International Conference on Smart Computing and Electronic Enterprise. (ICSCEE2018) ©2018 IEEE

Inflectional Review of Deep Learning on Natural Language Processing

SK Ahammad Fahad
Faculty of Computer and Information Technology
Al-Madinah International University
Shah Alam, Malaysia
bl308@lms.mediu.edu.my

Abdulsamad Ebrahim Yahya
Faculty of Computing and Information Technology
Northern Border University
Rafha, KSA
Abdulsamad.qasem@nbu.edu.sa

***Abstract*— In the age of knowledge, Natural Language Processing (NLP) express its demand by a huge range of utilization. Previously NLP was dealing with statically data. Contemporary time NLP is doing considerably with the corpus, lexicon database, pattern reorganization. Considering Deep Learning (DL) method recognize artificial Neural Network (NN) to nonlinear process, NLP tools become increasingly accurate and efficient that begin a debacle. Multi-Layer Neural Network obtaining the importance of the NLP for its capability including standard speed and resolute output. Hierarchical designs of data operate recurring processing layers to learn and with this arrangement of DL methods manage several practices. In this paper, this resumed striving to reach a review of the tools and the necessary methodology to present a clear understanding of the association of NLP and DL for truly understand in the training. Efficiency and execution both are improved in NLP by Part of speech tagging (POST), Morphological Analysis, Named Entity Recognition (NER), Semantic Role Labeling (SRL), Syntactic Parsing, and Coreference resolution. Artificial Neural Networks (ANN), Time Delay Neural Networks (TDNN), Recurrent Neural Network (RNN), Convolution Neural Networks (CNN), and Long-Short-Term-Memory (LSTM) dealings among Dense Vector (DV), Windows Approach (WA), and Multitask learning (MTL) as a characteristic of Deep Learning. After statically methods, when DL communicate the influence of NLP, the individual form of the NLP process and DL rule collaboration was started a fundamental connection.**

Keywords—Deep Learning; Natural language processing; Deep nural Network; Multitask Learning

I. INTRODUCTION

In the age of information, Natural Language Processing (NLP) create its demand by a comprehensive area of application. To present the significant knowledge to non-programmer from computer system Natural Language Processing was deliberate as a working field from 1950. Non-Subject material experts obtain an answer of simple queries by improvement of NLP. Previously NLP was dealing with statically data. Recent year NLP is doing well with the corpus, lexicon database, Neural Network. Since Deep Learning (DL) method allow artificial Neural Network (NN) to nonlinear process, Natural Language processing tools enhance more accurate and valuable that make a revolution. Multi-Layer Neural Network expanding the influence of the natural language processing for its capability with decent acceleration and reliable producing.

In the field of NLP, DL completely takeover Text Classification and Categorization, Named Entity Recognition (NER), Part-of-Speech Tagging (POST), Semantic Parsing and Question Answering, Paraphrase Detection, Language Generation and Multi-document Summarization, Machine Translation, Speech Recognition, Character Recognition, Spell Checking etc. Hierarchical representations of data operate complicated processing layers to learn and have by this imagined Deep Learning methods dominate many realms. Those tools are powered by DL and applied on Natural Language to complete the NLP process to achieve the goal. Text Summarization, Caption Generation, Question Answering, Text Classification, Machine Translation, Language Modeling, Speech Recognition all NLP tools are working with DL go obtain desire accurate result. Several methods are proposed and the system was implemented to NLP by DL and they are doing well.

NLP methods are modified when DL was associated. NLP process like POS, NER, Morphology, Syntactic Parsing, Coreference resolution is discussed in subsequent section. Different type of Neural Network like; ANN, TDNN, RNN, CNN are discussed with the relation and in NLP process was discussed in this paper. Different technique and tools of DL like; LSTM, MTL, DV, CBOW, VL, WA, SRL, Non-Linear Function are discussed with the possible relation with NLP. In this paper, it was trying to make a review of the tools and the basic methodology. After statically methods, when DL take over the control to NLP, a new form of the NLP process and DL process collaboration was presented with the basic relation.

II. NATURAL LANGUAGE PROCESSING

During a document record recognized for accomplishing besides, this is not continued to direct accomplishment. It requires processing background for authentic execution. There has some obligatory progression in contemporary Natural Language Processing (NLP). Deep Learning (DL) is a superior form of Neural Network (NN) and it deals with preprocessed data. Before NN applied in NPL subsequent processing the document. DL also operated among concocted document file. That's why unusual primary actions of the process of textual

documents are extremely valuable. There have six compulsory rounds should ensue to perceive more sustainable and accurate result by implementing DL. Those are; Splitting and Tokenization, Part-of-Speech Tagging, Morphological Analysis, Named Entity Recognition, Syntactic Parsing, and Conference Resolution [7,8].

Splitting and Tokenization segment clean document from outcast tags and absolute splitting document in to token. The input file has any tags for designing the text. NLP consider simply real fresh text for processing. Tags require cleaning before performing a better and correct result in NLP. Tokenization is that the approach of separating a stream of text into words, phrases, symbols, or different principal elements specified as tokens. How to split and character of the token is should describe as per the demand for output. Succeeding cleaning the writing from unwanted content and splitting text to token, NLP proceeding by tokens [3].

