Natural Language Inference with Recurrent Neural Networks

David Torrejon
MSc in Artificial Intelligence. 2016

KU LEUVEN

Outline

- Natural Language Inference
    - Goals
- Word Representations and Deep Learning
    - Word Embeddings
    - RNN/LSTMs
- Task definition & Approach
- Results
    - Dataset
    - Results
    - Representation issues
- Conclusion & Future Work

Natural Language Inference

- Two boys, one wearing green, the other in yellow, are playing with sidewalk chalk, drawing pictures on the ground. *Premise*
- Two boys are drawing pictures with pink chalk. *Hypothesis*

Many applications:

- Information Retrieval
- Question Answering Systems
- Summarization tasks.

Natural Language Inference with RNNs

- Why use Neural Networks?
  - o Allow us to elegantly represent sentences.
  - o No domain knowledge is required.
- Problems with Neural Networks.
  - o Needs of data.
    - Partially solved with Stanford Natural Language Inference dataset.
  - o Representation inputs need to capture all the meaning of sentences.
    - Ambiguity
    - Common sense.
    - Anaphora resolution, quantifier resolution …

KU LEUVEN

Natural Language Inference: Goals

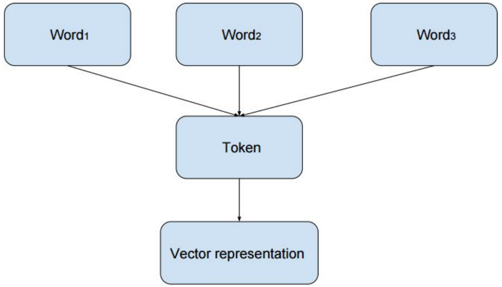
- Reproduce results from the best existing model.
- Study results achieved.
    - Why are there misclassified pairs of sentences?
    - Representation issues.
- Improve the results of the original paper.
    - Break the asymmetric relation

Word Representations

Mark Zuckerberg *represents* facebook at the convention. Speak or act for someone.

This new report *represents* the current situation. Describe something or someone.

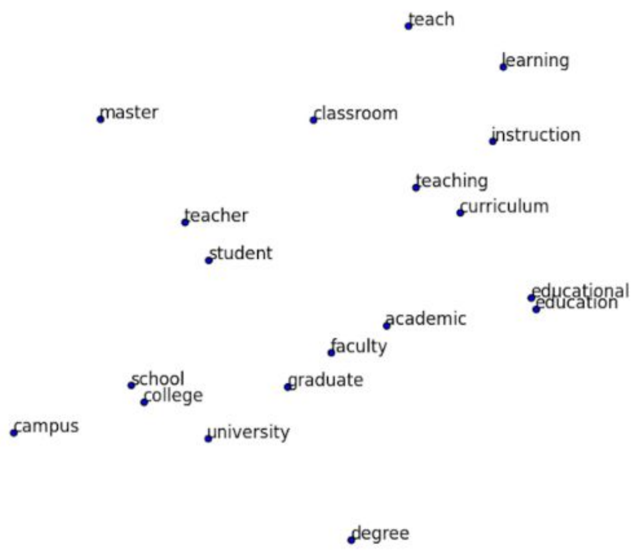
The book *represents* years of thoughts and research. To be the result of something.



* Sentences modified from: http://dictionary.cambridge.org/dictionary/english/represent

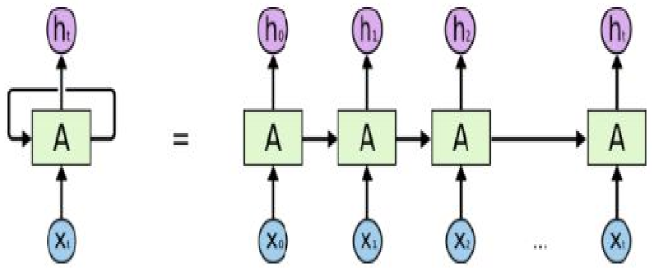
GloVe model

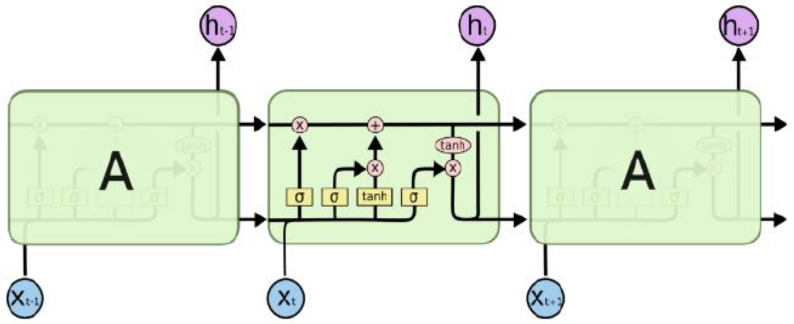
- Why? Coherent with best existing paper!
- Count based method that generates word embeddings.



