

Flow of Semantics in Narratives

Daniel Khashabi, Chris J.C. Burges

Microsoft Research

Introduction

- MCTest-160 [1] contains pairs of narratives and multiple choice questions.
- An example of narrative QA:
Story: "Sally had a very exciting summer vacation. She went to summer camp for the first time. She made friends with a girl named Tina. They shared a bunk bed in their cabin. Sally's favorite activity was walking in the woods because she enjoyed nature. Tina liked arts and crafts. Together, they made some art using leaves they found in the woods. Even after she fell in the water, Sally still enjoyed canoeing. ..."
Q: Who went to the beach with Sally?
1) her sisters
2) Tina
3) her brothers, mother, and father
4) herself

An early observation: pruning stories helps!

- Suppose we manually remove irrelevant sentences from the story. Does this help improve the performance of QA baseline?
- **Experiment:** Given set of (Story, Question, Options) answered wrongly by the baseline filtered the stories:

- **Scheme 1:** Keep the sentences needed to answer the question.
- **Scheme 2:** Keep the sentences needed to answer the question and relevant to options.

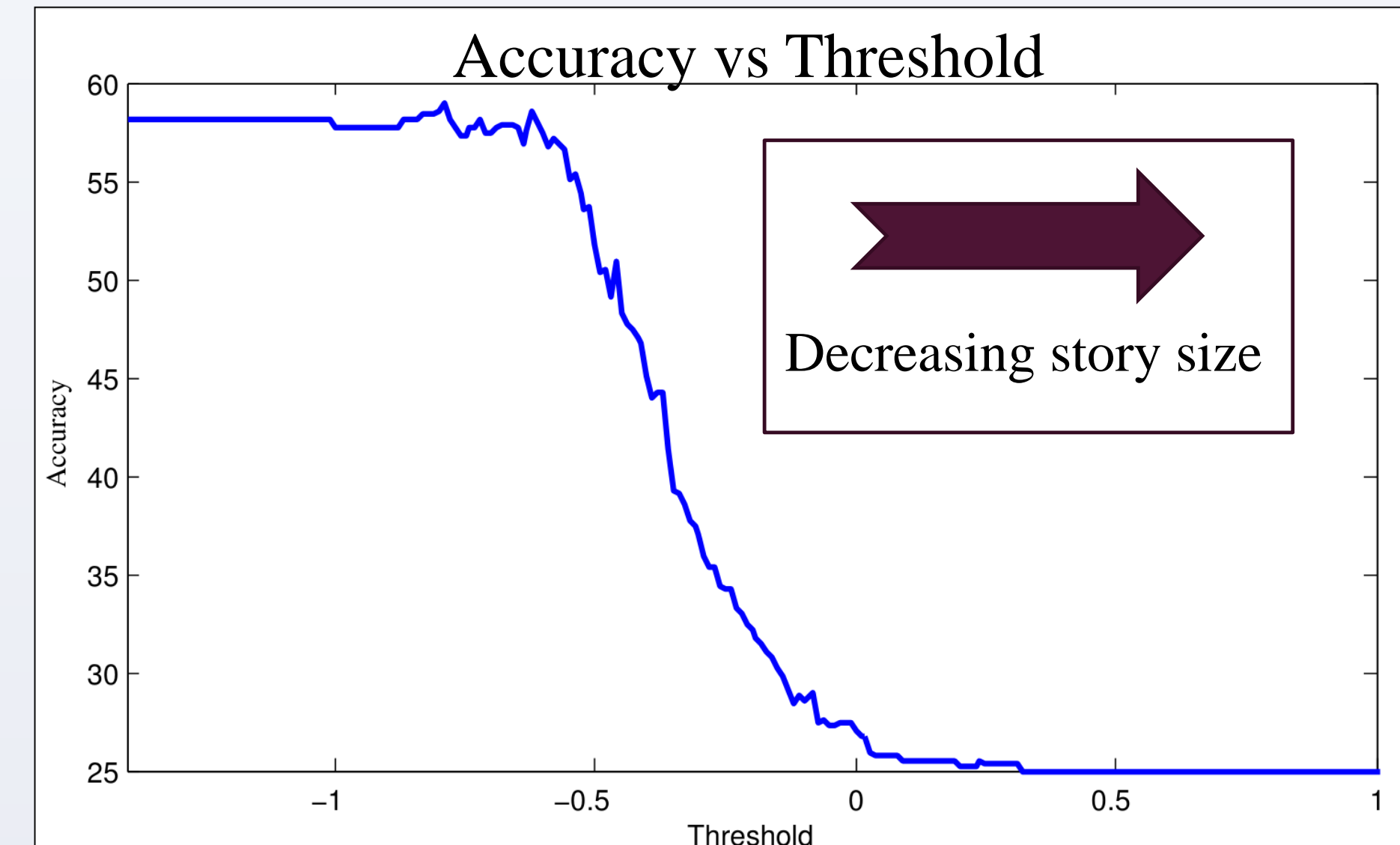
| Accuracies on baseline error | | |
|------------------------------|----------|----------|
| No pruning | Schema 1 | Schema 2 |
| 0.0 % | 38.77 % | 14.13 % |

