Paper presentation

Learning Continuous Phrase Representations and Syntactic Parsing with Recursive Neural Networks

Richard Socher, Christopher D. Manning, Andrew Y. Ng
Department of Computer Science
Stanford University
richard@socher.org, manning@stanford.edu, ang@cs.stanford.edu

Marco Damonte - 1333293

- Proper syntactic representations are of importance to tasks such as relation extraction and semantic role labeling

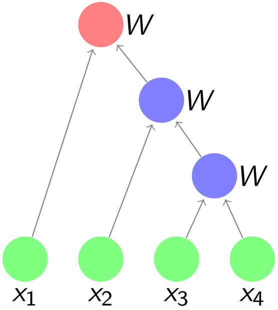
- *Recursive Neural Networks* (RNNs) can provide us with vector space representations that can be exploited during parsing

- It is even possible to jointly learn representation and parse using the same deep network

- *Deep networks* are commonly used to learn vector space representation of words $\rightarrow$ *Word Embeddings*

- RNNs generalize these embeddings for entire sentences

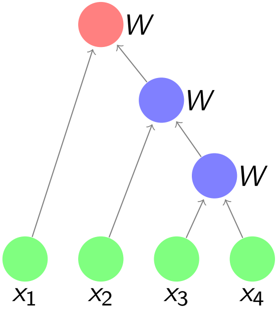
They are deep networks where we use the same set of weights recursively over a deep hierarchical structure



Methodology

Recursive Neural Networks

They are deep networks where we use the same set of weights recursively over a deep hierarchical structure

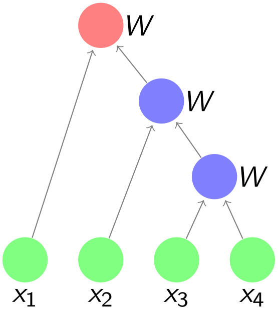


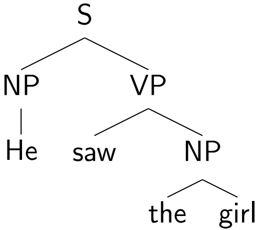
$$score = U^T p$$

$$p = tanh\left(W[x; y] + b\right)$$

You might ask: *Why do we believe RNNs are good for this?*

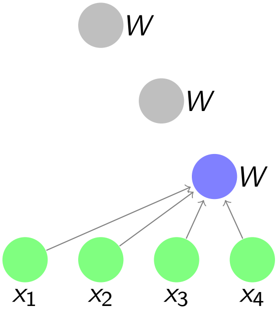
You might ask: *Why do we believe RNNs are good for this?*





Methodology

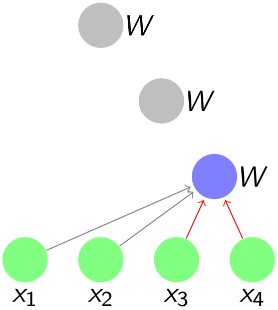
Model 1: Greedy RNN



$$\forall i, j$$

$$score_{i,j} = U^T p_{i,j}$$

$$p_{i,j} = tanh\left(W[x_i; x_j] + b\right)$$



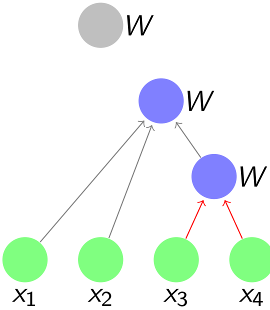
$$\forall i, j$$

$$score_{i,j} = U^T p_{i,j}$$

$$p_{i,j} = tanh\left(W[x_i; x_j] + b\right)$$

Methodology

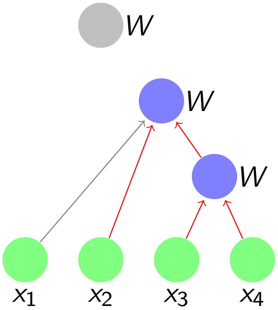
Model 1: Greedy RNN



$$\forall i, j$$

$$score_{i,j} = U^T p_{i,j}$$

$$p_{i,j} = tanh\left(W[x_i; x_j] + b\right)$$

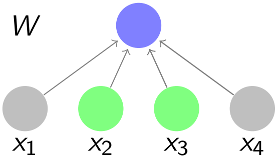


$$\forall i, j$$

$$score_{i,j} = U^T p_{i,j}$$

$$p_{i,j} = tanh\left(W[x_i; x_j] + b\right)$$

Context introduced in the first layer by allowing the representations of context words to modify the parsing decision



$$score_{2,3} = U^T p_{2,3}$$

$$p_{2,3} = tanh\left(W[x_1; x_2; x_3; x_4] + b\right)$$

Extension to greedy CRNN model: adding to each node a layer to predict class labels.

$$P(y = c|\mathbf{x}) = softmax(\mathbf{w}_c^\mathsf{T}\mathbf{x}) = \frac{e^{\mathbf{w}_c^\mathsf{T}\mathbf{x}}}{\sum_{c=1}^{C} e^{\mathbf{w}_c^\mathsf{T}\mathbf{x}}}$$

Instead of greedily collapse the best pairs:

- Formulate a global objective function that penalises choices far from the correct choice

$$\sum_i s(x_i, y_i) - \max_y (s(x_i, y) + \Delta(y, y_i)))$$

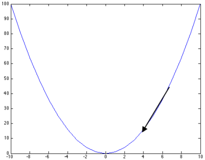
- *Beam Search* instead of *Greedy Search* to find the best parse

We need to determine the set of weights W:

- Backpropagation

$$(f(g(x)))' = f'(g(x))g'(x)$$

- Gradient descent

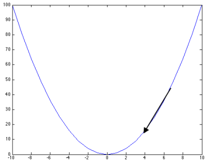


We need to determine the set of weights W:

- Backpropagation

$$(f(g(x)))' = f'(g(x))g'(x)$$

- Gradient descent



Actually..

- Backpropagation Through Structure (BTS)
- Subgradient method

Word	Initial Collobert&Weston embedding	After RNN training
the	a, its, an, this, his, their, UNK	an, The, no, some, A, these, another
and	,, or, but, as, –, ., that, for, in	or, but, And, But, &, least
in	at, on, from, for, over, after, and, as	In, from, into, since, for, For, like, with,
that	which, but, ,, and, –, as, for, or, about, if	what, who, if, this, some, which, If
said	added, says, –, while, but, reported, on	says, fell, added, did, rose, sold, reported
he	she, it, they, which, also, now, who, we	they, I, we, you, It, she, it, He, We, They
share	high, higher, business, market, current, stock, lower, increase, price, financial	increase, bank, income, industry, issue, state, sale, growth, unit, president
when	after, while, before, if, but, where, as	where, how, which, during, including

Center Phrase and Nearest Neighbors
(A) Sales grew almost 2 % to 222.2 million from 222.2 million.
1. Sales surged 22 % to 222.22 billion yen from 222.22 billion.
2. Revenue fell 2 % to 2.22 billion from 2.22 billion.
3. Sales rose more than 2 % to 22.2 million from 22.2 million.
4. Volume was 222.2 million shares, more than triple recent levels.

Method	F1
Model 1 (Greedy RNN)	76.55
Model 2 (Greedy, context-aware RNN)	83.36
Model 3 (Greedy, context-aware RNN + category classifier)	87.05
Model 4 (Beam, context-aware RNN + category classifier)	92.06
Left Corner PCFG, [MC97]	90.64
Current Implementation of the Stanford Parser, [KM03]	**93.98**

Conclusion

- Better word embeddings
- Sentence embeddings entailing syntactic and semantic information
- Almost state-of-the-art parsing