Part-of-Speech Tagging using Contextual Word Embeddings

Juhani Dickinson
jdickin9@uwo.ca

Paul Moore
email address or ORCID

***Abstract*—ABSTRACT**

***Index Terms*—pos tagging, part of speech, nlp, word embedding, contextual word embedding**

I. Introduction

A. Parts of Speech

In a sentence, each word has a syntactic role to play. In the sentence "The brown dog runs quickly towards food.", "dog" denotes a thing, "brown" modifies or describes "dog", "runs" denotes an action, and so on. These syntactic roles are commonly referred to as "parts of speech", and are mapped to words using part-of-speech tags. However, this is not a one-to-one mapping. Many words have multiple meanings, and in many cases, these different meanings play different syntactic roles, mapping them to different parts of speech. For example, the word "orange", spelled and pronounced identically, can be either a noun or an adjective, as seen in sentences (1) and (2).

"I often eat an orange after working out." (Noun) (1)

"The orange car has arrived." (Adjective) (2)

Figuring out what part of speech a word has is not an unsolvable task, however. Where a word is in the sentence and what words surround it play a major role in disambiguating between syntactic roles. In (1), the word after "orange" is "after", which is an adverb. "Orange" must is either an adjective or a noun, and since the following word is an adverb, it cannot be an adjective. In (2), the words surrounding "orange" are "the" and "car", which are a determiner and a noun respectively. In this position, "orange" describes "car", making it an adjective.

B. Part-of-Speech (PoS) Tagging

The task of PoS tagging (often simply called tagging) is as follows: given a sentence, label each word in the sentence with the PoS tag that describes its syntactic role, seen below for sentence (3).

(3) Can you spot the large spot on your nose ?
AUX D VB D ADJ N PP D N .

What makes the task of tagging challenging is the contextual requirement: without it, both instances of "spot" in sentence (3) would be tagged as either VB or N. Because of this, non-naive tagger implementations leverage some level of context to perform their task. Simple implementations like n-gram taggers assign tags to words based on both the word itself and the tags of the preceding n words.

C. Word Embeddings

A common issue with simple tagging implementations is resistance to parallelisation. Any operation below the level of sentence must be linear, as the tag of the current word depends on the tag(s) of the previous words. To get around the linear-time requirement, some taggers use word embeddings, which represent words as vectors. This embedding takes into account the surrounding words, so that the meaning and context of each word is stored in the embedding. Using embeddings, tagging can be parallelised at the level of individual words.

1) Static Word Embeddings: pass

2) Contextual Word Embeddings: pass

II. Related Work

pass

III. Motivation

Part-of-speech tagging plays a pivotal role in machine translation. ¡¿ For any two languages, it is highly likely that the word order in one is different in some way from the word order in the other. For instance, adjectives in French come after the noun they modify ("les ciseaux rouge"), while in English they come before that same noun ("the red scissors"). Since syntactic roles and thus parts of speech are invariant between languages, part-of-speech tags are useful for knowing how and where to move words and phrases when translating.

IV. Methods

pass

V. Experimental Results

pass

VI. Conclusions

pass

References

[1] Pass