

USING DEEP LEARNING MODELS IN TEXT CLASSIFICATION

Urakov Doston Zarif oglu

Samarkand Branch of Tashkent University of Information Technologies

Named After Muhammad al-Khwarizmi, Samarkand, Uzbekistan

Abstract

Text classification is a key task in several areas of natural language processing, including semantic word classification, sentiment analysis, question answering, or dialog management. In this paper, we consider three main types of deep learning architectures for text classification tasks: deep belief neural (DBN), convolutional neural network (CNN), and recurrent neural network (RNN). DBN has excellent learning ability for feature extraction and is suitable for general purposes. CNN is useful for determining the location of different features when RNN is modeled in a long-term relationship sequence.

Keywords: Deep belief neural(DBN), convolutional neural network(CNN), recurrent neural network(RNN), GRU, MLP.

Introduction

Text classification is one of the important tasks of machine learning and is widely used in several areas of natural language processing. The goal is to develop appropriate algorithms that allow computers to extract features and classify texts automatically. Deep learning, which was developed from artificial neural networks and has now become a popular field of machine learning that attempts to obtain a higher level of abstraction from data using a hierarchical mechanism. It is an emerging technique and is widely used in several areas such as pattern recognition, semantic analysis, speech recognition, computer vision, and natural language processing. DNN has become a very interesting research area in the last few years. It is important to use neurons to develop real-valued activations by adjusting the weights to build traditional neural networks.

Deep learning models typically adopt a hierarchical architecture to combine layers. The output of a lower layer can be considered as the input of a higher layer through simple linear or nonlinear connections. These models can transform low-level word-wise features of the data into high-level abstract feature-wise features. Based on features, deep learning methods are less computationally intensive than machine learning approaches in feature representation. The performance of existing machine learning approaches is usually dependent on user knowledge, but deep learning approaches are dependent on datasets. Therefore, we have found that deep learning models minimize the need for users and rapidly improve performance in computer vision domains.

CNNs are very sophisticated and widely used deep architectures that perform very well in areas with large datasets and have achieved early success in number classification tasks. While a

DBN is a generative probabilistic architecture, it consists of a single visible layer and multiple hidden layers of a deep architecture. Over the past five years, RNNs have achieved good results in various machine learning applications and are an extension of traditional neural networks that can handle variable-length sequential inputs.

Two successful types of RNN models have been developed to reduce the input gaps of the approximation mechanisms: long short-term memory (LSTM) introduced in 1997 and Gated recurrent unit (GRU) introduced in 2014. In other words, we can say that CNNs are hierarchical, DBNs are general-purpose, and RNNs are sequential architectures.

Recently, deep learning models have achieved remarkable results in various areas of NLP, such as text classification. Now the question is, which one should we choose among them for text classification tasks. Based on the previous results and the description of these models, we choose CNN against hierarchical model (CNN) and general purpose model (DBN) and sequential model (RNN) for difficult NLP classification tasks, such as text classification. Sentiment analysis classification, which is a sentiment analysis task usually defined by a key phrase; Recently, CNN has shown high performance with LSTM engine on classification and sequential language modeling tasks.

On the other hand, DBN model selection performs similar tasks to NLP such as text classification, with the ability to learn multiplex features with hidden layers and derive more complex functions to represent data. Each hidden layer unit learns the statistical relationship between the units in the lower layer, while the upper layer representation tends to be more complex. RNN model selection is a sequential modeling task such as language modeling, and is a flexible sequential modeling that requires context dependencies.

For example, an RNN model performs very well on many sequential tasks such as NLP, text classification, web classification, spam filtering, document-level sentiment classification, and any audio dataset.

This study systematically compared CNN, DBN, and two types of available RNNs, namely LSTM and GRU, on classification tasks. In this study, we identified two main findings from our research experience: CNN and RNN provide additional information for text classification tasks. Which architecture performs better depends on how important it is to understand the semantics of the entire sequence.

However, based on our research experience, we have found that some of the shortcomings of standard RNNs are gradient decay and explosion problems. This makes training RNNs difficult in two ways: (i) it cannot handle very long sequences when using a hyperbolic tan activation function; (ii) it uses a rectified linear unit (ReLU) as the activation function. Types of RNNs such as LSTM and GRU are able to overcome this problem. The learning rate changes the performance relatively smoothly, while the batch size and hidden layer size represent large differences in the results.

