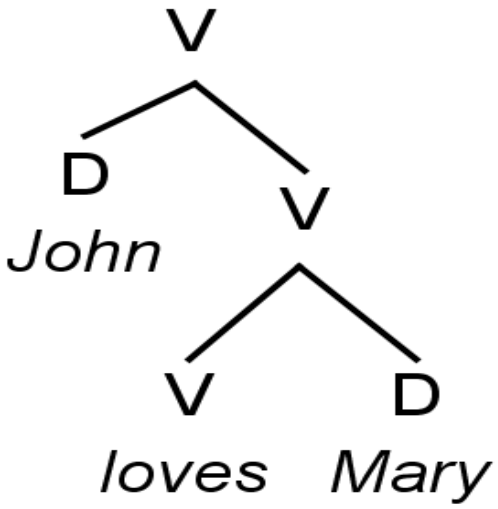
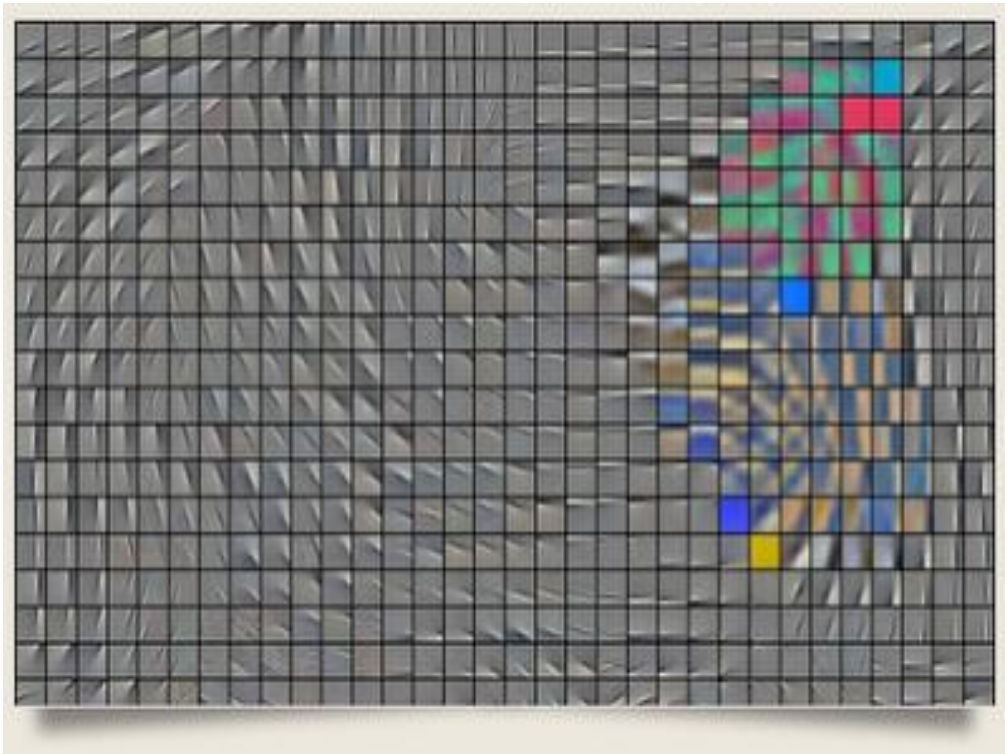
Compositionality through recursive neural nets

Philip Schulz





Some Facts about Trees

- Tree representations are directed graphs; so are neural nets!!!
- There is a one-to-one mapping from trees to neural nets
- This means that we don‘t need to come up with any fancy new methods

Some Facts about Neural Nets

- Because of their built-up they are sensitive to the sequential structure of their input
- They are extremely flexible since the neurons can implement many different functions

The Present Approach

- The lexicon consisted of 50-dim vectors which had been trained on text extracted from wikipedia
- Unlike Baroni and Zamparelli each word consists of a vector AND a matrix in this model

The Present Approach

- Therefore we have no type distinctions in this model
- In fact, all terminals and non-terminals have the same built-up
- However, by tweaking the parameters, the model of Baroni and Zamparelli can be replicated

How to Compute Non-terminals

$$p = f_{A,B(a,b)} = f(Ba, Ab)$$
$$= g(W \begin{pmatrix} Ba \\ Ab \end{pmatrix})$$

- Clearly, f depends on the input
- W is a global (|word| x 2*|word|)-matrix that reduces dimensionality
- g allows to introduce non-linearity

