Joseph Jinn

Notes on “SemEval-2016 Task 6: Detecting Stance in Tweets”

* F-score?
* Stance detection:
  + Task of automatically determining from text whether the author of the text is in favor of, against, or neutral towards a proposition or target.
  + Difference from stance analysis – determine favorability towards given (pre-chosen) target of interest.
* Automatic stance detection – applications:
  + Information retrieval
  + Text summarization
  + Textual entailment?
* Task:
  + Given tweet text and target entity (person, organization, movement, policy, etc.)
  + Automatic natural language systems 🡪 determine tweeter in favor of or against given target OR neither inference is likely
* Task A – access to training data
* Task B – no training data
* Standard text classification features
  + N-grams
  + Word vectors
  + Sentiment lexicons
* Neutral stance:
  + No explicit neutral class
  + Merge all classes other than “favor” and “against” into one “neither” class
* Sentiment analysis:
  + Determine whether a piece of text is positive, negative, neutral.
  + OR determine from text speaker’s opinion and the target of the opinion.
* Tweet-target pairs:
  + Tweet and target are commonly understood by a wide # of people in the U.S.
  + Must be large amount of data for 3 classes of: favor, against, neither.
  + Dataset should include large # of tweets expressing opinion towards target without direct referral by name.
  + Dataset should include tweets that explicitly mention the target.
  + Dataset should include large # of tweets where target of opinion is different from given target of interest.
* Query hashtags”
  + Favor hashtags – occur in tweets expressing favorable stance towards target
  + Against hashtags – occur in tweets expression opposition towards target
  + Stance-ambiguous hashtags – occur in tweets about the target but not explicitly indicative of stance.
* Pre-processing:
  + Poll Twitter API for tweets containing query hashtags
  + Discard retweets and tweets with URL’s
  + Keep only tweets with query hashtags at the end
  + Removed query hashtags from tweets to exclude obvious cues for classification
  + Manually annotate tweet-target pairs
* Targets:
  + Sample equal # of tweets pertaining to each of the 3 classes: favor, against, stance-ambiguous
  + Max of 1000 tweets
* Stance annotation:
  + We can infer from the tweet that the tweeter supports the target
  + We can infer from the tweet that the tweeter is against the target
  + We can infer from the tweet that the tweeter has a neutral stance towards the target
  + There is no clue in the tweet to reveal the stance of the tweeter towards the target
* Stance Dataset:
  + 5 of 8 annotators per tweet must agree
  + Partition into training/test set based on timestamps of tweets
  + First 70% are training set and last 30% are test set.
* Task A (supervised framework):
  + 5 targets:
    - Atheism
    - Climate Change is a Real Concern
    - Feminist Movement
    - Hillary Clinton
    - Legalization of Abortion
  + 2914 labeled training data instances.
  + 1249 test data instances.
* Task B (weakly supervised framework):
  + Tested stance towards one target: Donald Trump
  + No training data.
  + 78,000 tweets given associated with target. (domain corpus)
    - Not labeled for stance
* Common Evaluation Metric:
  + Macro-average of F1-score for “favor” and F1-score for “against” as bottom-line evaluation metric
  + F(average) = F(favor) + F(against) / 2
  + F(favor) = 2\*P(favor)\*R(favor) / P(Favor) + R(favor)
  + F(against) = 2\*P(against)\*R(against) / P(against) + R(against)
  + “neither” as class not of interest (negative class in IR – information retrieval terms)
  + F-microT – micro-average of F-scores across targets.
  + F-macroT – macro average across targets.
* Task A Baselines:
  + Majority class – classifier that labels every instance with the majority class (‘favor’ or against’) for the corresponding target.
  + SVM-unigrams: 5 SVM classifiers (one per target) trained on corresponding training set for the target using word unigram features.
  + SVM-ngrams – 5 SVM classifiers (one per target) trained on corresponding training set for target using word n-grams (1-, 2-, 3-gram) and character n-grams (2-, 3-, 4-, 5-gram) features.
  + SVM-ngrams-comb: 1 SVM classifier trained on combined (all 5 targets) training set using word n-grams (1-, 2-, 3-gram) and character n-grams (2-, 3-, 4-, 5-gram) features.
  + 5-fold cross-validation tuning on training data?
* RNN – recurrent neural network classifiers?
* Cross-target balancing?
* Task A results on 2 subset of the test set:
  + Subset where opinion is expressed towards the target
  + Subset where opinion is expressed towards some other entity
* Sentiment and emotion lexicons?
* Standard text classification features and standard sentiment analysis features
* Task B baselines:
  + Majority class – classifier that labels every instance with majority class (favor or against) for corresponding target.
  + SVM-ngrams-comb – 1 SVM classifier trained on combined (all 5 targets) Task A training set, using n-grams (1-, 2-, 3-gram) and character n-grams (2-, 3-, 4-, 5-gram) features.
* Noisy labels: