Natural Language Processing: The Brain Way

Ronan Collobert and Jason Weston
NEC Laboratories America, Princeton, NJ
ronan@collobert.com, jaseweston@gmail.com

The field of Natural Language Processing (NLP) aims to convert human language into a formal representation that is easy for computers to manipulate. Current end applications include information extraction, machine translation, summarization, search and human-computer interfaces.

While complete semantic understanding is still a far-distant goal, researchers have taken a divide and conquer approach and identified several sub-tasks useful for application development and analysis. These range from the syntactic, such as part-of-speech tagging, chunking and parsing, to the semantic, such as word-sense disambiguation, semantic-role labeling, named entity extraction and anaphora resolution.

Currently, most research analyzes those tasks *separately*. Many systems possess few characteristics that would help develop a unified architecture which would presumably be necessary for deeper semantic tasks. In particular, many systems possess three failings in this regard: (i) they are *shallow* in the sense that the classifier is often linear, (ii) for good performance with a linear classifier they must incorporate many hand-engineered features specific for the task; and (iii) they cascade features learnt separately from other tasks, thus propagating errors.

We propose instead a completely different approach: we advocate the *brain way*, where features for a task are *implicitly trained* in a very deep architecture. We describe a general end-to-end system based on a single convolutional neural network architecture. Given a sentence, *without any engineered features*, it outputs a host of language processing predictions: part-of-speech tags, chunks, named entity tags, semantic roles, semantically similar words and the likelihood that the sentence makes sense (grammatically and semantically) using a language model.

The entire network is trained *jointly* on all these tasks using weight-sharing, an instance of *multitask learning*. All the tasks use labeled data except the language model which is learnt from unlabeled text and represents a novel way of performing *semi-supervised learning* for the shared tasks. We used for this purpose Wikipedia (`http://en.wikipedia.org`), leveraging *631M* unlabeled words into our system.

We show how both *multitask learning* and *semi-supervised learning* improve the generalization of the shared tasks, resulting in a learnt model with state-of-the-art performance for semantic role labeling, chunking, part of speech tagging and name entity recognition. Amazingly, this is achieved with *no prior knowledge* (we only use the words as input) whereas all existing approaches use a raft of hand-crafted features. As a plus, we give a taste in Table 1 of the impressive embedding obtained by our language model, which managed to learn semantic information even for very uncommon words (for a dictionary size of $30,000$ words).

Table 1: Performance of our language model. Embedding in $wsz = 50$ dimensions, dictionary size: $30,000$. For each column the queried word is followed by its index in the dictionary (higher means more rare) and its 10 nearest neighbors (using the Euclidean metric).

DEEP	LEARNING	ROCKS	VLADIMIR	DESERTED
1392	1642	3571	6582	10612
SHALLOW	THINKING	FORESTS	VIKTOR	DROWNED
DENSE	KNOWLEDGE	SEDIMENTS	ALEKSANDR	CURSED
BRIGHT	EDUCATION	VALLEYS	MIKHAIL	BESIEGED
MUDDY	TEACHING	RIDGES	ALFRED	WRECKED
MOIST	MEDITATION	DUNES	NIKOLAI	FALLEN
DEEPER	PSYCHOLOGY	MOUNTAINS	OSKAR	RESCUED
DARK	PERFORMING	CLOUDS	JOSEF	FORTIFIED
ROCKY	CHANGING	CLIFFS	ANDREI	RUINED
NARROW	MATHEMATICS	DROPLETS	GIUSEPPE	BURNED
WARM	DRAWING	HILLS	SERGEI	DECAPITATED

Topic: natural language processing
Preference: oral