Recurrent Neural Networks (LSTM)

- Elegantly encode sentences.





Task definition

- Formal definition of the inference task: p -> h

A woman with a green headscarf, blue shirt and a very big grin.

⇩

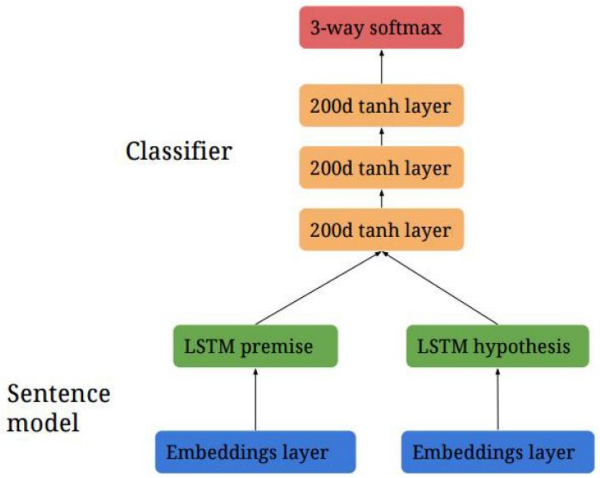
The woman is very happy.

- Given a premise and a hypothesis, allow the network to learn whether the relation is entailment, contradiction, or neutral. Neutral relations could be true, but can't be inferred. (According to the definition given by the authors of the dataset).

KU LEUVEN

Approach (1)

- Reproduce best performing model in SNLI.
- Network architecture differentiation between learning sentence representations and solving NLI.
- Network architecture modification
  - o Slight improvements if sentences encoded in 300d.
- Discuss influence of word embeddings
  - o As input or as parameters in the model
  - o Random vs pre-trained.

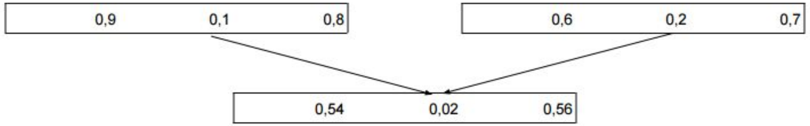


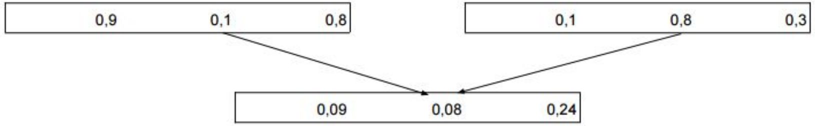
KU LEUVEN

Approach (2)

Breaking the asymmetry.

- In the original paper asymmetry is achieved by concatenating the sentence representations.
- Break this by multiplying the sentence representations.
- Highly differentiated inputs to the classifier between similar and non-similar sentence representations.





Dataset & Implementation

- Dataset format
- Dataset Distribution

Premise	Hypothesis	Manual Labels	Ground Truth
Premise 1	Hypothesis 1	[E E E E E]	E
Premise 1	Hypothesis 2	[E N E N N]	N
Premise 1	Hypothesis 3	[C C C N C]	C
Premise 2	Hypothesis 1	[E E C N E]	E

