المجلة السعودية للدراسات التربوية والنفسية
المجلد (٣) العدد (١)
الإصدار السادس
(٢٠١-٢١١) ٢٠٢٢

Middle East Institution of scientific Publishing

Natural Language Processing problems that can be solved using neural networks

Hamad mubarak Musameh AL-Ajmi
ماجستير هندسة الكترونيات، وزارة الإعلام **دولة الكويت**
Email: Hamad04@live.com
Published on: 13 May 2022

Abstract

neural networks have had its limitations and shortcomings in certain areas. Its representations must be formal which cannot be said about humans who have common sense and use it. Rule-based computers are limited in their abilities to accommodate inaccurate information and logic machines are not good at dealing with images and making analogies. In this study, we described Natural Language Processing problems that can be solved using neural networks. As we showed, neural networks have many applications such as text classification, information extraction, semantic parsing, question answering, paraphrase detection, language generation, multi-document summarization, machine translation, and speech and character recognition. In many cases, neural networks methods outperform other methods.

Keywords: Artificial intelligence, Natural Language Processing, neural networks

* Introduction

Attempts were first made at describing the anatomy of the brain and how its different parts work. These attempts and others eventually led computer scientists, after many phases of development, to try and benefit from them in developing computer systems that function in a way similar to the brain, hence hoping to be able to make these systems perform or simulate tasks unthought-of in the computer world previously. Part of these efforts led to artificial intelligence, another part led to artificial neural networks (ANN)

ANN have been defined in many ways. In relation to their biological origins, they are said to be crude electronic models based on the neural structure of the brain, or simple mathematical constructs that loosely model biological nervous systems, or even highly simplistic abstractions of the human brain. From a more technical perspective, they are defined as parallel computing devices consisting of many interconnected simple processors, or equivalently as an interconnected assembly of simple processing elements. Another way to describe them is as an information processing paradigm composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. And finally, in relation to their specific function they have been termed as data classifiers, associators, predictors, conceptualizers and filters.

Artificial neural networks (Bandy, 1997; Haykin, 1999) are information processing structures providing the (often unknown) connection between input and output data (Honkela, Duch, & Girolami, 2011) by artificially simulating the physiological structure and functioning of human brain structures.

Artificial neural networks are composed of elementary computational units called neurons (McCulloch & Pitts, 1943) combined according to different architectures. For example, they can be arranged in layers (multi-layer network), or they may have a connection topology.

The neural network is not programmed directly but it is explicitly trained through a learning algorithm for solving a given task, a process that leads to "learning through experience". The learning algorithm helps to define the specific configuration of a neural network and, therefore, conditions and determines the ability of the network itself to provide correct answers to specific problems.

Now that we have given a basic idea about ANN, we can talk about the reasons that drove scientists towards using them. The parallelism described in the definition and shown in the architecture is one of the most important reasons; sequential processing has faced the problems of speed and what is called the whole-picture problem. The speed problem can be understood by using the analogy of a single door as opposed to many. There is a limit to the number of people entering a single door at once while you can allow multiples of that number to enter using many doors (parallel processing). Supercomputers might have a solution to the speed problem but parallel processing can take in more factors at one time giving them the ability to look at any problem as a whole-picture and not only small pieces.

*** Definition of Artificial Neural Networks**

Definition of Artificial Neural Networks (ANNs) is made by computer scientists, artificial intelligence experts and mathematicians in various dimensions. Many of the definitions explain ANN by referring to graphics instead of giving well explained mathematical definitions; therefore, misleading weighted graphs (as in minimum cost flow problem networks) fit the definition of ANN. This study aims to give a clear definition that will differentiate ANN and graphical networks by referring to biological neural networks. The proposed definition of ANN is a mathematical definition, from the point of graph theory which defines ANN as a directed graph. Then differences between ANNs and other networks will be explained by examples using proposed definition.

The proposed definition also contains restrictions about connectivity of an ANN, through which the input and the output are mapped. This subject will be explained as the output must be obtained from given input variables. If an SE is not connected to at least one of the EEs, than this means it does not have an effect on output. In a similar way if an EE is not connected to at least one of the SEs, than this means it is not effect

The proposed definition also contains restrictions about connectivity of an ANN, through which the input and the output are mapped. This subject will be explained as the output must be obtained from given input variables. If an SE is not connected to at least one of the EEs,

than this means it does not have an effect on output. In a similar way if an EE is not connected to at least one of the SEs, than this means it is not effect.