Deep neural networks (DNNs) are the basis of deep learning and use complex mathematical techniques to train various models. They contain many hidden layers, which is why they are sometimes called multilayer perceptrons (MLPs).



In machine learning, a deep neural network (DBN) is a generative graph model, or alternatively, a class of deep neural network, that consists of multiple layers of hidden variables ("hidden units"), with connections between the layers, but not between the units within each layer.

When trained on an unsupervised set of examples, a DBN can learn to reconstruct its inputs probabilistically. The layers then act as feature detectors. After this learning phase, the DBN can be further trained under supervision to perform classification.

DBNs can be viewed as a composition of simple, unsupervised networks, such as restricted Boltzmann machines (RBMs) or autoencoders, in which the hidden layer of each subnetwork serves as the visible layer for the next. An RBM is an undirected, generative energy-based model with a "visible" input layer and a hidden layer, and between layers, but not within layers. This composition leads to a fast, layer-by-layer unsupervised learning process, where contrastive divergence is applied to each subnetwork in turn, starting with the "lowest" pair of layers.

The observation that DBNs can be trained rapidly, one layer at a time, led to one of the first effective deep learning algorithms. Overall, there are many attractive applications and uses of DBNs in real-life applications and scenarios (e.g., electroencephalography, drug discovery).

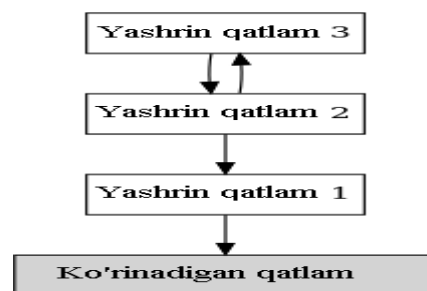


Figure 1. Schematic representation of a deep belief network. The arrows represent the directed connections in the graphical model that the network shows .

Recurrent neural network

The RNN proposal is based on the intuitive notion that "Human cognition is based on experience and memory." In an RNN, there are sequential relationships within a sequence, and neighboring elements are related to each other. The network combines the input features of the previous and current time intervals to predict the output of the next time step. In particular, the hidden layer nodes of the RNN are connected to each other. The input of the hidden layer consists of the input layer and the output of the previous hidden layer. The structure of the RNN is shown in Figure 2. Given an input sequence $D=(w_1, w_2, \dots, w_{|D|})$, where $w_t (t \leq |D|)$ denotes the input token at time step t , the RNN can output the winning image of D , which is $h=(h_1, h_2, \dots, h_{|x|})$.

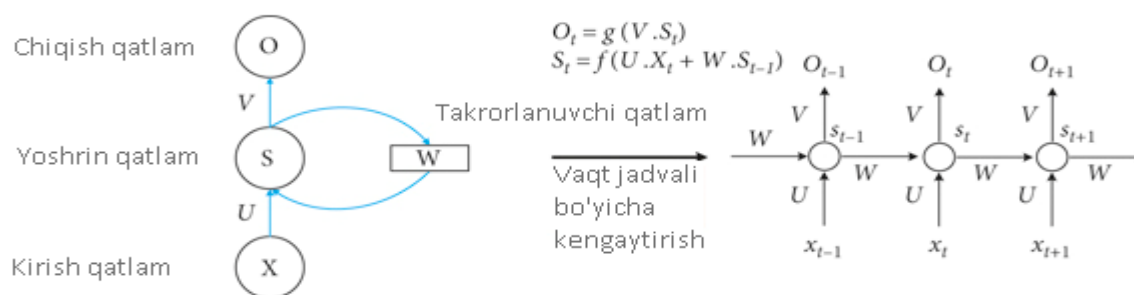


Figure 2. RNN timeline expansion diagram.