TABLE 5.1: SNLI corpus format

- Implementation done with Keras framework *

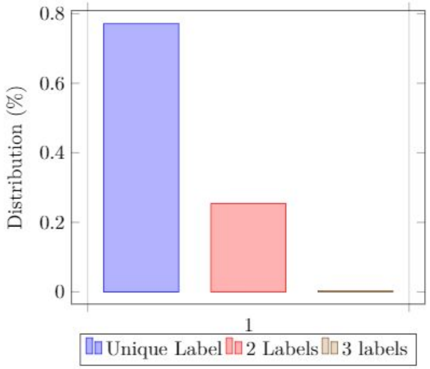


FIGURE 5.1: Data distribution for training set

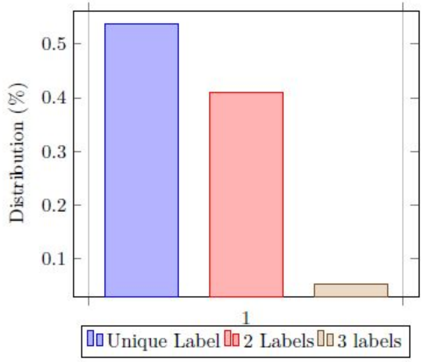
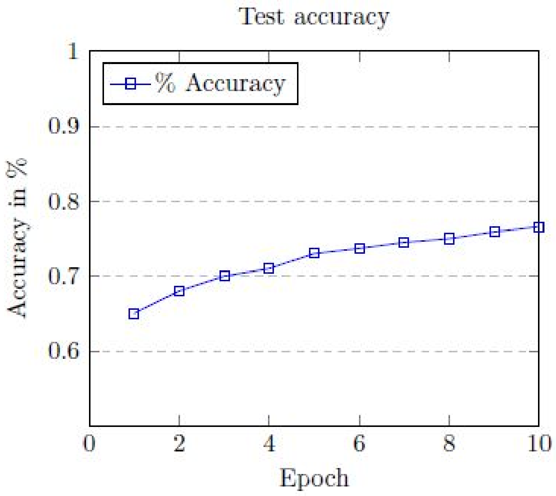


FIGURE 5.2: Data distribution for test set

* https://keras.io/

Results (1)

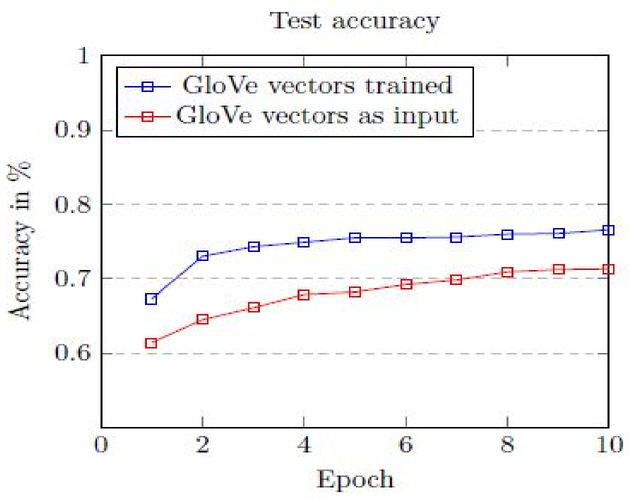
- Best performing model presented had an accuracy of 77.8%
- My implementation with the same network architecture is at 76.6%

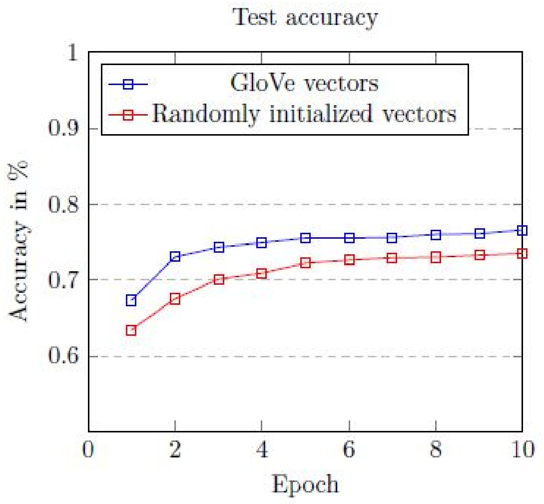


KU LEUVEN

Results (2) (word embeddings related)

- Tuning the parameters of the word embeddings help improve the accuracy.
- Study whether word embeddings perform better than random vectors.

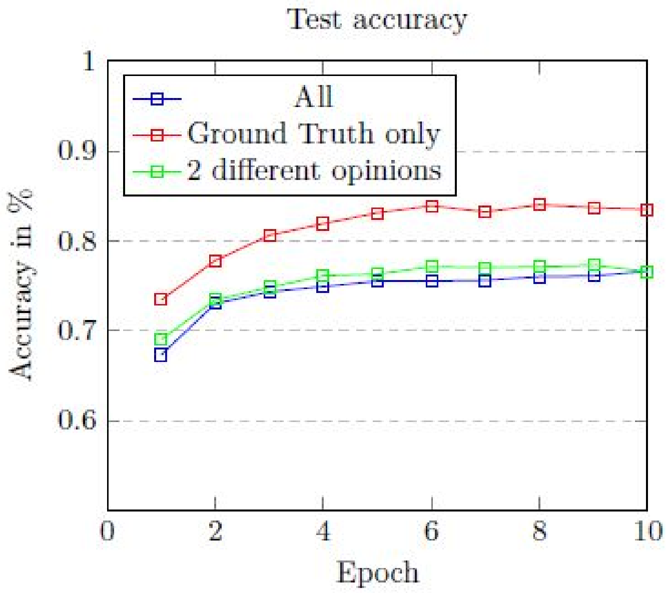




Results (3)