How to Compute Non-terminals

$$p = f_{A,B(a,b)} = f(Ba, Ab)$$

$$= g(W \begin{pmatrix} Ba \\ Ab \end{pmatrix})$$

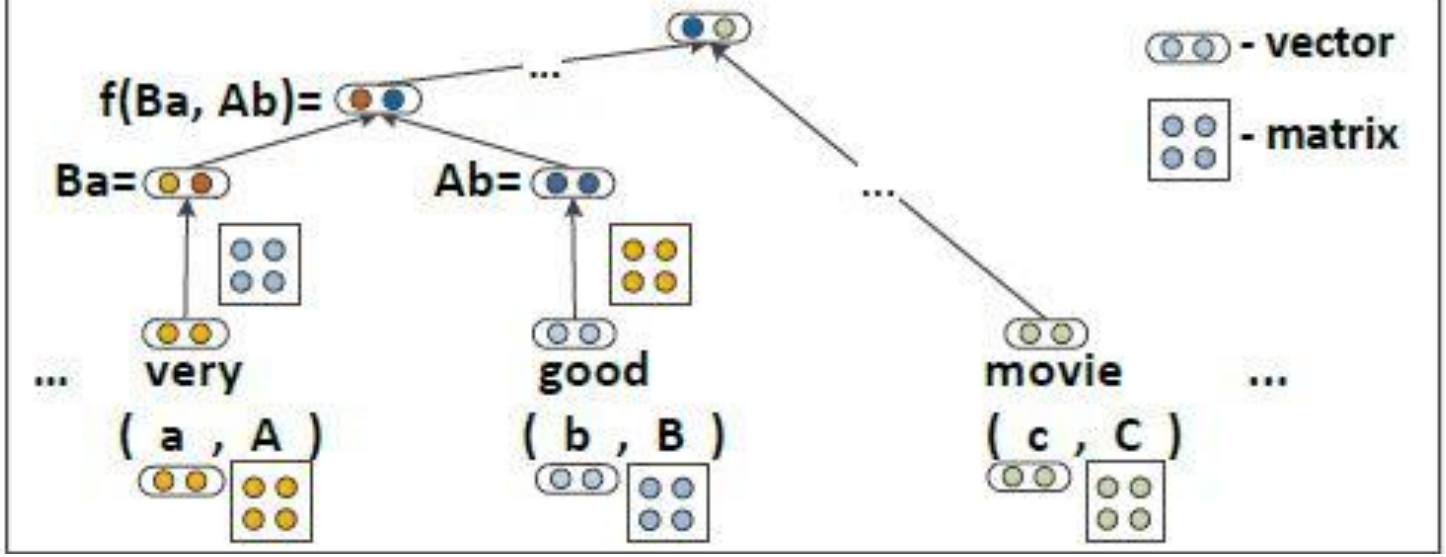
- This composition function is equivalent to a neuron!!!

… and now the matrices

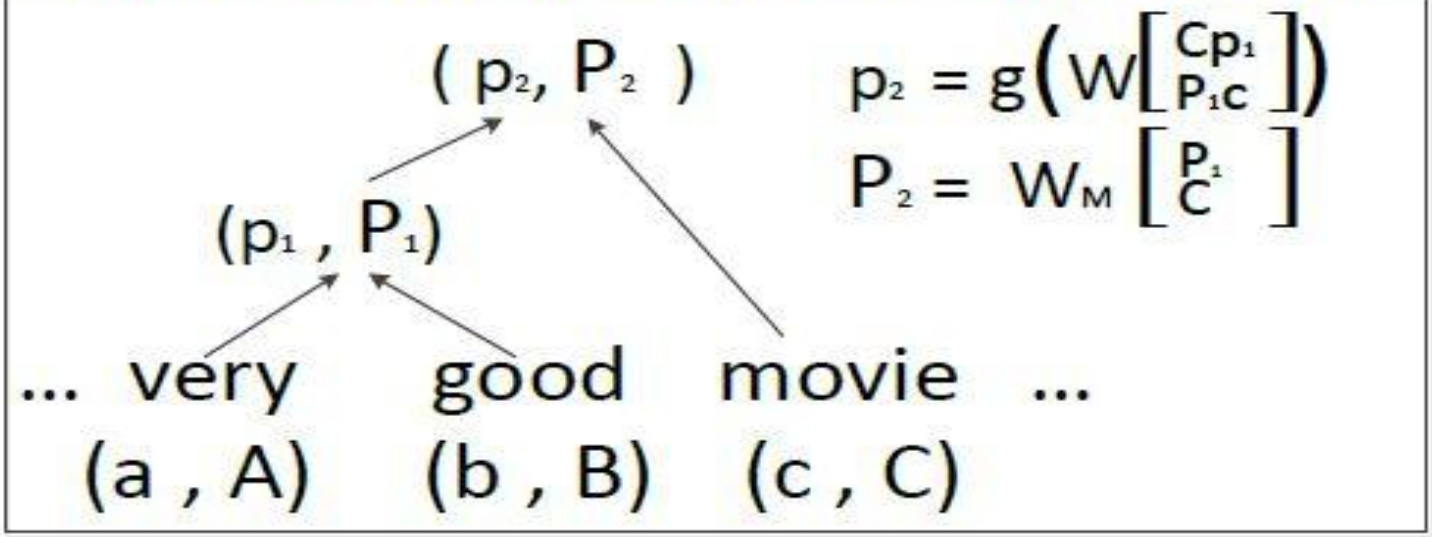
$$P = f_M(A, B) = W_M \begin{pmatrix} A \\ B \end{pmatrix}$$

