



# Character-level Convolutional Networks for Text Classification



Pavel Surmenok · Follow

3 min read · Jun 13, 2016



216



1



One of the common natural language understanding problems is text classification. Over last few decades, machine learning researchers have been moving from the simplest “bag of words” model to more sophisticated models for text classification.

Bag of words model uses only information about which words are used in the text. Adding TFIDF to the bag of words helps to track relevancy of each word to the document. Bag of n-grams enables using partial information about structure of the text. Recurrent neural networks, like LSTM, can capture dependencies between words even if they are far from each other. LSTM learns structure of sentences from the raw data, but we still have to provide a list of words. Word2vec algorithm adds knowledge about word similarity, which helps a lot. Convolutional neural networks can also help to process word-based datasets.

A trend is to learn using raw data, and provide machine learning models with an access to more information about text structure. A logical next step

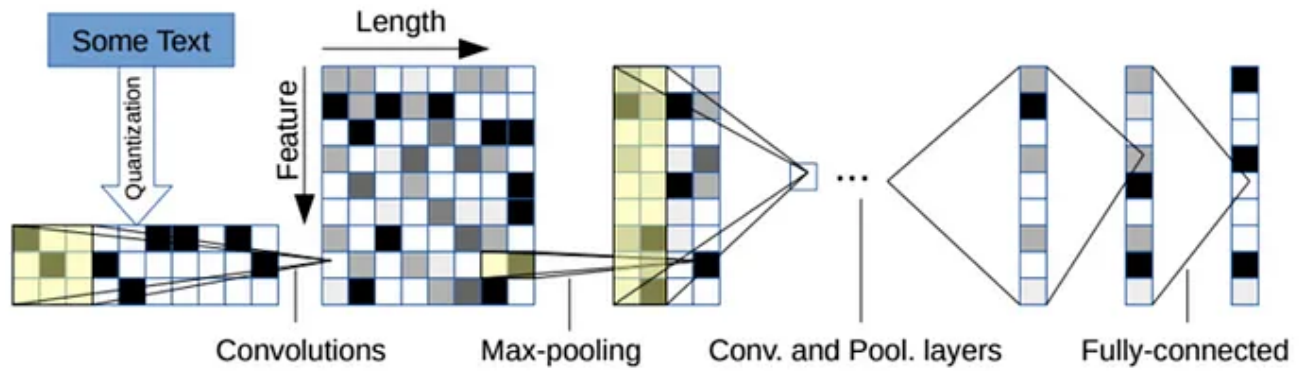
would be to feed a stream of characters to the model and let it learn all about the words. What can be cruder than a stream of characters? An additional benefit is that the model can learn misspellings and emoticons. Also, the same model can be used for different languages, even those where segmentation into words is not possible.

The article “Character-level Convolutional Networks for Text Classification” (Xiang Zhang, Junbo Zhao, Yann LeCun) explores usage of character-level ConvNet networks for text classification. They compare performance of a few different models on several large-scale datasets.

Datasets contain from 120,000 to 3,600,000 samples in the training set, from 2 to 14 classes. The smallest dataset is AG’s News: news articles divided into 4 classes, 30,000 articles for each class in the training set. The largest is Amazon Review Polarity: 2 polarity segments with 1,800,000 reviews for each of them.

Character-level ConvNet was compared with state-of-the-art models: Bag-of-words and its TFIDF, Bag-of-ngrams and its TFIDF, Bag-of-means on word embedding, word-based ConvNet, word-based LSTM.

Character-level ConvNet contains 6 convolutional layers and 3 fully-connected layers.



Source: <http://arxiv.org/pdf/1509.01626v3.pdf>

Results are quite interesting. N-gram and N-gram TFIDF models have are the best for smaller datasets, up to several hundreds of thousands of samples. But when dataset size grows to several million we can observe that character-level ConvNet performs better.

ConvNet tends to work better for texts which are less curated. For example, ConvNet works better on Amazon reviews dataset. Amazon Reviews are raw user inputs, whereas users could be more careful in writings on Yahoo! Answers.

Choice of alphabet matters. ConvNet works better is not distinguishing between upper and lower case characters.

[Another overview of this paper](#)

[Discussion on Reddit](#)

[Torch 7 code from Xiang Zhang](#)

[Keras \(Theano\) implementation](#)

Other useful links:

## Understanding Convolutional Neural Networks for NLP

### A Set of Character-Based Models for Sentiment Analysis, Ad Blocking and other tasks

*This article was originally published on*

<http://pavel.surmenok.com/2016/06/12/character-level-convolutional-networks-for-text-classification/>

Machine Learning

Deep Learning

Natural Language

#### More from the list: "NLP"

Curated by Himanshu Birla



Jon Gi... in Towards Data ...

#### **Characteristics of Word Embeddings**

★ · 11 min read · Sep 4, 2021



Jon Gi... in Towards Data ...