The proposed definition also contains restrictions about connectivity of an ANN, through which the input and the output are mapped. This subject will be explained as the output must be obtained from given input variables. If an SE is not connected to at least one of the EEs, than this means it does not have an effect on output. In a similar way if an EE is not connected to at least one of the SEs, than this means it is not effect in which three graphical representations are used for defining a neural network: x with a block diagram to describe the network functionally, x with signal-flow graph to describe signal flow in the network, x with the architectural graph to describe the network layout. Description of ANN in [1] as a directed graph is not complete as long as it excludes the learning process, input and output sets (number of input or output records and number of attributes of inputs and outputs) and parallel structure. An interesting and highly mathematical definition we found in [5]. In [5] ANN is defined from the point of graphs as follows: "A neural network model is defined as a directed graph with the following properties: x A state variable ni is associated with each node i, x A real valued weight wik is associated with each link (ik) between two nodes i and k, x A real valued bias vi is associated with each node i, x A transfer function fi[nk,wik,vi,(i□k)] is defined for each node i, which determines the state of the node as a function composed of its bias, the weights of incoming links, and the states of nodes connected to it.

*** Application of Artificial Neural Networks in organizations**

They have also been used for the estimation of heating-loads of buildings, for the prediction of air flows in a naturally ventilated test room and for the prediction of the energy consumption of a passive solar building. In all such models, a multiple hidden-layer architecture has been used. Errors reported when using these models are well within acceptable limits, which clearly suggests that artificial neural-networks can be used for modelling in other fields of energy production and use. The work of other researchers in the field of energy is also reported. This includes the use of artificial neural-networks in heating, ventilating and air-conditioning systems, solar radiation, modelling and control of power-generation systems, load-forecasting and refrigeration.

Computer technology has been advanced tremendously and the interest has been increased for the potential use of 'Artificial Intelligence (AI)' in medicine and biological research. One of the most interesting and extensively studied branches of AI is the 'Artificial Neural Networks (ANNs)'.

Basically, ANNs are the mathematical algorithms, generated by computers. ANNs learn from standard data and capture the knowledge contained in the data. Trained ANNs approach the functionality of small biological neural cluster in a very fundamental manner. They are the digitized model of biological brain and can detect complex nonlinear relationships between dependent as well as independent variables in a data where human brain may fail to detect. Nowadays, ANNs are widely used for medical applications in various disciplines of medicine especially in cardiology. ANNs have been extensively applied in diagnosis, electronic signal analysis, medical image analysis and radiology. ANNs have been used by many authors for modeling in medicine and clinical research. Applications of ANNs are increasing in pharmacoepidemiology and medical data mining. In this paper, authors have summarized various applications of ANNs in medical science.

Since artificial neural networks allow modeling of nonlinear processes, they have turned into a very popular and useful tool for solving many problems such as classification, clustering, regression, pattern recognition, dimension reduction, structured prediction, machine translation, anomaly detection, decision making, visualization, computer vision, and others. This wide range of abilities makes it possible to use artificial neural networks in many areas. In this article, we discuss applications of artificial neural networks in Natural Language Processing tasks (NLP).

NLP includes a wide set of syntax, semantics, discourse, and speech tasks. We will describe prime tasks in which neural networks demonstrated state-of-the-art performance.

*** Text Classification and Categorization**

Text classification is an essential part in many applications, such as web searching, information filtering, language identification, readability assessment, and sentiment analysis. Neural networks are actively used for these tasks.

In Convolutional Neural Networks for Sentence Classification by Yoon Kim, a series of experiments with Convolutional Neural Networks (CNN) built on top of word2vec was presented. The suggested model was tested against several benchmarks. In Movie Reviews (MR) and Customer Reviews (CR), the task was to detect positive/negative sentiment. In Stanford Sentiment Treebank (SST-1), there were already more classes to predict: very positive, positive, neutral, negative, very negative. In Subjectivity data set (Subj), sentences were classified into two types, subjective or objective. In TREC the goal was to classify a question into six question types (whether the question is about person, location, numeric information, etc.) The results of numerous tests described in the paper show that after little tuning of hyperparameters the model performs excellent suggesting that the pre-trained vectors are universal feature extractors and can be utilized for various classification tasks [1].