- Again, W just reduces dimensions

Recursive Matrix-Vector Model



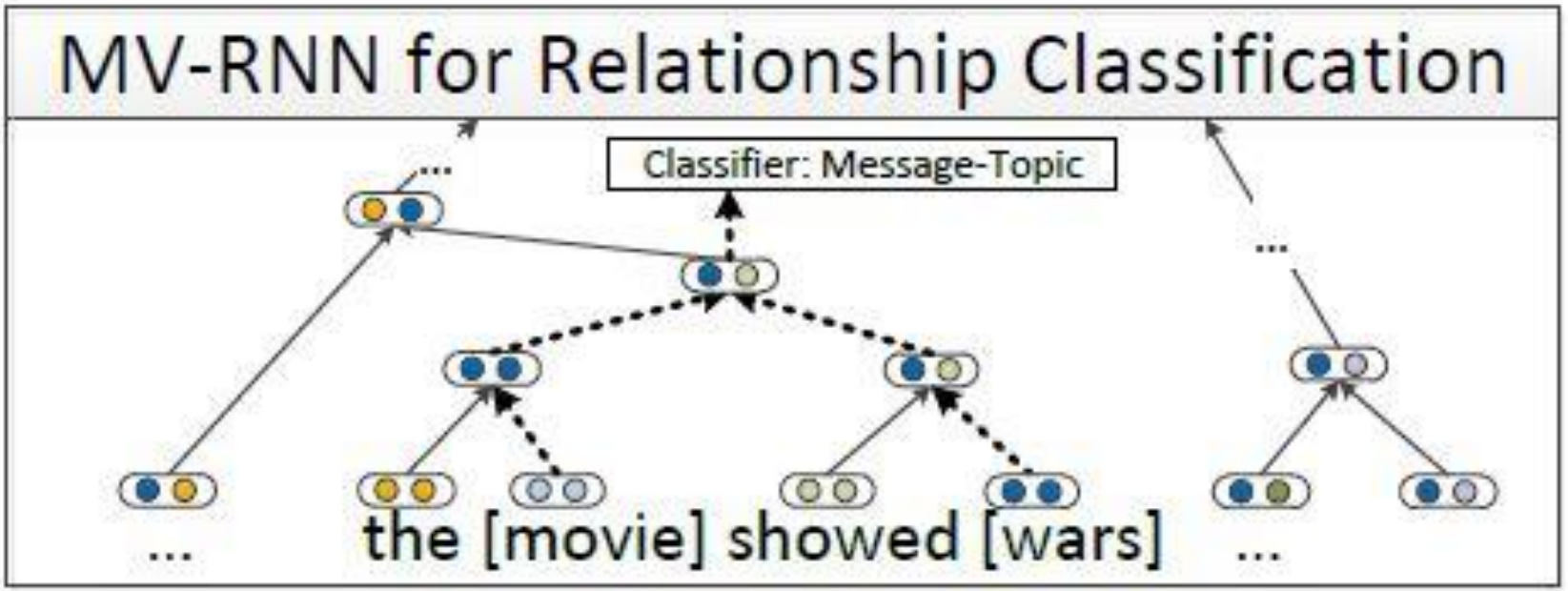
Matrix-Vector Recursive Neural Network



Evaluation

- Evaluation took place on three tasks; we will focus on the semantic relations task
- Semantic role labeling is done by finding the smallest constituent containing both targets

MV-RNN for Relationship Classification



Evaluation

- Interestingly, the classification is done in one step by only looking at the top node of the constituent
- This may possibly lead to methods for investigating long distance dependencies (e.g. in questions)

Summary

- We have seen a new approach that is completely general and in fact subsumes foregoing models
- Since we can map parse trees onto neural nets hardly any additional processing is required

Summary

- Neural Networks are well studied and currently there is a surge in interest in them; hence NLP may profit from results in other fields

Resources

- Socher et al. (2012): Semantic Compositionality through Recursive Matrix-Vector spaces
- Socher.org (including data and tutorial)
- Neural Networks for Machine Learning on Coursera (taught by Geoffrey Hinton!)