

Determining Support from Transcripts of Congressional Floor Debate

Cheuk Hang Lee

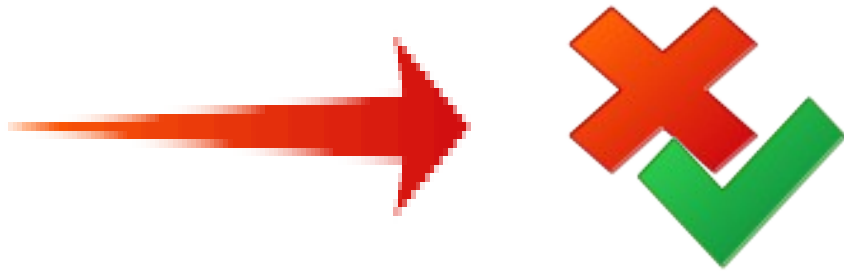
Input: transcripts of Congress floor debates

(sample speech 2346007)

Many of the accountability principles that were outlined in the gentleman's bill, H.R. 778, have been included in the School Readiness Act, and these reforms I think are critical if ...



Output: whether the speech supports the legislation



speech 2346007: support
speech 2346009: oppose

...

- sentiment analysis
- binary classification problem

Procedure

[from Thomas, Pang, Lee (EMNLP 2006)]

1. Use SVM to score each speech segment individually
- 2a. Find references to other speakers in speeches
- 2b. Train another SVM using (dis)agreements of speech pairs as label, and words near the reference as feature vector
3. Use max-flow/min-cut algorithms to find assignment that minimizes

$$\sum_s ind(s, \bar{c}(s)) + \sum_{s, s' : c(s) \neq c(s')} \sum_{l \in L(s, s')} str(l)$$

where ind , str are cost functions and l is an agreement link between 2 speeches (s is over all speeches)

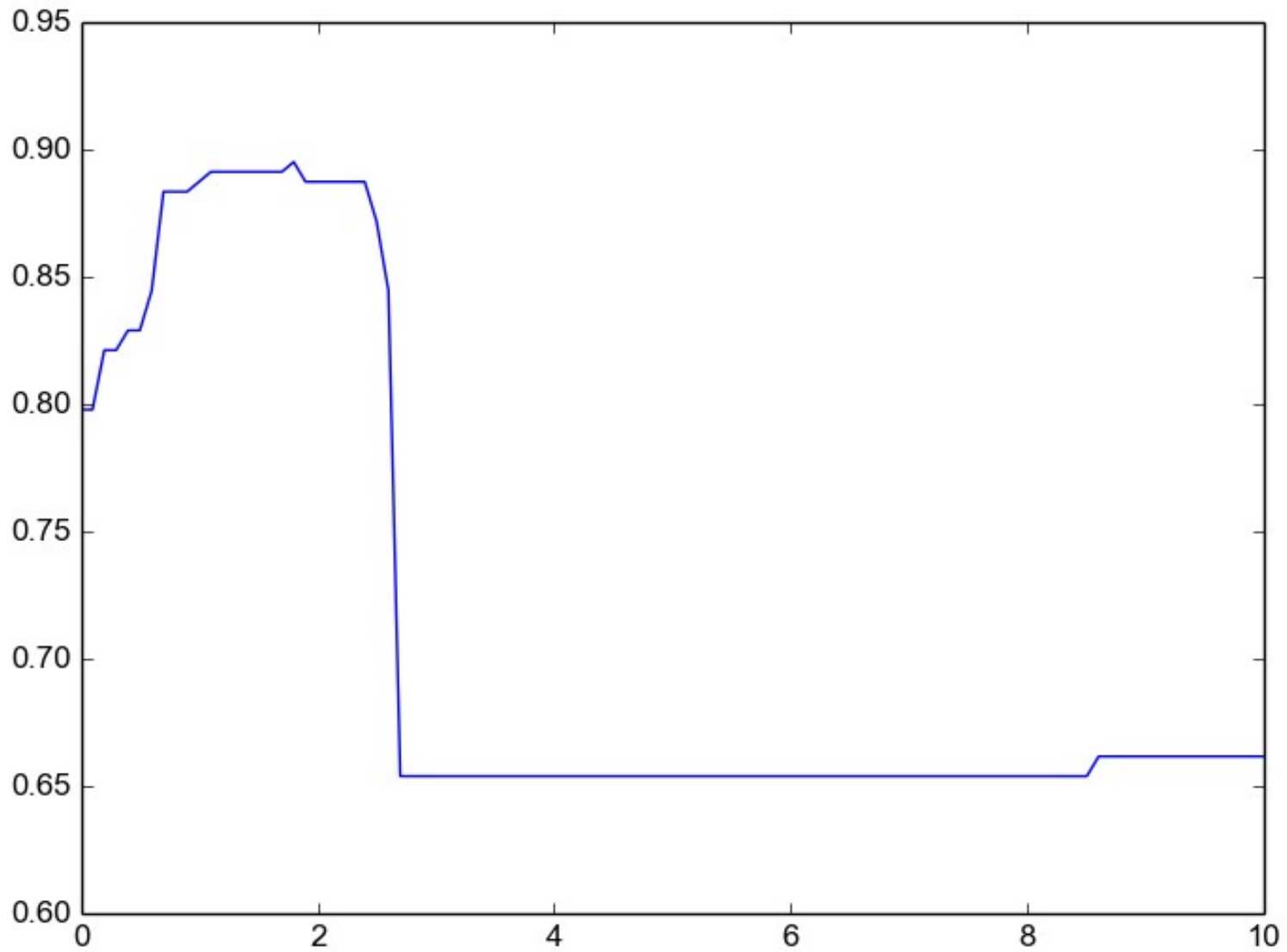
4. We can weight the agreement links (i.e. k times $\text{str}(l)$ instead of $\text{str}(l)$); find the optimal weight
5. re-insert yield statements before running the algorithm on test set

Possible Variations

- assign higher importance to speeches that mentions the legislation directly
- only include links that we are confident of

Key Ideas

- exploiting the connection between speeches gives better result than analyzing each speech individually
- same-speaker constraint: assume that speeches by a Congressman either all support or all oppose the legislation
- hard vs. soft constraint: two speeches with agreement link between each other does not have to receive same label; forcing the same label on the speeches significantly decreases accuracy, even though using appropriate weight will increase accuracy



(Accuracy of development set as a function on agreement links in the total cost function)

Majority baseline
(Yes if most speeches say yes)

SVM on speech segments

SVM + same-speaker constraint

SVM + same-speaker + agreement links

SVM + same-speaker + agreement links
+ mention links

Development set	Test set
54.1%	58.4%
70.0%	66.1%
79.8%	67.2%
89.5%	69.5%
89.5%	72.0%