- **Next:** With this observation, can we come up with an efficient way of pruning sentences for QA?

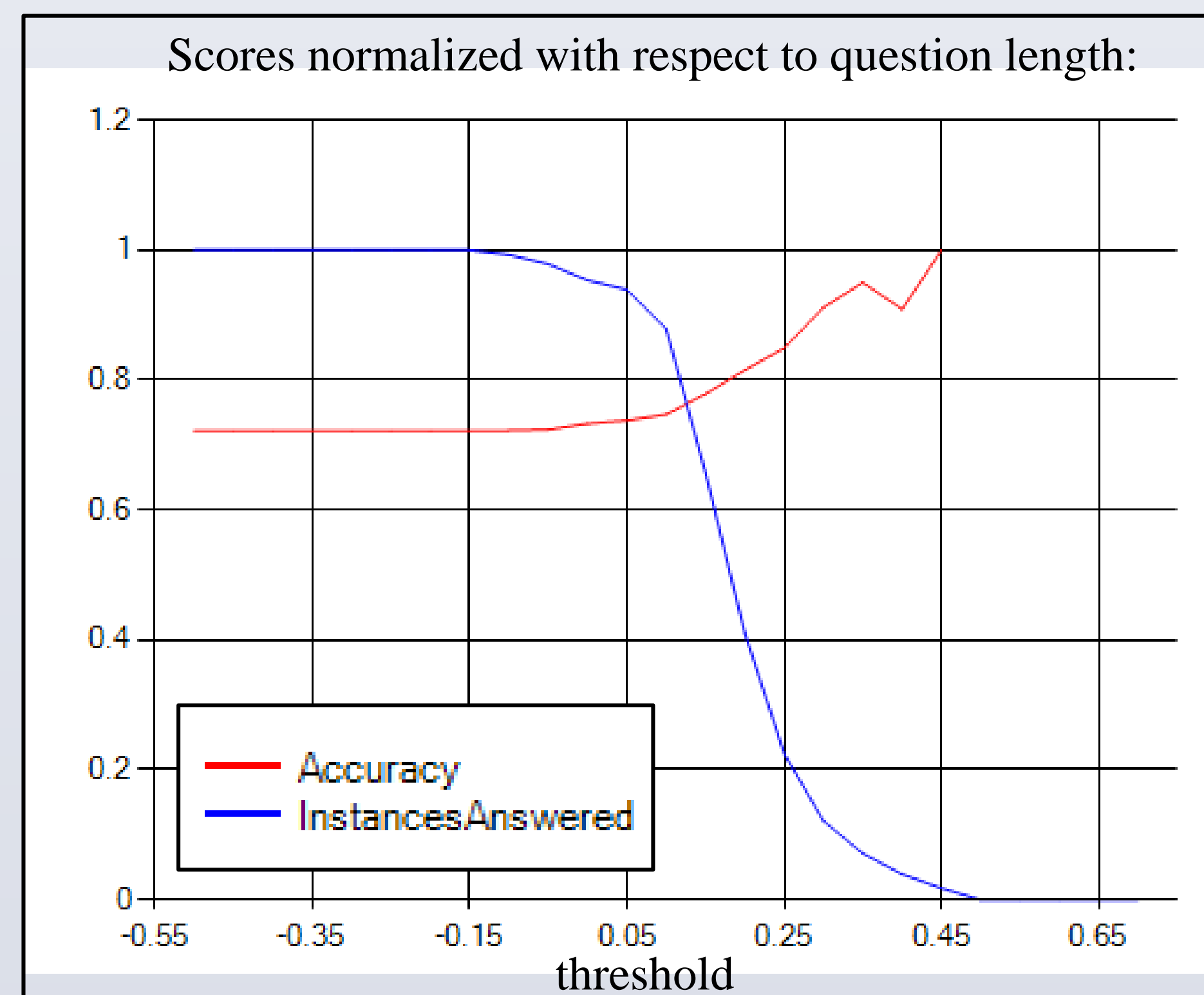
System 1: Finding the Right Sentence(s)

