowdFlower
- each instance was annotated by at least eight respondents
- quality control
  - 5% of the data annotated internally

In subsequent work, we also annotated the data for target of opinion and sentiment.

Stance and Sentiment in Tweets. Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. 2016b. Special Section of the ACM Transactions on Internet Technology on Argumentation in Social Media, Submitted.

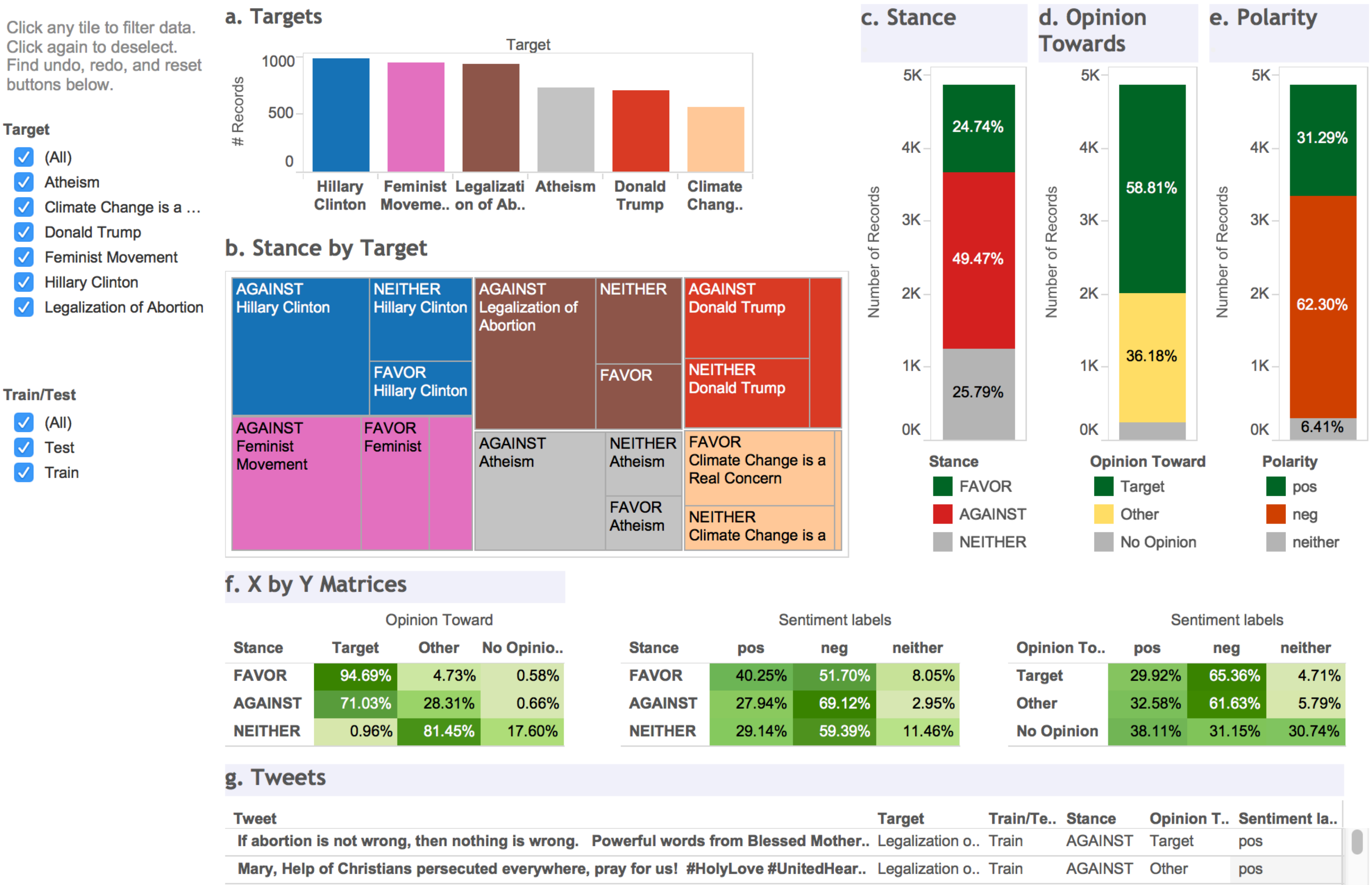
Stance Data: Test and Training

- 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

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

a. Targets

b. Stance by Target

c. Stance

d. Opinion Towards

e. Polarity

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

