A large annotated corpus for learning natural language inference

"NLP Reading group (school of information"

Mohamed Salem Elaraby

3rd October, 2018

Outline

1. NLI
2. Applications of NLI
3. Stanford NLI
4. Data Collection
5. Modeling
6. Results

What is Natural Language Inference (NLI)

Why NLI

The semantic concepts of entailment and contradiction are central to all aspects of natural lan- guage meaning

What is NLI?

Given two sentences (a Hypothesis and a premise) we want to know if they contradict, infer or there is no relation between them

NLI applications

Note

NLI is a core NLP task, which means its a block that can be used in different applications

Popular Applications that uses NLI:

- Text Simplification
- Summarizing
- Paraphrasing

SNLI dataset

- The stanford NLI dataset is benchmark evaluation data for evaluating Natural Language Inference.
- Addresses the issues of size, quality, and indeterminacy by collecting a large scale dataset and carefully annotate it.

Data collection

- Using Amazon Mechanical Turk, annotators have the chance to annotate both sentences as `entailment`, `contradiction`, and `neurtal`
- A fourth Option to report problems was added to the task.

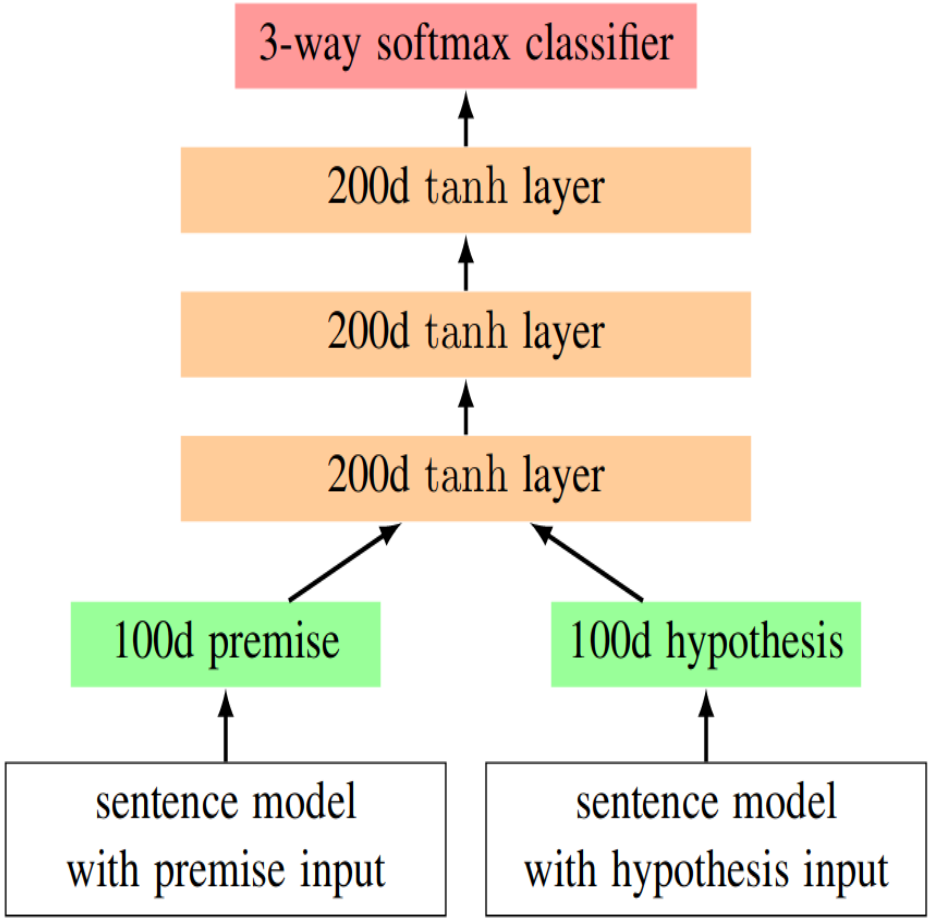
Examples from the dataset

A man inspects the uniform of a figure in some East Asian country.	**contradiction** C C C C C	The man is sleeping
An older and younger man smiling.	**neutral** N N E N N	Two men are smiling and laughing at the cats playing on the floor.
A black race car starts up in front of a crowd of people.	**contradiction** C C C C C	A man is driving down a lonely road.
A soccer game with multiple males playing.	**entailment** E E E E E	Some men are playing a sport.
A smiling costumed woman is holding an umbrella.	**neutral** N N E C N	A happy woman in a fairy costume holds an umbrella.

Data Stats

Data set sizes:	
Training pairs	550,152
Development pairs	10,000
Test pairs	10,000
Sentence length:	
Premise mean token count	14.1
Hypothesis mean token count	8.3

NLI Model


3-way softmax classifier
200d tanh layer
200d tanh layer
200d tanh layer
100d premise
100d hypothesis
sentence model with premise input
sentence model with hypothesis input

Results

- Results are obtained on both the SNLI and a standard dataset called SICK.
- In the SICK experiment, the SNLI data used to train the neural model and then obtained weights are used as initialization of the trained SICK models.

Table Results

Results of SNI dataset

Sentence model	Train	Test
100d Sum of words	79.3	75.3
100d RNN	73.1	72.2
100d LSTM RNN	84.8	**77.6**

SICK results

Training sets	Train	Test
Our data only	42.0	46.7
SICK only	100.0	71.3
Our data and SICK (transfer)	99.9	**80.8**