The article Text Understanding from Scratch by Xiang Zhang and Yann LeCun shows that it's possible to apply deep learning to text understanding from character-level inputs all the way up to abstract text concepts with help of temporal Convolutional Networks (ConvNets) (CNN). Here, the authors assert that ConvNets can achieve excellent performance without the knowledge of words, phrases, sentences and any other syntactic or semantic structures with regards to a human language [2]. To prove their assertion several experiments were conducted. The model was tested on the DBpedia ontology classification data set with 14 classes (company, educational institution, artist, athlete, office holder, mean of transportation, building, natural place, village, animal, plant, album, film, written work). The results indicate both good training (99.96%) and testing (98.40 %) accuracy, with some improvement from thesaurus augmentation. In addition, the sentiment analysis test was performed on the Amazon Review data set. In this study, the researchers constructed a sentiment polarity data set with two negative and two positive labels. The result is 97.57% training accuracy and 95.07% testing accuracy. The model was also tested on Yahoo! Answers Comprehensive Questions and Answers data set with 10 classes (Society & Culture, Science & Mathematics, Health, Education & Reference, Computers & Internet, Sports, Business & Finance, Entertainment & Music, Family & Relationships, Politics & Government) and on AG's corpus where the task was a news categorization into four categories (World, Sports, Business, Sci/Tech.). Obtained results confirm that to achieve good text understanding ConvNets require a large corpus in order to learn from scratch.

Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao introduced recurrent convolutional neural networks for text classification without human-designed features in their document Recurrent Convolutional Neural Networks for Text Classification [3]. The team tested their model using four data sets: 20Newsgroup (with four categories such as computers, politics, recreation, and religion), Fudan Set (a Chinese document classification set that consists of 20 classes, including art, education, and energy), ACL Anthology Network (with five languages: English, Japanese, German, Chinese, and French), and Sentiment Treebank (with Very Negative, Negative, Neutral, Positive, and Very Positive labels). After testing, the model was compared to existing text classification methods like Bag of Words, Bigrams + LR, SVM, LDA, Tree Kernels, RecursiveNN, and CNN. It turned out that neural network approaches outperform traditional methods for all four data sets, and the proposed model outperforms CNN and RecursiveNN.

*** Named Entity Recognition (NER)**

The main task of named entity recognition (NER) is to classify named entities, such as Guido van Rossum, Microsoft, London, etc., into predefined categories like persons, organizations, locations, time, dates, and so on. Many NER systems were already created, and the best of them use neural networks.

In the paper, Neural Architectures for Named Entity Recognition, two models for NER were proposed. The models require character-based word representations learned from the supervised corpus and unsupervised word representations learned from unannotated corpora [4]. Numerous tests were carried on using different data sets like CoNLL-2002 and CoNLL-2003 in English, Dutch, German, and Spanish languages. The team concluded that without a requirement of any language-specific knowledge or resources, such as gazetteers, their models show state-of-the-art performance in NER.

*** Part-of-Speech Tagging**

Part-of-speech (POS) tagging has many applications including parsing, text-to-speech conversion, information extraction, and so on. In the work, Part-of-Speech Tagging with Bidirectional Long Short-Term Memory Recurrent Neural Network a recurrent neural network with word embedding for part-of-speech (POS) tagging task is presented [5]. The model was tested on the Wall Street Journal data from Penn Treebank III data set and achieved a performance of 97.40% tagging accuracy.

*** Semantic Parsing and Question Answering**

Question Answering systems automatically answer different types of questions asked in natural languages including definition questions, biographical questions, multilingual questions, and so on. Neural networks usage makes it possible to develop high performing question answering systems.

In Semantic Parsing via Staged Query Graph Generation Question Answering with Knowledge Base Wen-tau Yih, Ming-Wei Chang, Xiaodong He, and Jianfeng Gao described the developed semantic parsing framework for question answering using a knowledge base. Authors say their method uses the knowledge base at an early stage to prune the search space and thus simplifies the semantic matching problem [6]. It also applies an advanced entity linking system and a deep convolutional neural network model that matches questions and predicate sequences. The model was tested on WebQuestions data set, and it outperforms previous methods substantially.

*** Paraphrase Detection**

Paraphrase detection determines whether two sentences have the same meaning. This task is especially important for question answering systems since there are many ways to ask the same question.