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

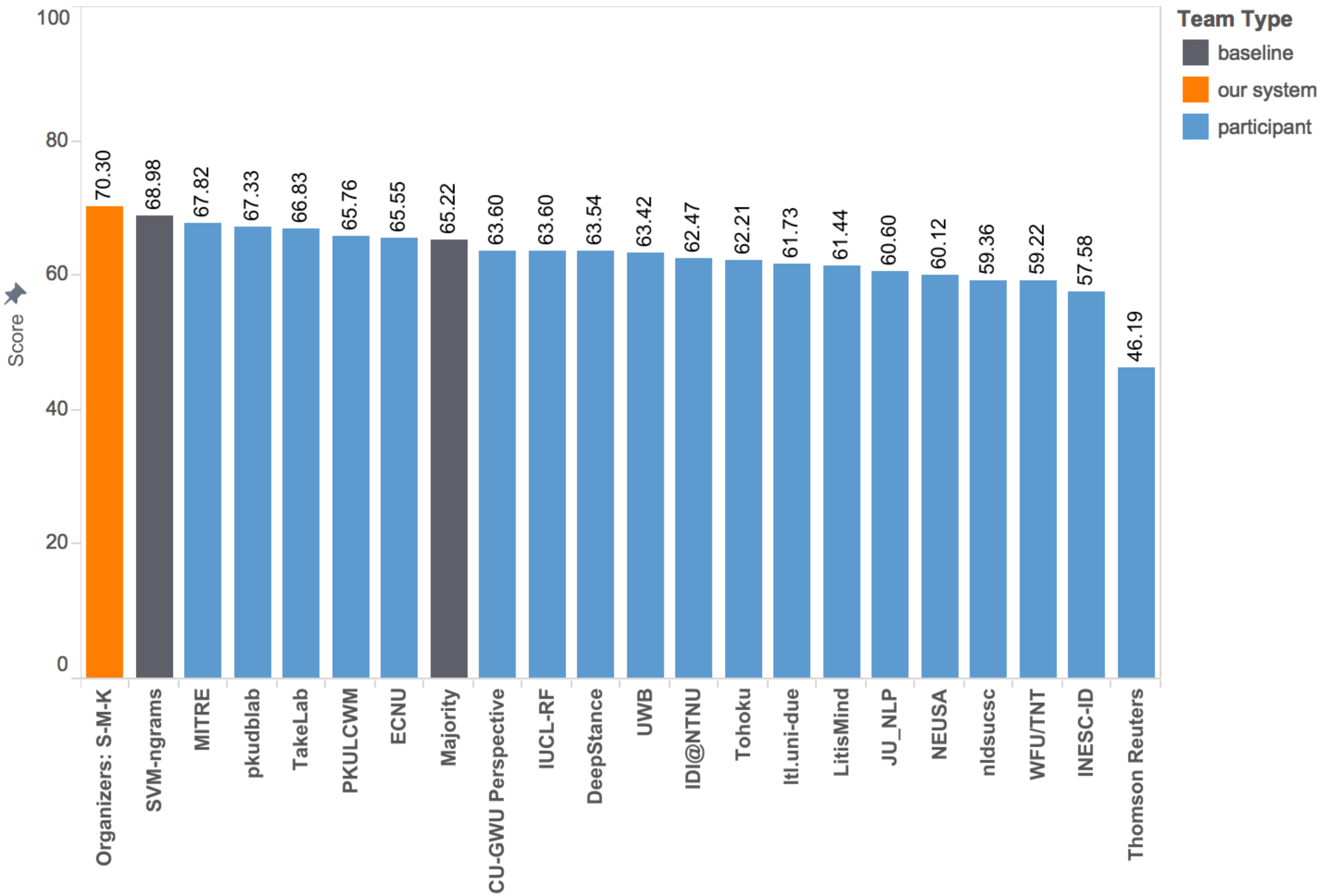
Evaluation Metric

- 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)

Results: Task A (19 teams participated)



Scores on subsets of test set:

where opinion is expressed towards the target, some other entity, or on the whole set

Team	Opinion Towards Target	Opinion Towards Other	All Test Data
SVM-ngrams	**74.54**	43.20	**68.98**
MITRE	72.49	44.48	67.82
pkudblab	71.07	**46.66**	67.33
TakeLab	73.66	37.47	66.83

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

Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. 2016b. Stance and sentiment in tweets. Special Section of the ACM Transactions on Internet Technology on Argumentation in Social Media, Submitted.

