

Wenpeng Yin¹, Hinrich Schütze², Dan Roth¹

¹University of Pennsylvania, USA; ²LMU Munich, Germany

End-task oriented TE

TE: textual entailment;
 ⇒ Given a premise (**P**), if a hypothesis (**H**) is true or not

End-task oriented TE ... Why?

A TE problem that can contribute to a downstream NLP problem is more worthy of studying.

SciTail (Khot et al., 2018)

– A TE task derived from a multi-choice QA problem.

A performance gain on SciTail

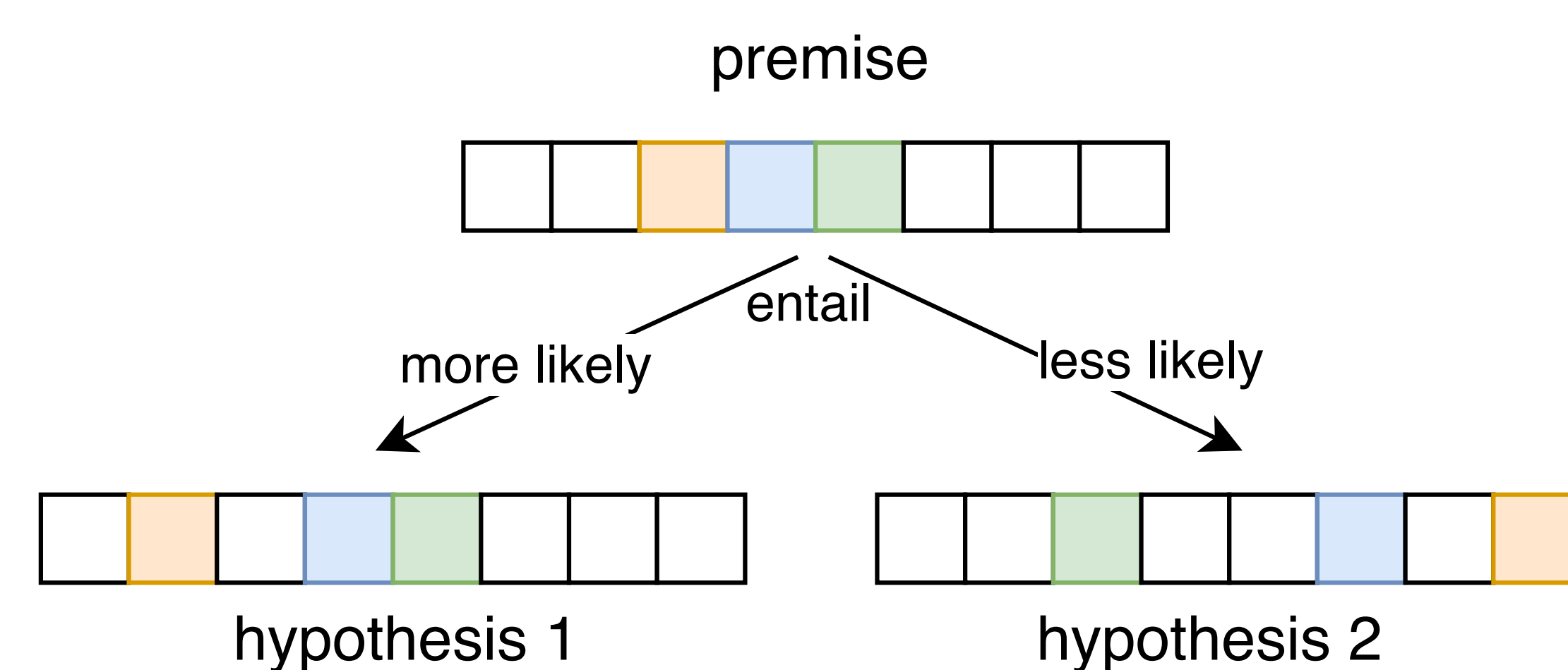
⇒ A better QA performance.