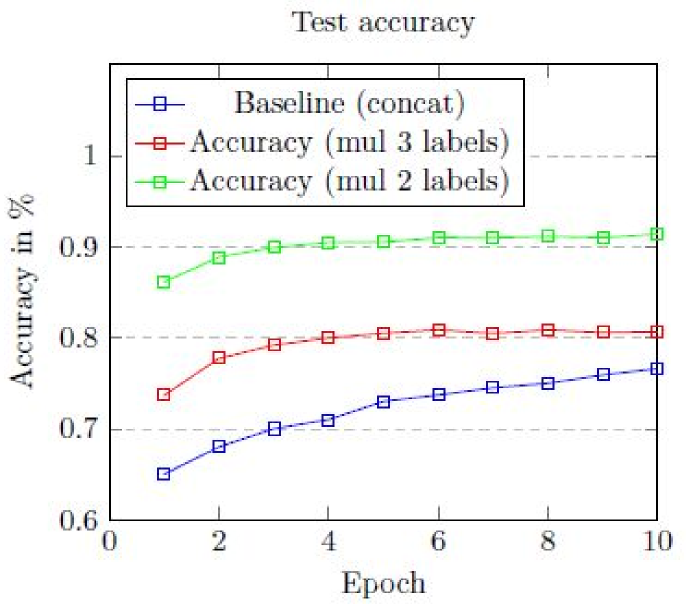
- Where do the mistakes come from?
    - When there is fully human agreement, the classification accuracy is better.

- Not able to deal with different sentence lengths. When there is a big difference in the length of the sentence pair, the model does not know what to do.



Results (4)

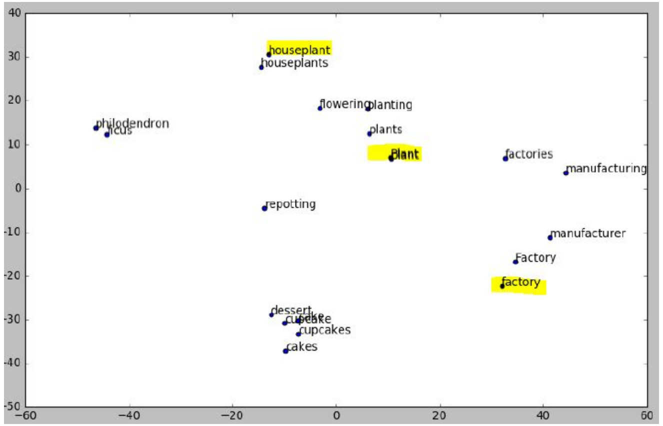
- When working with symmetric relations the accuracy gets 3% higher.
- No need to increase vector dimensions!
- In the SNLI dataset, if we remove Neutral labeled sentences, the model achieves over 90% accuracy.



KU LEUVEN

Representation Issues

- Natural language is hard to represent. In image processing, the raw input used are the actual values of the pixels. Still haven't found a way to do so in NLP tasks.
- Using one vector to encode all the possible meanings may not be good enough.
  - Already working on this. (Making Sense of Word Embeddings in ACL 2016).*
  - Instead of raw ¨token¨ embeddings, make word-sense embeddings.



* https://aclweb.org/anthology/W/W16/W16-1620.pdf

Conclusions

- NLI it is a complex task that involves topics like WSD (word-sense disambiguation). It is no easy task.
- Combination of word embeddings and LSTMs provides with a good baseline at dealing with the task.
- Training word embeddings along the network improves the results for the particular task.
- When we find human agreement, the model performs better.
- Bringing a symmetry relation helped the network perform better.

KU LEUVEN

Future work

- A main issue regarding all NLP tasks is the way we represent data.
  - o Introduce senses to the embeddings and other kind of relations.
  - o Study figures of speech to be grouped in vectors as if they were one token.
- Work with the sentence representation. Long sentences = mistakes!
- Improve & expand the datasets used to train this models.
- Interesting to study how a system trained in the SNLI dataset, performs in a real world task.

Thanks!

Code available in:

https://github.com/tmdavid/NLI_Code

People in the audience have questions. *Premise*

Someone is asking a question. *Hypothesis*