RNN is very effective in processing sequential data. It can collect temporal and semantic information in the data. Therefore, DL models based on RNN have made progress in solving some difficult problems in NLP, such as information retrieval, recommendation system, machine translation, text generalization, and time analysis. However, when the sequence is too long, the RNN gradient burst will start to appear and disappear. To overcome this problem, P. Cheng and L. Dong designed a new neural network called Long Short Term Memory (LSTM). Unlike RNN, LSTM selectively stores information through input, forget, and output gates, which solves the problem of long-term correlation. Based on LSTM, K. Cho et al. further simplified the network structure. They used the update gate to replace the input and forget gates, and proposed a new Gate Recurrent Unit (GRU). In addition, bidirectional RNNs are proposed by multiplying the information flow from back to front, which are designated as Bi-RNN, Bi-LSTM, and Bi-GRU.

Convolutional neural network. CNN is a deep feedforward neural network composed of many convolution operations. The neurons in CNN are arranged in three dimensions, namely depth, width, and height. Neurons in different layers are no longer fully connected, but are connected between small areas. The most notable features of CNN are equivariant views, sparse interactions, and parameter sharing, which provide a way to handle inputs of different sizes to neural network models. The basic CNN consists of three structures: convolution, activation, and pooling. CNN uses a convolution kernel to extract features from a data object and uses maximum pooling over the features obtained in an interval that can obtain features of different levels from simple to complex. The convolution filter and pooling operations not only identify important features of the input matrix, but also greatly simplify the complexity and reduce the parameters. One convolutional block consists of a sequence of M convolutional layers and b convolutional layers. In CNN, N convolutional blocks can be placed in series, and K is connected at the end of the fully connected layer. Typically, M is set to 2-5, b is 0 or 1, N is set to 1-100 or more, and K is set to 0-2. The typical CNN structure commonly used is shown in Figure 3. As the core technology of CV, CNN plays an important role in the imaging field. Classical CNNs include Lenet, Alexnet, GoogleNet, VGG, etc. In recent years, CNN has been expanding in the fields of face recognition, machine translation, motion analysis, and NLP, and has achieved good results.

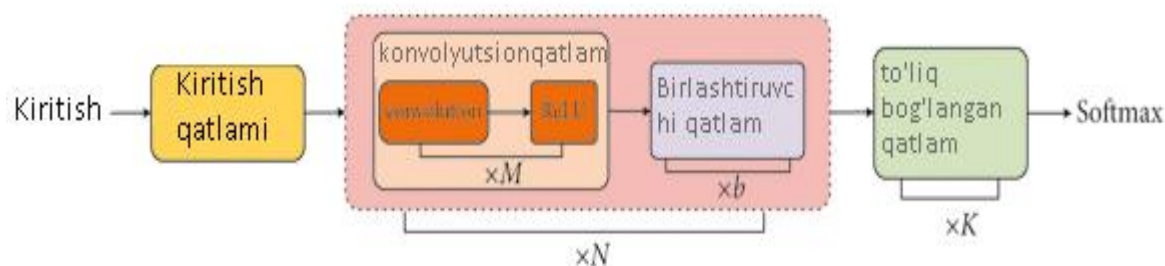


Figure 3. Framework of convolutional neural networks.

References

1. Zhang, M., Zhou, G., Yu, W., Huang, N., & Liu, W. (2022). A Comprehensive Survey of Abstractive Text Summarization Based on Deep Learning. Computational Intelligence and Neuroscience , 2022 .
2. Zulqarnain, M., Ghazali, R., Hassim, YMM, & Rehan, M. (2020). A comparative review on deep learning models for text classification. Indonesia. J. Electr. Most. Comput. Sci , 19 (1), 325-335.
3. A. Khan, MA Gul, M. Zareei et al., "Movie Review Summarization Using Supervised Learning and Graph-Based Ranking Algorithm," Computational intelligence and neuroscience , vol. 2020, Article ID 7526580, 2020.
4. WS El-Kassas, CR Salama, AA Rafea, and HK Mohamed, "Automatic text summarization: a comprehensive survey," Expert Systems with Applications , vol. 165, Article ID 113679, 2021.
5. Rakhimov, R., Primova, H., & Ruziyeva, Z. (2021, November). Methods of recognizing texts in different images. In 2021 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-4). IEEE.
6. Primova, HA, Sotvoldiyev, DM, Rakhimov, RT, & Bobabekova, X. (2020). Computing fuzzy integral of the basis of fuzzy measurement. In Journal of Physics: Conference Series (Vol. 1441, No. 1, p. 012161). IOP Publishing.