	#instances	#entail	#neutral
train	23,596	8,602	14,994
dev	1,304	657	647
test	2,126	842	1,284
total	27,026	10,101	16,925

Our model attempts to

Make full use of interactions between (**P**, **H**). E.g.,

- Detect important words and let the final sentence representation express mainly those keywords;
- Find the aligned words that are most indicative;
- For a window of words in *P* or *H*, the relative locations of their best aligned words in the other sentence provide clues:

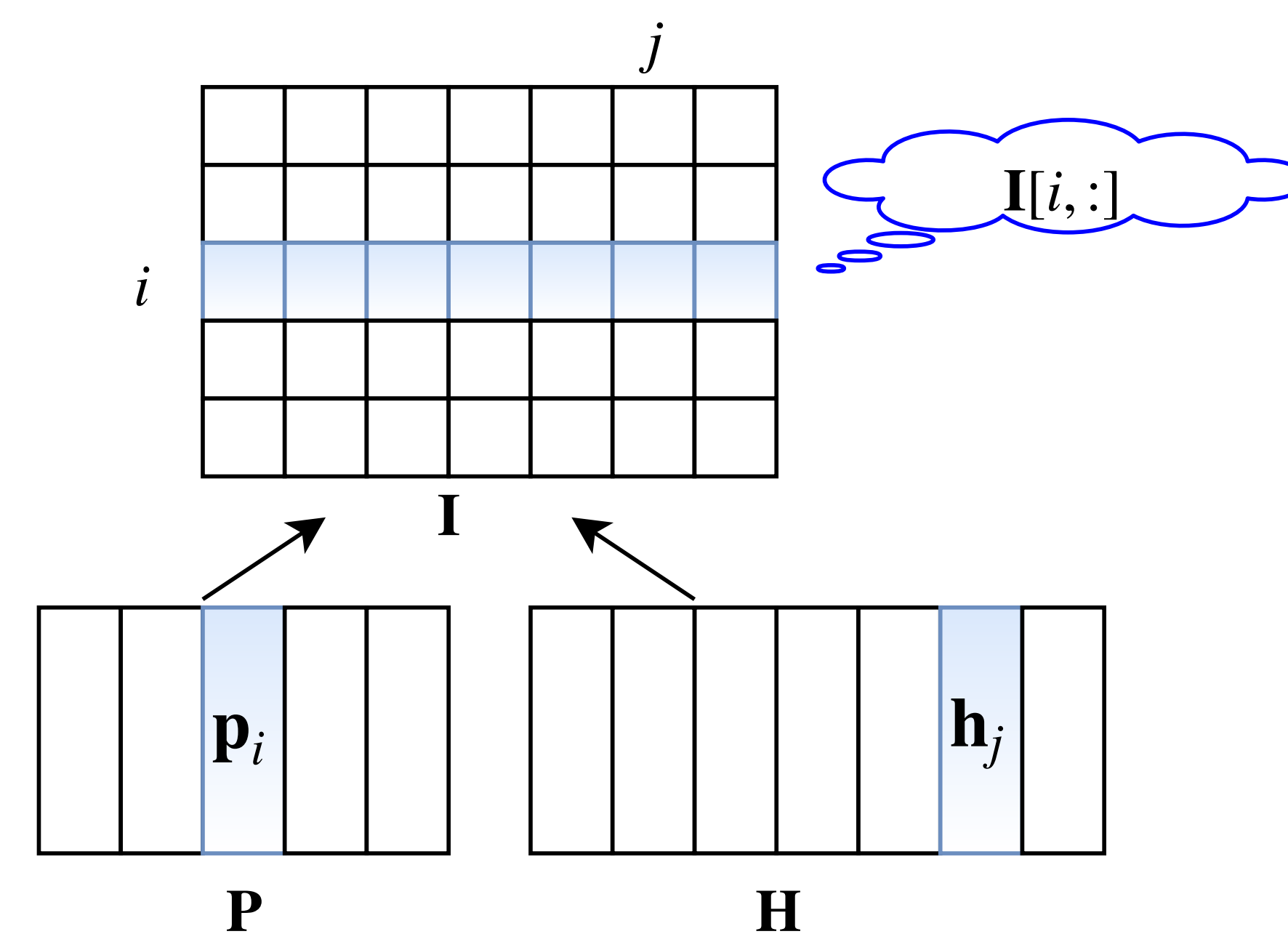


Basic principles

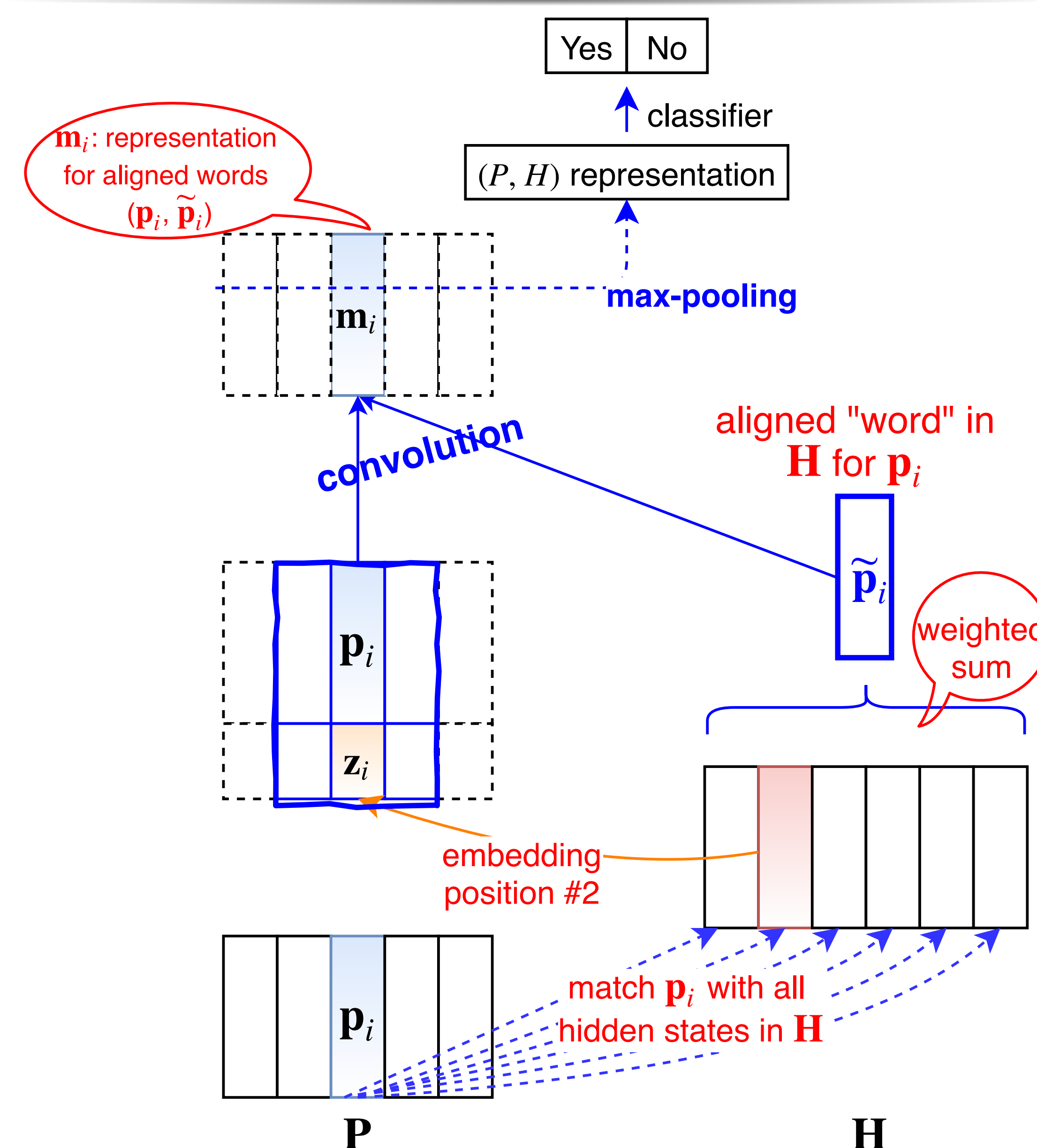
Given : **I** = interactions(**P**, **H**)

Learn :

