ECRONICON OPEN ACCESS

EC MICROBIOLOGY
Review Article

Deep Learning for Medical Named Entity Recognition

Suny Raja[1] and Anitha S Pillai[2]*

[1]*Department, Bachelor of Computer Applications, KMM College of Arts and Science, Thrikakkara, Ernakulam, India*
[2]*Professor, School of Computing Sciences, Hindustan Institute of Technology and Science, Chennai, India*

***Corresponding Author**: Anitha S Pillai, Professor, School of Computing Sciences, Hindustan Institute of Technology and Science, Chennai, India.

Received: July 06, 2017; **Published:** August 04, 2017

Abstract

Deep Learning is a novel branch of Machine Learning study, which has been pioneered with the intention of pushing Machine Learning technology nearer to one of its unique aim: Artificial Intelligence. Named Entity Recognition (NER) is a critical and essential step in many biomedical information mining jobs. But conventional NER systems are mostly dependant on composite hand written features which are specific to a particular area. This paper, presents the different deep learning methods used by researchers in the medical domain.

Keywords: *Deep Learning; Machine Learning; Convolutional Neural Network (CNN); Named Entity Recognition (NER); Deep Convolutional Neural Network (DCNN)*

Introduction

In Artificial Intelligence, to have the efficiency and strength of a human brain artificially is the core challenge faced by the researchers for decades. This challenge is made possible through machine learning where its concepts are based on the functions of a human brain in which multitude of sensory data received every second of the day are captured and analyzing the critical aspects of this data in a way that allows for its future use in a concised manner.

Deep Learning (DL) is a division of machine learning which is based on a set of algorithms learning multiple levels of representation and abstraction that help to make sense of data such as images, sound and text. Machine Learning helps to extract data features for solving and predicting many tasks like named entity recognition (NER), forecasting, detecting anomalies, classifying data, ranking the sites, decision making etc.

Currently, neural networks and deep learning give the most excellent solutions to the challenges in image recognition, speech recognition, and natural language processing.

The world is now in the period of big data. To deal with enormous medical words usefully and well-organized has become a vital problem for the people who are closely associated with medical data. Named Entity Recognition (NER) is a very important task in the field of natural language processing. It is the ability to identify names of person, place, organization, diseases etc from a document.

Machine learning is an area of science for getting systems to mimic without being plainly programmed. The explosive developments of machine learning are self-driving cars, practical speech recognition, effective web search, and an infinitely enhanced understanding of the human genome. Humans use machine learning technology on a daily basis without knowing it. Scientists assume that this is most excellent way to make advancement towards human-level AI. Using some Deep Learning techniques, doctors are getting trained to make quick and fast decision thereby making a shift in biomedical field.

Description about the paper

The paper gives an aerial view about the information extracted from different research papers which are based on Named Entity Recognition (NER) using Deep Learning concepts in medical field. The paper is divided into five sections. In the first section, the paper gives a brief introduction followed by the description of the paper in second section, literature review in the third section, a conclusion in the fourth section followed by references in the fifth section.

Deep Learning in Medical domain

The machine learning techniques used to analyze the medical data provide effective solution for handling vast amount of patient data in the near future is becoming a reality. They are used to identify, classify and thereby understand the pattern in which the medical data are stored.

To understand and identify different types of word embeddings from patient records using deep learning tools are a major achievement in medical field. For acquiring the right entity attribute, appropriate method to be used in deep learning. B Zhong., *et al*. [1] presents a novel method which is entity attribute extraction using the power of deep learning method. At first, conditional random field (CRF) model was used to solve named entities problem and then extracted attributes were identified by deep belief network (DBN) model. Using this method, entity attributes were mined precisely with very less manual involvement when evaluated with conventional methods.