#### **The Word2vec Hyperparameters**

★ · 6 min read · Sep 3, 2021



Jon Gi... in

#### **The Word2vec**

★ · 15 min read



[View list](#)



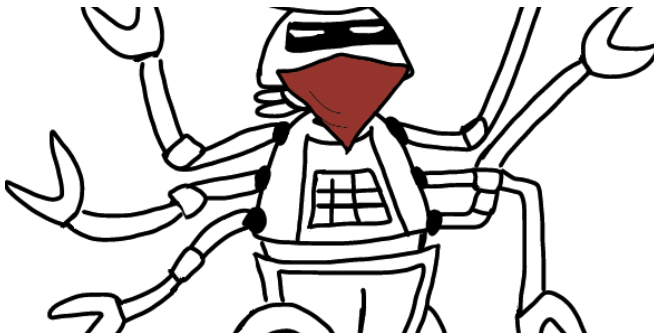
## Written by Pavel Surmenok

[Follow](#)

2.6K Followers

Machine learning engineering and self-driving cars. Opinions expressed are solely my own and do not express the views or opinions of my employer.

### More from Pavel Surmenok



Pavel Surmenok in Towards Data Science

## Contextual Bandits and Reinforcement Learning

If you develop personalization of user experience for your website or an app,...

6 min read · Aug 27, 2017



1.3K



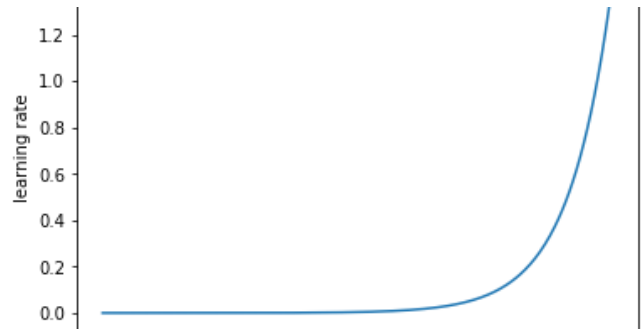
9



3K



18

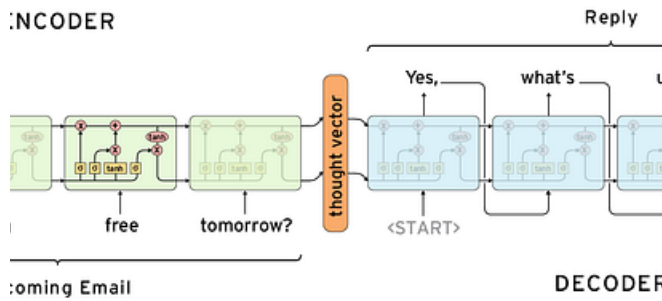


Pavel Surmenok in Towards Data Science

## Estimating an Optimal Learning Rate For a Deep Neural Network

The learning rate is one of the most important hyper-parameters to tune for training deep...

5 min read · Nov 13, 2017



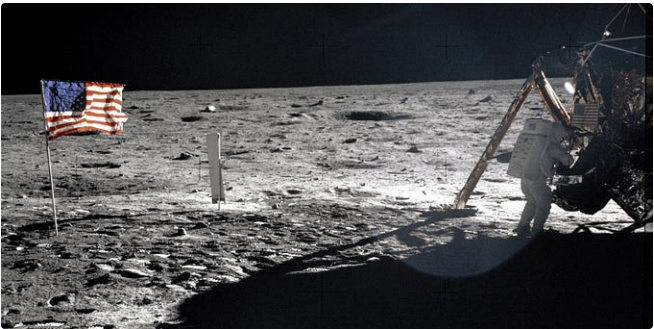
 Pavel Surmenok


**How to Run Text Summarization with TensorFlow**

Text summarization problem has many useful applications. If you run a website, you can...

6 min read · Oct 16, 2016

 667  14  



 Pavel Surmenok

**Learning Kalman Filter**

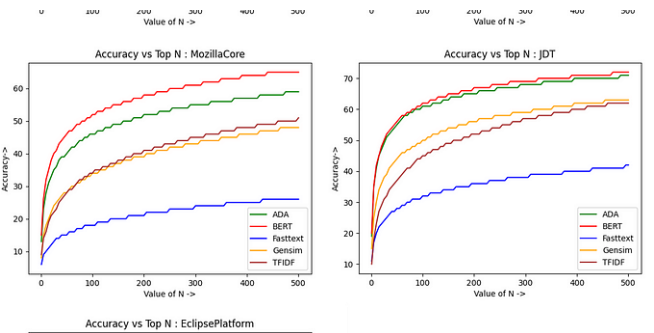
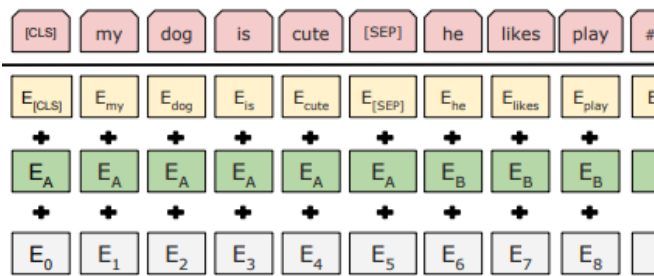
Kalman filter is an algorithm that combines information about the state of a system usin...

