Seminar assignment: Natural language inference

Group-3

First paper:

NLI datasets: FraCaS and SMLI dataset:

FraCaS is a hand-crafted database that contains different entailment problems covering multiple semantic phenomena (such as references, quantifiers, etc.) and minimizing the impact of syntactic problems like and word sense ambiguity. This database is built to be used as a performance measure for inference systems that use logical relations between premises and hypotheses. It is small in nature and was built to test systems that relay on logical entailment.

NLI systems that utilize data-driven methods need a large dataset. Also, it contains real-world language where the relationship between pair of sentences is not strictly logical. The paper also discusses the issue with indeterminacy.

	FraCaS	SNLI
Entailment	Based on the logical relation between p and h	Based on statistical learning method to learn the relation between p and h
Usage	Testing tool for systems that rely on logical entailment	Training and validating data-driven systems
Dataset features	Tailored and designed by experts.	Written by humans; not automatic.
	Not everyday usage of language	More naturalistic corpus
	Small in size	Large in size

Interdemecy and SNLI data collection through image caption:

To ensure that the relationship between the produced premise and the provided hypnosis (image caption) is based solely on the available text, not on some assumptions on events and entity coreference that could be induced by the image.

Different NLI systems performance:

Both neural-based and lexicalized based systems have achieved the same accuracy, while the unlexicalized system achieves the worst. The neural-based system, unlike both lexicalized and lexicalized based systems, can learn from unstructured representations of sentence meaning. However to be able to achieve this performance they need a larger dataset. Yet the performance of all systems is limited to their ability to handle syntactic complexity and semantic problems like word sense disambiguation and compositionality representation.

Points to consider for discussion:

Since the annotators are humans, there might be a human bias due to their cultural backgrounds, personal motives, belief about the world and so on. Bias usually affects when we design a model and also when we evaluate a model.

Second paper:

Applications of NLI

In NLI, the goal is to model the relations between two pairs of sentences/phrases (premise/hypothesis); judging whether the hypothesis can be inferred from the premise or cannot be true. Modeling the relations between two sentences can be useful in paraphrase detection, automatic text summarization, and question answering.

HBMP heuristics

The reason for combining the representation of two sentences and feeding to a linear layer is to aggregate the relationships between these two sentences. This aggregation or relationship modeling is done by using simple heuristics which are:

- Concatenation of the two sentences vector; this is known as “Siamese” architecture which learns to differentiate between the two feed inputs.
- Element-wise product and difference to capture the similarities between the two sentences.

HBMP Advantages

Instead of using the standard configuration of stacked BLSTM, the authors separate each layer and initialize them with the hidden and cell states of the previous layer. This method

helps to improve the prediction as we go up to a higher level. Also, each layer is fed with the original word embedding so each layer remembers the input words, tuning the outputs as we go to a higher level and improves the performance.

Neural networks and NLI

The paper focuses on sentence encoding-based models. These models embed each sentence separately and then combine them using a linear layer. The main idea is to train the neural network to learn important linguistic properties of sentences and not to fall into the common problem of neural networks tending to capture a shallow statistical pattern from the training data. Examples of such shallow patterns that can be captured are assuming the entailment between two sentences if all words of the hypothesis are shown in the premises (word overlap), and assuming contradiction because of the word “not” is shown in the most of the pair of sentences that are labeled by contradiction in the training set. However, some of these problems may be avoided so the neural network can capture general abstractions of sentences that could be used for other NLP tasks.

Points to consider for discussion:

Can we use character encoding in an NLI neural-based system, if yes, what information can we capture from character by character sequence that we can use to model text entailment?