

Recent Trends in Deep Learning Based Natural Language Processing

IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE | AUGUST 2018

Tom Young Devamanyu Hazarika Soujanya Poria

Citations: 603

Summary and notes by: Abir Azad | 160041024

Distributed Representation

A. Word Embeddings

Implementing Vectors to capture the characteristics of the neighbors of a word. They are efficient in capturing context similarity.

B. Word2vec

Word embeddings was popularized by Mikolov who proposed the continuous bag-of-words (CBOW) and skip-gram models. CBOW computes the conditional probability of a target word given the context words surrounding it across a window of size k . On the other hand, the skip-gram model does the exact opposite of the CBOW model, by predicting the surrounding context words given the central target word.

Sometimes words such as good and bad share almost the same embedding [23], which is problematic if used in tasks such as sentiment analysis.

C. Character Embeddings

Word embeddings are able to capture syntactic and semantic information, yet for tasks such as POS-tagging and NER, intra-word morphological and shape information can also be very useful. It produces better results on morphologically rich languages like Portuguese and Spanish.

Convolutional Neural Networks

A. Basic CNN

1. Sentence Modeling

An architecture to allow modeling of complete sentences into sentence representations.

2. Window Approach

NLP tasks, such as NER, POS tagging, and SRL, require word-based predictions. To adapt CNNs for such tasks, a window approach is used, which assumes that the tag of a word primarily depends on its neighboring words.

B. Applications

Here different applications of CNN models over the years are shown

Recurrent Neural Networks

RNNs use the idea of processing sequential information. The term “recurrent” applies as they perform the same computation over each token of the sequence and each step is dependent on the previous computations and results.

A. Need for Recurrent Networks

it has the ability to capture the inherent sequential nature present in language, where units are characters, words or even sentences. Words in a language develop their semantical meaning based on the previous words in the sentence. A simple example stating this would be the difference in meaning between “dog” and “hot dog”.

While RNNs try to create a composition of an arbitrarily long sentence along with unbounded context, CNNs try to extract the most important n-grams

B. RNN Models

1. Simple Run
2. long short-term memory
3. gated Recurrent units

C. Applications

1. RNN for Word-level Classification
2. Rnn for sentence-level Classification
3. Rnn for generating language

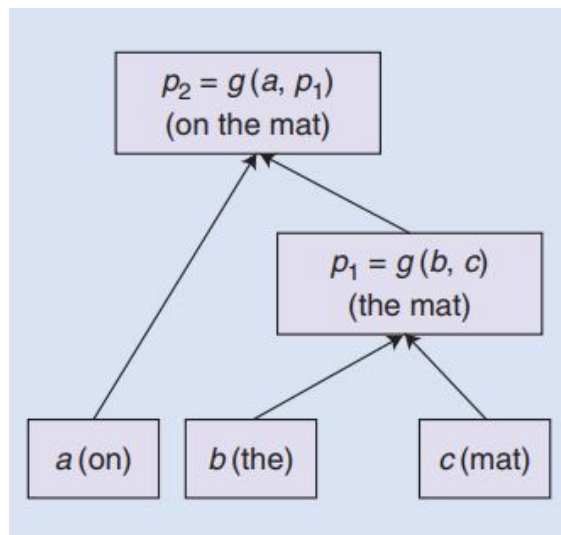
D. Attention Mechanism

Task of summarization, machine translation requires it

Recursive Neural Networks

It is tree structured model. Language exhibits a natural recursive structure, where words and sub-phrases combine into phrases in a hierarchical manner. Thus this tree models are used to better make use of such syntactic interpretations of sentence structure.

A. Basic Model



B. Applications

A scoring function is defined on the phrase representation to calculate the plausibility of that phrase.

Deep Reinforced Models and Deep Unsupervised Learning

A. Reinforcement Learning for Sequence Generation

Reinforcement learning is a method of training an agent to perform discrete actions before obtaining a reward.

B. Unsupervised Sentence Representation Learning

C. Deep Generative Models

Memory-Augmented Networks

For producing better result in QA type tasks.

Performance of Different Models on Different NLP Tasks

(**Possible Sectors**)

A. POS Tagging

Detecting the Parts of Speech within a sentence

B. Parsing

Mostly the implementation of grammar of a language in sentences. Whether “Man bites dog” or “Dog bites man”

C. Named-Entity Recognition

(wikipedia)...

A subtask of information extraction that seeks to locate and classify named entity mentioned in unstructured text into predefined categories such as person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc.

Most research on NER systems has been structured as taking an unannotated block of text, such as this one:

Jim bought 300 shares of Acme Corp. in 2006.

And producing an annotated block of text that highlights the names of entities:

[Jim]_{Person} bought 300 shares of [Acme Corp.]_{Organization} in [2006]_{Time}.

In this example, a person name consisting of one token, a two-token company name and a temporal expression have been detected and classified.

D. Semantic Role Labeling

Semantic role labeling (SRL) aims to discover the predicate-argument structure of each

predicate in a sentence. For each target verb (predicate), all constituents in the sentence which take a semantic role of the verb are recognized

E. Sentiment Classification

Basic task in sentiment analysis is classifying the *polarity* of a given text in the document.

F. Machine Translation

G. Question Answering

H. Dialogue Systems

Maintaining an effective dialogue.

Conclusion

Supervised learning is the most popular practice in recent deep learning research for NLP. Authors expect to see more development in

.reinforcement learning methods, e.g., dialogue system

.multimodal learning [164] as, in the real world, language is often grounded on (or correlated with) other signals

.Natural Language Understanding

Points

LSTM

gated recurrent units (GRUs)

residual networks (ResNets)