3 min read · Jun 11, 2018

 140  1  

See all from Pavel Surmenok

Recommended from Medium





Zain ul Abideen

## A Comparative Analysis of LLMs like BERT, BART, and T5

Exploring Language Models

6 min read · Jun 26



20



1



...



Avinash Patil

## Embeddings: BERT better than ChatGPT4?

In this study, we compared the effectiveness of semantic textual similarity methods for...

4 min read · Sep 19



3

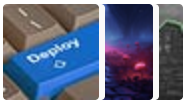


1



...

### Lists



#### Predictive Modeling w/ Python

20 stories · 452 saves



#### Practical Guides to Machine Learning

10 stories · 519 saves



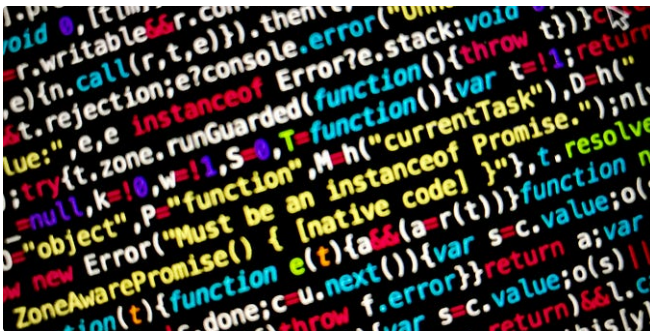
#### Natural Language Processing

669 stories · 283 saves



#### The New Chatbots: ChatGPT, Bard, and Beyond

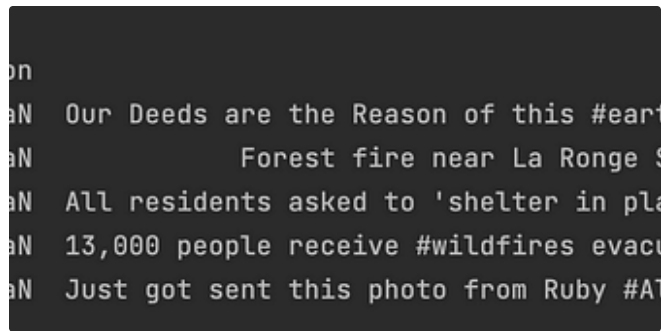
13 stories · 133 saves



sandeep pandey

## Text Classification using Custom Data and PyTorch

Text classification is a fundamental natural language processing (NLP) task that involve...



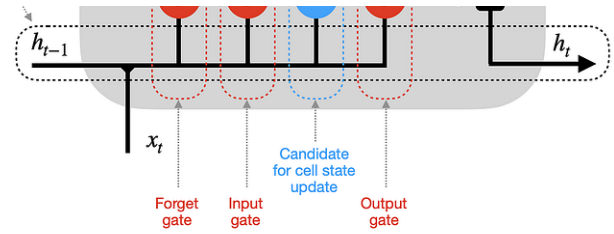
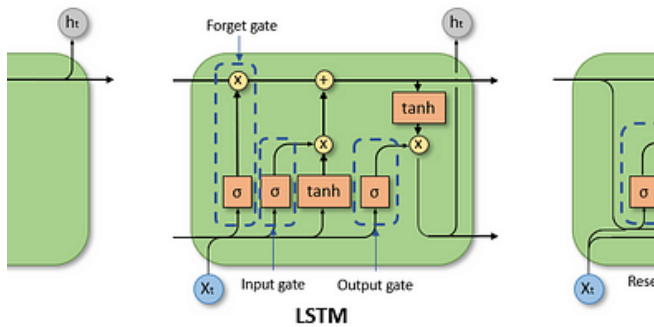
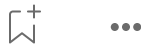
Hatice Şeyma Koç

## Fasttext & Doc2Vec for Text Classification

Hello everyone,

5 min read · May 18

5 min read · Jul 7



$h_{t-1}$  - hidden state at previous timestep t-1 (short-term memory)  
 $c_{t-1}$  - cell state at previous timestep t-1 (long-term memory)

Jonte Dancker in Towards Data Science

Muhammad A in AI Mind

## A Brief Introduction to Recurrent Neural Networks

An introduction to RNN, LSTM, and GRU and their implementation

12 min read · Dec 26, 2022



## Next word prediction using LSTM/RNN

In this article, you will learn an end-to-end implementation of next-word prediction or...

7 min read · Jul 25

See more recommendations