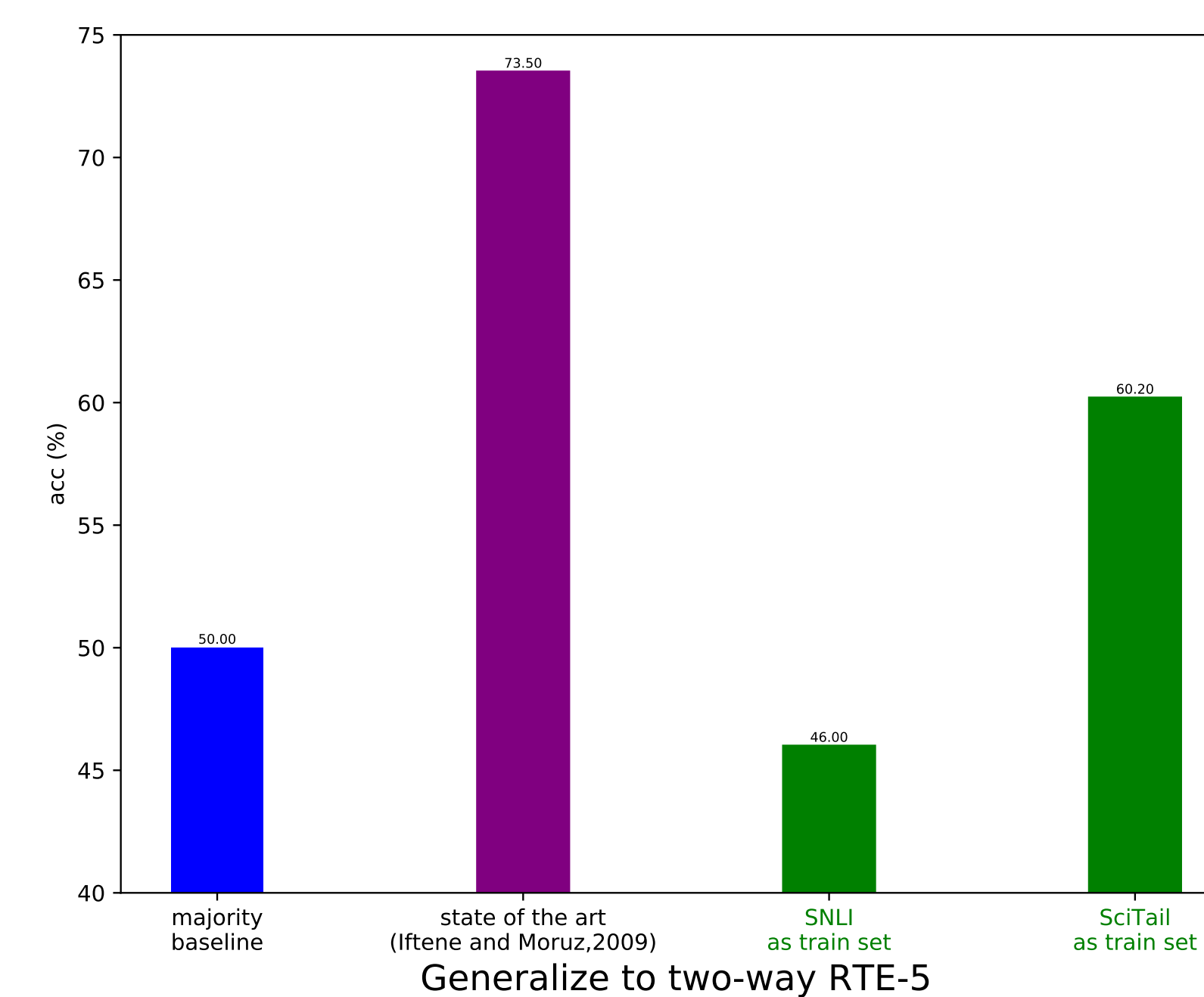
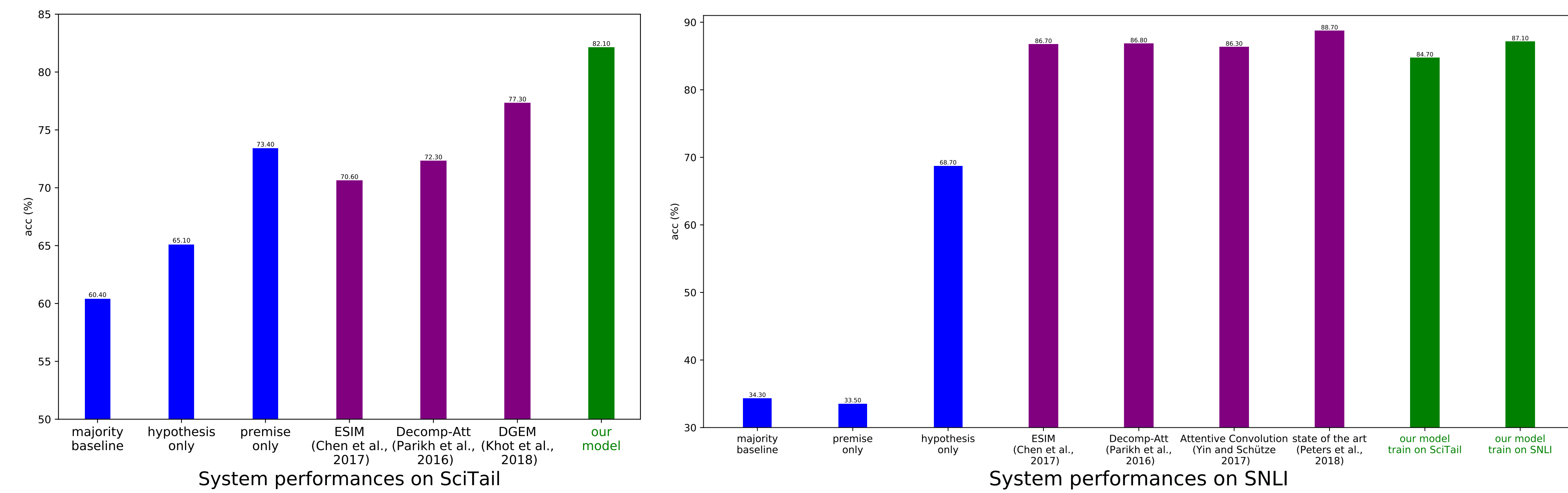
- (1) $\frac{1.0}{1.0 + \max(\mathbf{I}[i, :])}$: importance of \mathbf{p}_i
- (2) $\mathbf{H} \cdot \text{softmax}(\mathbf{I}[i, :])$: soft best match of \mathbf{p}_i
- (3) $\text{argmax}(\mathbf{I}[i, :])$: hard location of best match of \mathbf{p}_i



