Relation Extraction: Perspective from Convolutional Neural Networks

Thien Huu Nguyen and Ralph Grishman. 2015a. Relation extraction: Perspective from convolutional neural networks. In The NAACL Workshop on Vector Space Modeling for NLP (VSM).

Deep Learning Reading Group Presentation
by Çağıl Sönmez

Problem

- Relation Extraction
- Nguyen and Grishman, 2015, proposes a CNN for relation extraction, where they use only sentence representations and no features generated by linguistic analysis, knowledge sources, gazetteers, etc. In other words their target is to avoid using any external features and resources.

MODEL

Their NN for RE consists of 4 layers:

1) Lookup tables for word embeddings

2) Convolutional layer to recongnize grams, includes filters with multiple window sizes

3) Pooling layer that finds most relevant features

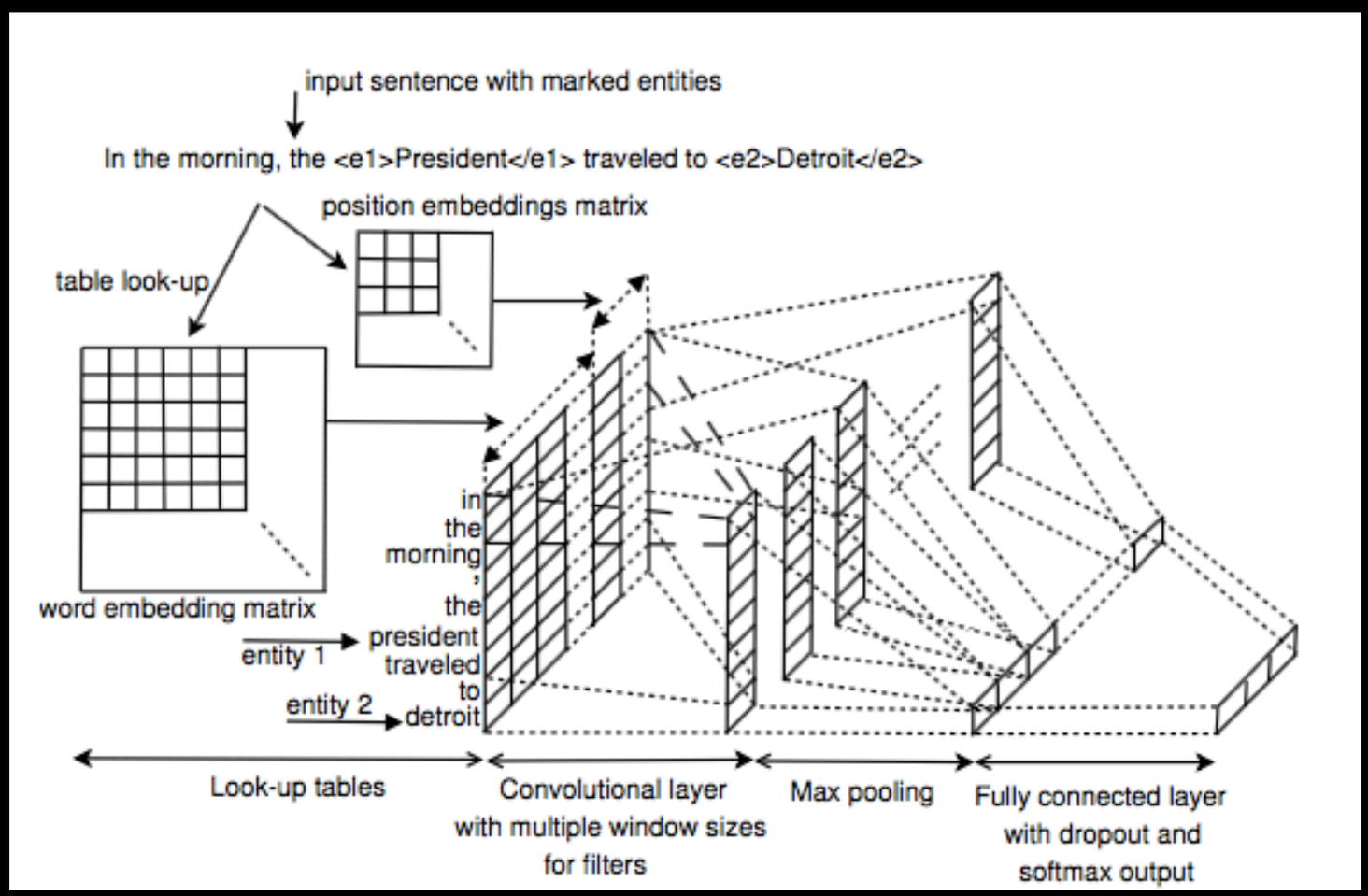
4) Logistic regression layer for classification

Word Representations

- The input consists of sentences marked with the two entity mentions of interest.
- Trimming and paddings
- Word embedding table
- Position embedding matrix (initialized randomly)

Convolution

- Filter f as representing some hidden class of the augmented n-grams and the scores si as measuring possibility of the corresponding hidden class



CNN

- Pooling
  - They use max-pooling to identify the most relavent features from the score sequences.
- Regularization and Classification
  - They uses dropout for the concatenated single feature vector from pooling scores of every filter before fed into a fully connected layer of neural network followed by a softmax layer.

Experiments

- tanh for the non-linear function
- 150 filters for each window size in the model
- position embedding vectors with $m^d = 50$
- dropout rate $\rho = 0.5$
- the mini-batch size of 50
- the hyperparameter for the l_2 is 3
- the pre-trained word embeddings word2vec (Mikolov et al. (2013))

EVALUATION

DATASETS

They evaluate their system with two datasets:

- SemEval-2010 Task 8 dataset for RC
- ACE 2005 dataset for RE.

Model Architectures

- the word embeddings and the position embeddings are randomly initialized and optimized during the training process (denoted by nonstatic.rand)

- the word embeddings are initialized by the pre- trained word embeddings; the position embeddings are initialized randomly and the two embeddings are kept unchanged during the training (denoted by static.word2vec)

- the two embeddings are initialized as in case (ii) but they are optimized as model parameters when the model is trained (denoted by nonstatic.word2vec)

Multiple Window Sizes

- The experiments show that using multiple window size combinations improves the results,
- they use a combination of 2,3,4,5 for the final setup

Balanced vs Unbalanced

- Their evaluation on balanced vs unbalanced dataset shows that systems perform their best when relation extraction task is reduced to relation classification.

Tasks

- When compared their proposed CNN with other **RE** systems, proposed CNN model outperforms the feature based systems if all systems uses no additional features
  - i.e, the words in the relation mentions, the positions of the two entity heads and the word embeddings.
- For **RC** the system achieves comparable results with other systems even though they use supervised and manual features
  - Again it outperforms the other systems when they do not use additional features such as POS, WordNet, name tagging, dependency parse, patterns etc.

Conclusions

- Easy reading, a very good intro paper
- Compares Relation Classification and Relation Extraction tasks, empirically shows that RE performs best when induced to RC task.
- Compares performance of system on balanced and unbalanced corpus, system performs best when positive/negative ratio is higher.
- Compares different settings for word embeddings, best performance is obtained when pre-trained embeddings are used and are optimised (non-static)
- Compares performance while using varying window sizes, best performance is obtained when all window sizes 2, 3, 4 and 5 (multiple window size model) are employed.