System 2: Answering the question, given the selected sentence(s)

- **Experiment:** We selected the system in [2] (one of the state-of-the-art QA systems).
- This system scores a pair of (Sentence, Question).
- Given any sentence $s_i \in \text{Story}$, and question q , run the scoring system on (s_i, q) and get the score.
- For the given threshold v_{th} , keep the sentences with $\text{Score}(s_i, q) \geq v_{th}$.
- **Result:** Just a little performance increase over the baseline.



- **Next:** Can we filter some of the hard questions?
- **Experiment:** Given a threshold v_{th} , abstain all of the questions for which the best option has confidence less than v_{th} :



- **Result:** Increased performance comes at the cost of losing substantial coverage!
- **Next:** Can we chunk events, based on their animate characters?

Coreference for creating timeline

- Suppose we are given coreference chains of a given story. Can we create a coherent set of event chains, for each animate entity?
- **Experiment:**
 - Extract coreference chains for each story
 - Separate chains into plural and singular chains, and find connection between them.
 - And create timeline of singular chains, with plural chains attached to them.
- An example is shown below. The output can be much noisier.
- **Next:** Can we use this timeline to improve QA?
- **Experiment:** Given question, find its main character, and choose the subset of sentences, related to this character, using the timeline.
- **Result:**
 - On average the reduced size of the stories is 83.2% (no. of words) of the original stories.
 - But the performance almost doesn't change.
- **Promising point:** although no improvement yet, but reduced text

Relgrams

- Define an event to be:
(Subject; Relation; Object)
- Ex.:
(Abramoff; was indicted in; Florida)
- Normalize the arguments:
(Person; be indicted in; State)
- In [4] the authors have gathered co-occurrence information for the normalized events, for 1.8 M words, gathering about 11M pair of events.

Creating timeline with event probabilities

- We can use the co-occurrence information between the normalized-event.
- We are given:
 $p(e_j|e_i), \forall e_i \in \mathcal{E}_1, \forall e_j \in \mathcal{E}_2$
such that:
$$\sum_{\forall e_j \in \mathcal{E}_2} p(e_j|e_i) = 1, \forall e_i \in \mathcal{E}_1$$
- Given a story, extract all of the events in the story: \mathcal{E}_{story}
- For any pair of events in the story extract event-transition probability: $p(e_2|e_1)$
- The indicator variable $x(e_1 \rightarrow e_2)$ shows the directed connection from e_1 to e_2 , if it is 1.
$$\begin{cases} \max_x \sum_{e_1, e_2 \in \mathcal{E}_{story}} x(e_1 \rightarrow e_2)(p(e_2|e_1) - \theta) \\ x(e_1 \rightarrow e_3) \geq x(e_1 \rightarrow e_2) + \\ \quad x(e_2 \rightarrow e_3) - 1, \forall e_1, e_2, e_3 \\ x(e_1 \rightarrow e_2) \in \{0,1\}, \quad \theta \in (0,1) \end{cases}$$
- The ILP is inspired from [3].
- θ is a handle to control number of clusters.
- **Problem:** the coverage of the Relgram dataset is not enough to give reasonable set of transition probabilities between events.
- **Alternative:** Training a classifier which approximates the transition probabilities $p(e_1 \rightarrow e_2)$ for any pair e_1 and e_2 .
 - In progress ...

Future work

- Using the timeline information for improving coreference
- QA on the subset chosen by event timeline
- Neural models for transition between events
- SRL-gram

Acknowledgment

Scott Yih, Erin Renshaw, Andrzej Pastusiak, Matt Richardson, and Chris Meek

References

- [1] Richardson, Matthew, et al "MCTest: A Challenge Dataset for the Open-Domain Machine Comprehension of Text." EMNLP. 2013.
- [2] Chang, Wen-tau Yih Ming-Wei, and Christopher Meek Andrzej Pastusiak. "Question answering using enhanced lexical semantic models." (2013).
- [3] Chang, Kai-Wei, et al. "Inference protocols for coreference resolution." CoNLL: Shared Task, 2011.
- [4] Balasubramanian, Niranjan, et al. "Generating Coherent Event Schemas at Scale." EMNLP. 2013.