In modern electronic medical records (EMR), much of the clinically important data are not provided in structured data fields, but are encoded in clinician generated narrative text. The research papers give an outlook of NLP systems on the basis of a unified architectural view. NLP system architecture consists of two main components which are the background knowledge that includes biomedical knowledge resources and a framework that integrates NLP tools to process text. Xinbo Lv., *et al*. [2] explains the importance of relation extraction which is named entity recognition (NER) on clinical documents (EHR). For this relation classification, conditional random field (CRF) model is used and then extracted features were used to the best by a deep learning model. The results outcome showed the efficiency of the proposed model when compared with the baseline model.

In the same way of EMR, Biomedical information extractions (BioIE) intend to mechanically release structured semantics out of unstructured biomedical text data. F Liu., *et al*. [3] provide an extensive knowledge related to biomedical information extraction (BioIE) and the exceptional role of deep learning (DNN) to learn internal distributed representations of data, big data processing and in named entity recognition (NER).

In the current years, mining corresponding data from biomedical and clinical texts such as articles related to research, patients' summaries after they leave the hospital, or electronic health records (EHR) have been in the limelight of various research fields and challenges. Relation extraction is a method of identifying and segregating the semantic relation among entities in a recommended section of texts. S K Sahu., *et al*. [4] focus their paper on relation extractions from patients' discharge summaries by proposing a new structure based on convolutional neural network (CNN) for automated feature learning. Their results signify that CNN is a superior model when compared to the SVM model for extracting the relations from clinical data without manual expertise on precision factors and improved performance.

Without manual intervention, recognising a disease name in biomedical and doctors narrative texts are very crucial for the expansion of powerful and refined NLP systems such as information extraction, question answering, text summarization etc. In the paper, S K Sahu., *et al*. [5] propose a recurrent neural network(RNN) model for disease name recognition to reduce the word dependencies in a sentence and along with exploiting the power of convolutional neural network(CNN) for achieving enhanced performance on classification job of disease labels.

To understand the interactions of protein to protein in biomedical field, Z Zhao., *et al*. [6] proposed a deep learning-based method, deep neural network (DNN) for extracting protein-protein interactions (PPI) information from biomedical documents. This method will automatically understand composite and conceptual features at higher layers illustrations with much increased performance than neural network (NN). This identification is very important in biomedical field.

Biomedical named entity recognition (Bio-NER) intends to discover entities in biomedical texts. Y Zhang., *et al*. [7] proposed a machine-learning-based method to overcome the difficulty of identifying drug information from the biomedical texts. Using the dictionary look-up and conditional random field (CRF) model with lexical features, drug names are used to identify which achieves an F-score of 92.54% on the test data set.

Understanding the importance of disease classification by analyzing social media through deep learning concepts are very much prevalent in the society today. N Limsopatham., *et al*. [8] give an understanding about learning and exploiting the semantic comparison between the messages in social media and medical field with the power of deep neural networks, i.e convolutional neural network (CNN) and recurrent neural network (RNN).The experimental results proved that the proposed approach considerably surpassed already available methods with a performance enhancement by 44%.

Even though, the named entity recognition is based in the clinical narrative texts especially in the common language ,i.e; English, Y Peng and T Jiang., *et al*. [9] proved machine-learning methods outperform in named entity recognition (NER) based on Conditional Random Field (CRF) and assertion determination using Support Vector Machine (SVM) classifier for the Chinese electronic medical records (EMRs).Same in the case of Wu., *et al*. [10] proposed a deep learning based NLP systems which can automatically understand important attributes illustrations from unlabeled corpora through unsupervised learning in Chinese medical manuscripts.

Precision in the medical data is very important to identify the risk factors (NER) in high risk patients. J Urbain [11] presented a new distributional semantic model to confine the context thereby increasing the accuracy to much higher precision than expected for picking out heart disease risk factor events in diabetic patients over time. This accuracy level will surpass many of the ignored or left out readings from the patients EHRs. Ford E., *et al*. [12] presents in their paper about the review of patients' accurate information which are texts obtained from EMRs using case-detection algorithms.

To get an insight of the paper, a comparison is made with the deep learning methods used here for named entity recognition in tabular form in table 1 above.

