# Ranking Based on the paper 'Learning to Rank Short Text Pairs with Convolutional Deep Neural Networks'

## A) CNN Architecture

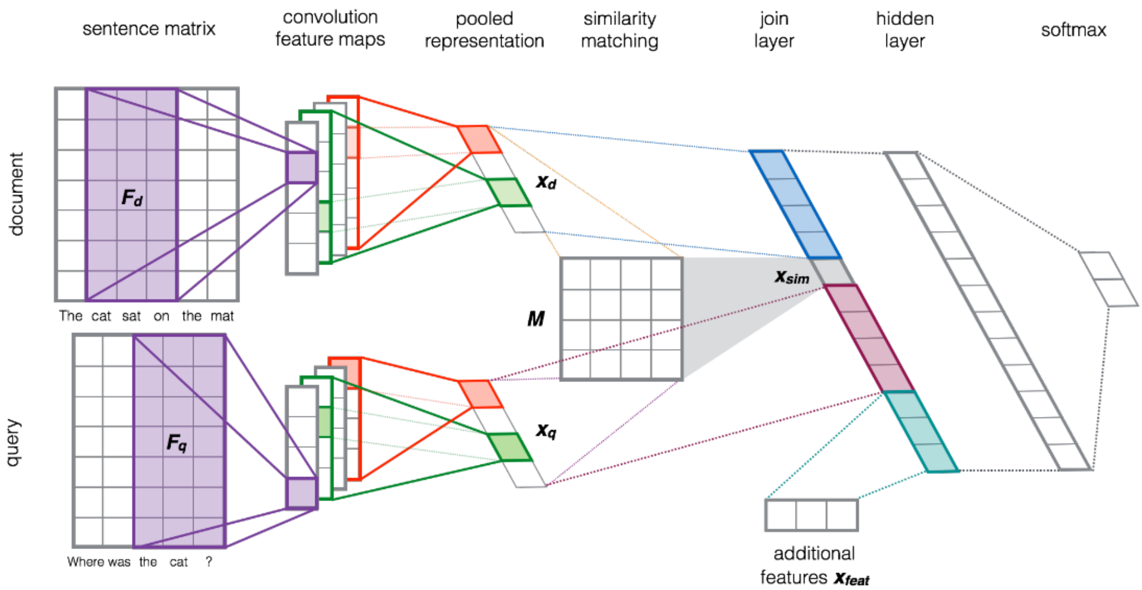


**Figure 2: Our deep learning architecture for reranking short text pairs.**

**1) Entity(document) and Query (Type) matrix:**

The input to our sentence model is a sentence s treated as a se-
quence of words: [w 1 , .., w |s| ]. For entity, it will be the context words in order of their positions. Where

Words are represented by distributional vectors w ∈ R^d looked up in a word embeddings matrix W ∈ R^d×|V |. For each input sentence s we build a sentence matrix S ∈ R d×|s| ,where each column i represents a word embedding w i at the corresponding position i in a sentence

**2) Convolution:** [Same as paper]
**3) Activation:** [Same as paper]
**4) Pooling:** [Same as paper]

Till here we will get the representation of type and entity
**5) Matching Type and Entity**

Given the output of our sentence ConvNets for processing queries and documents, their resulting vector representations x q and x d , can be used to compute a query-document similarity score.

$$sim(\mathbf{x}_q, \mathbf{x}_d) = \mathbf{x}_q^T \mathbf{M} \mathbf{x}_d, \qquad (2)$$

where $\mathbf{M} \in \mathbb{R}^{d \times d}$ is a similarity matrix. The Eq. 2 can be viewed

**Output Layer:** Last layers will be the probability distribution over labels.

## B) Non-linguistic features

Entity Salience Task could be used as non linguistic features

**1. A New Entity Salience Task with Millions of Training Examples:**

i) Index of the sentence in which the first mention of the entity appears.
ii) Number of times the head word of the entity's first mention appears.
iii) Number of mentions of that entity.

**Entity centrality:** All the features described above use only information available within the document. But articles are written with the assumption that the reader knows something about at least some of the entities involved. An entity may be mentioned just once but may be closely related to other entitites.

**Entity Centrality using Weighted page rank** as a measture. Where a directed edge between E1 -> E2 represents P(E2|E1), the probability of observing E2 in a document given that we have observed E1. We estimate P (E 2 |E 1 ) by counting the number of training documents in which E 1 and E 2 co-occur
and normalizing by the number of training documents in which E1 occurs.
The nodes' initial PageRank values act as a prior, where the uniform distribution, used in the classic PageRank algorithm, indicates a lack of prior knowledge. Since we have some prior signal about salience, we initialize the node values to the normalized mention counts of the entities in the document. We use a damping factor d, allowing random jumps between nodes with probability 1 − d, with the standard value d = 0.85.

**2. Understanding document aboutness-Identifying Salient Entities**

i) tf.idf of entity in D.
ii) length of D
iii) Norm frequency of entity

**3. SEL: a Unified Algorithm for Entity Linking and Saliency Detection**

Table 2: **Light Features for Supervised Candidate Pruning: features are relative to a candidate entity $c_j$**

| | |
|---|---|
| 1. **positions** | first, last, average, and standard deviation of the normalized positions of the spots referring to $c_j$ |
| 2. **first field positions** | document $D$ is subdivided in 4 fields: *the title*, *the first three sentences*, *the last three sentences*, and *the middle sentences*; the normalized position of the first spot referring to $c_j$ is computed for each field |
| 3. **average position in sentences** | the average position of spots referring to $c_j$ across the sentences of the document (salient entities are usually mentioned early) |
| 4. **field frequency** | number of spots referring to $c_j$ computed for each field of the document |
| 5. **capitalization** | True *iff* at least one mention of $c_j$ is capitalized |
| 6. **uppercase ratio** | maximum fraction of uppercase letters among the spots referring to $c_j$ |
| 7. **highlighting** | True *iff* at least one mention of $c_j$ is highlighted in bold or italic |
| 8. **average lengths** | average term- and character-based length of spots referring to $c_j$ |
| 9. **idf** | maximum Wikipedia inverse document frequency among the spots referring to $c_j$ |
| 10. **tf-idf** | maximum document spot frequency multiplied by *idf* among the spots referring to $c_j$ |
| 11. **is title** | True *iff* at least one mention of $c_j$ is present in the document title |
| 12. **link probabilities** | maximum and average *link probabilities* of the spots referring to $c_j$ |
| 13. **is name/person** | True *iff* at least one mention of $c_j$ is a common/person name (based on Yago – `http://goo.gl/glfBYN`) |
| 14. **entity frequency** | total number of spots referring to $c_j$ |
| 15. **distinct mentions** | number of distinct mentions referring to $c_j$ |
| 16. **not ambiguity** | True *iff* at least one mention of $c_j$ for which $c_j$ is the only candidate entity |
| 17. **ambiguity** | minimum, maximum and average ambiguity of the spots referring to $c_j$; spot ambiguity is defined as 1 minus the reciprocal of the number of candidate entities for the spot |
| 18. **commonness** | maximum and average *commonness* of the spots referring to $c_j$ |
| 19. **max commonness** $\times$ **max link probability** | maximum *commonness* multiplied by the maximum *link probability* among the spots referring to $c_j$ |
| 20. **entity degree** | in-degree, out-degree and (undirected) degree of $c_j$ in the Wikipedia citation graph |
| 21. **entity degree** $\times$ **max commonness** | maximum *commonness* among the spots of $c_j$ multiplied by the degree of $c_j$ |
| 22. **document length** | number of characters in $D$ |