Architecture



Experiments



- Our model has **robust** and (close to) **top** performances in both SciTail and SNLI;
- SciTail enables the model to **generalize well** to RTE-5 and SNLI; SNLI fails to generalize;
- SciTail and SNLI both have some annotation artifacts -- refer to "premise-only" on SciTail, "hypothesis-only" on SNLI

Error cases

(Premise <i>P</i> , Hypothesis <i>H</i>) Pair	G/P	Challenge
(<i>P</i>) ... the notochord forms the backbone (or vertebral column). (<i>H</i>) Backbone is another name for the vertebral column.	1/0	language conventions
(<i>P</i>) ... animals with a vertebral column or backbone and animals without one. (<i>H</i>) Backbone is another name for the vertebral column.	1/0	ambiguity
(<i>P</i>) Heterotrophs get energy and carbon from living plants or animals (consumers) or from dead organic matter (decomposers). (<i>H</i>) Mushrooms get their energy from decomposing dead organisms.	0/1	discourse relation
(<i>P</i>) ... the SI unit... for force is the Newton (N) and is defined as (kg·m/s ⁻²). (<i>H</i>) Newton (N) is the SI unit for weight.	0/1	beyond text

Code & pretrained model: <https://github.com/yinwenpeng/SciTail>

References:

- Khot et al. *SciTail: A textual entailment dataset from science question answering*. In AACL'2018.