Papers Referred	Method Used	Accuracy/F-Score (%)*	Aim
[1]	Deep Belief Network	65	Entity Accuracy Extraction
[2]	Conditional Random Field	80	Clinical Relation Extraction
[3]	Deep Neural Network	50 (reduced annotation time)	BioMedical Information Extraction
[4]	Convolutional Neural Network	76.34	Clinical Relation Extraction
[5]	Convolutional Neural Network	79.13	Disease Name Extraction
[6]	Deep Neural Network	89.05	Protein-Protein Interaction Extraction
[7]	Conditional Random Field	92.54	Biomedical NER
[8]	Convolutional Neural Network	44	Disease Classification from Social Media
[9]	Conditional Random Field	89.07	NER in Clinical Texts
[10]	Deep Neural Network	92.8	Chinese Clinical Entities Extraction
[11]	Deep Neural Network	89	Chinese Clinical Entities Extraction
[12]	Case Detection Algorithm	66	EMR Information Extraction

Table 1: *Comparison of Methods.*

**Results shown here is according to either F-Score or Accuracy mentioned in the corresponding Paper on comparison.*

Conclusion

With the technology entering and residing in the world of machine learning and artificial intelligence, deep learning is moving into the limelight of international, biomedical, bioinformatics, academic and commercial domains. In each and every domain, there are major advances in machine learning algorithms.

In our paper, we have done an extensive comparative study especially in biomedical field introducing deep learning and gave an insight to the named entity recognition and features of frequently used deep learning architectures. That is, everything that humans can intelligently sense and that our technology can digitalize.

Using deep learning technology, we are basically giving society the ability to behave much more intelligently, by accurately interpreting what's happening in the world around us with software. We trust, this review will provide helpful ideas and will be a stepping stone for aspiring researchers to know more about deep learning in their area of research.

Bibliography

1. Bei Zhong., *et al*. "Extracting Attributes of Named Entity from Unstructured Text with Deep Belief Network". *International Journal of Database Theory and Application* 9.5 (2016): 187-196.

2. Xinbo Lv., *et al*. "Clinical Relation Extraction with Deep Learning". *International Journal of Hybrid Information Technology* 9.7 (2016): 237-248.

3. Feifan Liu., *et al*. "Learning for Biomedical Information Extraction: Methodological Review of Recent Advances". Published in Cornell University Library (2016).

4. Sunil Kumar Sahu., *et al*. "Relation extraction from clinical texts using domain invariant convolutional neural network". Published in Cornell University Library (2016).

5. Sunil Kumar Sahu and Ashish Anand. "Recurrent neural network models for disease name recognition using domain invariant features". Published in Cornell University Library (2016).

6. Zhao Z., *et al*. "A protein- protein interaction extraction approach based on deep neural network". *International Journal of Data Mining and Bioinformatics* 15.2 (2016): 145-164.

7. Linna He., *et al*. "Drug name recognition in biomedical texts: a machine-learning-based method". *Drug Discovery Today* 19.5 (2014): 610-617.

8. Nut Limsopatham and Nigel Collier. "Normalising Medical Concepts in Social Media Texts by Learning Semantic Representation". Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany. Association for Computational Linguistics (2016): 1014-1023.

9. Jianhong Wang., *et al*. "Extracting Clinical entities and their assertions from Chinese Electronic Medical Records Based on Machine Learning". 3rd International Conference on Materials Engineering, Manufacturing Technology and Control (ICMEMTC) (2016).

10. Wu Y., *et al*. "Named Entity Recognition in Chinese Clinical Text Using Deep Neural Network". *Studies in Health Technology and Informatics* 216 (2015): 624-628.

11. J Urbain. "Mining heart disease risk factors in clinical text with named entity recognition and distributional semantic models". *Journal of Biomedical Informatics* 58 (2015): S143-S149.

12. Elizabeth Ford., *et al*. "Extracting information from the text of electronic medical records to improve case detection: a systematic review". *Journal of the American Medical Informatics Association* 23.5 (2016): 1007-1015.

Volume 10 Issue 3 August 2017