Detecting Semantically Equivalent Questions in Online User Forums suggests a method for identifying semantically equivalent questions based on a convolutional neural network. The experiments are performed using the Ask Ubuntu Community Questions and Answers (Q&A) site and Meta Stack Exchange data. It was shown that the proposed CNN model achieves high accuracy especially when the words embedded are pre-trained on in-domain data. The authors compared their model's performance with Support Vector Machines and a duplicate detection approach. They demonstrated that their CNN model outperforms the baselines by a large margin [7].

In the study, Paraphrase Detection Using Recursive Autoencoder, a novel recursive autoencoder architecture is presented. It learns phrasal representations using recursive neural networks. These representations are vectors in an n-dimensional semantic space where phrases with similar meanings are close to each other [8]. For evaluating the system, the Microsoft Research Paraphrase Corpus and English Gigaword Corpus were used. The model was compared to three baselines, and it outperforms them all.

*** Language Generation and Multi-document Summarization**

Natural language generation has many applications such as automated writing of reports, generating texts based on analysis of retail sales data, summarizing electronic medical records, producing textual weather forecasts from weather data, and even producing jokes.

In a recent paper, Natural Language Generation, Paraphrasing and Summarization of User Reviews with Recurrent Neural Networks, researchers describe a recurrent neural network (RNN) model capable of generating novel sentences and document summaries. The paper described and evaluated a database of 820,000 consumer reviews in the Russian language. The design of the network permits users control of the meaning of generated sentences. By choosing sentence-level features vector, it is possible to instruct the network; for example, "Say something good about a screen and sound quality in about ten words" [9]. The ability of language generation allows production of abstractive summaries of multiple user reviews that often have reasonable quality. Usually, the summary report makes it possible for users to quickly obtain the information contained in a large cluster of documents.

*** Machine Translation**

Machine translation software is used around the world despite its limitations. In some domains, the quality of translation is not good. To improve the results researchers try different techniques and models, including the neural network approach. The purpose of Neural-based Machine Translation for Medical Text Domain study is to inspect the effects of different training methods on a Polish-English machine translation system used for medical data. To train neural and statistical network-based translation systems The European Medicines Agency parallel text corpus was used. It was demonstrated that a neural network requires fewer resources for training and maintenance. In addition, a neural network often substituted words with other words occurring in a similar context [10].

*** Speech Recognition**

Speech recognition has many applications, such as home automation, mobile telephony, virtual assistance, hands-free computing, video games, and so on. Neutral networks are widely used in this area.

In Convolutional Neural Networks for Speech Recognition, scientists explain how to apply CNNs to speech recognition in a novel way, such that the CNN's structure directly accommodates some types of speech variability like varying speaking rate [11]. TIMIT phone recognition and a large-vocabulary voice search tasks were used.

*** Character Recognition**

Character Recognition systems also have numerous applications like receipt character recognition, invoice character recognition, check character recognition, legal billing document character recognition, and so on. The article Character Recognition Using Neural Network presents a method for the recognition of handwritten characters with 85% accuracy [12].

*** Spell Checking**

Most text editors let users check if their text contains spelling mistakes. Neural networks are now incorporated into many spell-checking tools.

In Personalized Spell Checking using Neural Networks a new system for detecting misspelled words was proposed. This system is trained on observations of the specific corrections that a typist makes [13]. It outwits many of the shortcomings that traditional spell-checking methods have.

*** References**

Bandy, H. (1997). Developing a Neural Network System. AAII Working Pap.

Haykin, S. (1999). Neural Networks: a comprehensive foundation (2nd ed.). Upper Saddle River, NJ, USA: Prentice Hall PTR.

Honkela, T., Odzis Aw Duch, W., & Girolami, M. (Eds.). (2011). Artificial Neural Networks and Machine Learning: ICANN 2011, part 1-2. Springer.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics, 5, 115–133. doi:10.1007/ BF02478259

S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall, New Jersey, 1999.

2. R. Eberhart and Y. Shi, Computational Intelligence, Morgan Kaufmann 2007. 3. K. Gurney, An Introduction to Neural Networks, CRC Press, 2003.

4. R. Rojas, Neural Networks: A Systematic Introduction, Springer-Verlag Berlin Heidelberg, 1996. 5. B. Muller, J. Reinhardt, M.T. Strickland, Neural Networks an Introduction, Springer-Verlag Berlin Heidelberg, 1995