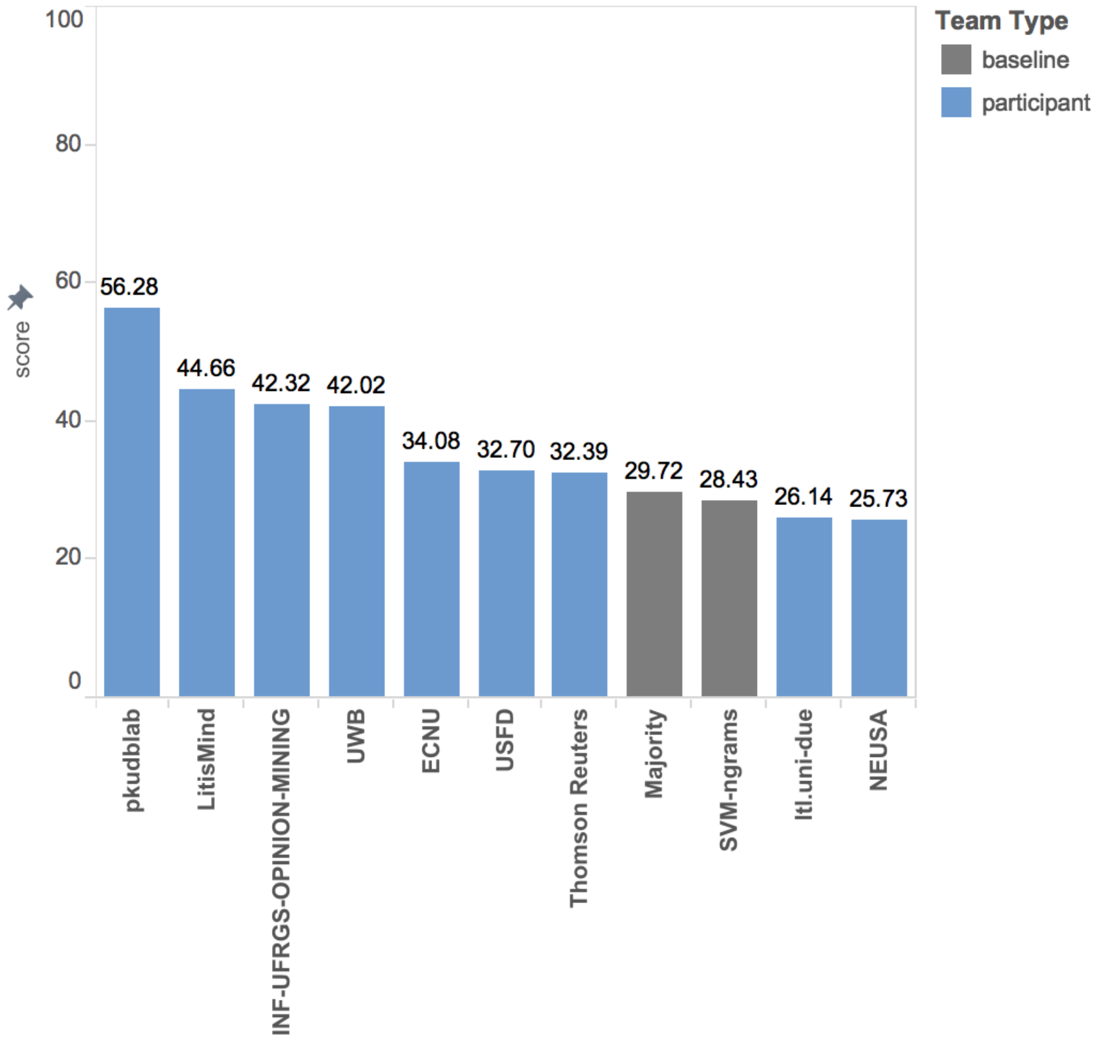
Task A teams

- 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

Results: Task B (9 teams participated)



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

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

favor

against

neither