Part of speech tagging (POST) or POS tagging is a mechanism of NLP, it performs the extremely significant role on text phrase, syntax, translation, and semantic analysis. Rule-Based tagging perform POS tagging by match everything on a lexical dictionary and match rest of words independently with each part of speech precept. In Stochastic POS tagging is accomplished by applying prospect including a tagged corpus. Both Rule-Based and Stochastic POS tagging are worked blended in Transformation-Based POS tagging [2].

Morphology allots with the connection between the structure and purpose of words. Morphological Analysis is the insignificant semantic systems including a lexical or a logical definition of words. Morphological analyzers and lemmatizes typically demand training data. The latter is practiced on character-aligned combinations of stems and lemmas, where stems are extorted and neophytes stem into lemmas [9].

Named Entity Recognition (NER) is the policy that recognizes the name and number and eliminates from tokens which are allowing for additional processing. Hand-Made NER sharpens on obtaining name and numbers applying the rules which are human-made. Those rules utilizing grammatical syntactic and Orthographic innovations in combination with dictionaries. Machine Learning-based NER method applies a classification analytical pattern to determine and converting description query into a distribution problem. As tike POS tagging there has a Hybrid scheme developed which one is the combination of human-made and machine-made rules [6].

Syntactic Parsing produces comprehensive syntactic analysis and Sentiment analysis including both ingredient and dependence description with a compositional pattern atop trees utilizing deep learning. WordNet® is an immense on-line database of English. Nouns, verbs, adjectives and adverbs section organized within collections of subjective peculiarity synonyms (synsets), individual expressing a particular representing. Synsets section interlinked with hints of conceptual-semantic and lexical associations [6].

Coreference resolution is the responsibility of defining semantic definitions that belong to the corresponding real-world existence in natural language. Coreference consists of two semantic appearances—antecedent and anaphor.

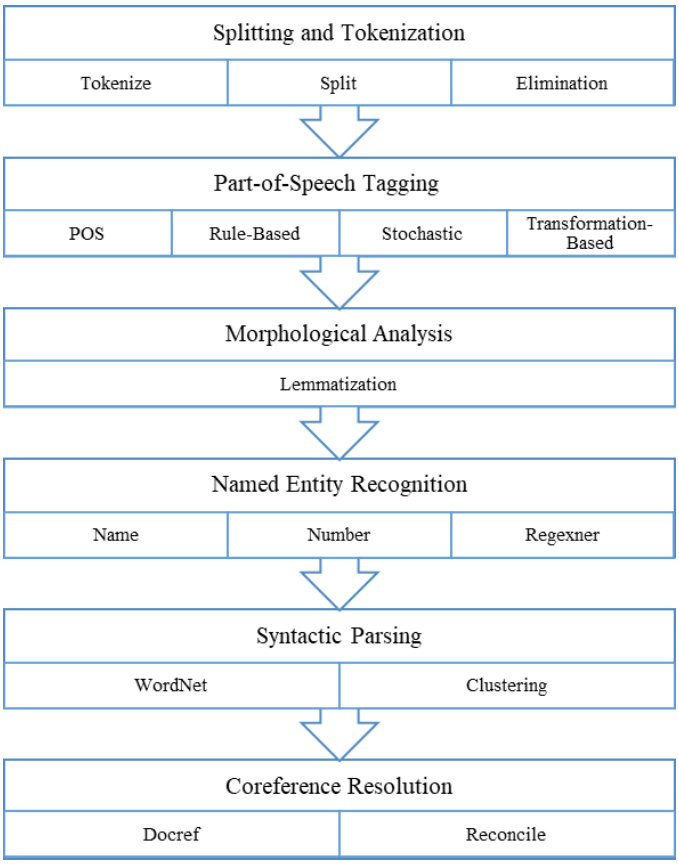

Fig. 1. *Basic steps for Natural Language Processing for Deep Learning Technique*

The anaphor is the appearance whose interpretation (i.e., comparing by both pavement and theoretical real-world article) depends on that of the differential phase. The predecessor is the semantic appearance on which an anaphor depends. Coreference resolution complete specifying disclosure and both pronominal from token and dismissed them [4,5].

III. Deep Learnig Technique on Natural Language Processing

A. Dense Vector

Artificial Neural Networks (ANN) Input textual data to the data vector x in-dimensions and produce the product in out-dimensions. DL distributes with raw words (tokens) and not engineered features, the first layer has to map words into real-valued vectors for processing by consequent layers of the NN. Alternatively, forming a unique dimension for individual and every feature, implant every feature into a D-Dimensional space and describe as a Dense Vector (DV) in the space [1]. If there exists a correlation distribution which can learn, that can be captured by DV. It is the principal influence of DV representation. DV training will cause comparable

characteristics have to experience vectors-information between them [7].

B. Feature Extraction

For deriving a feature from tokens; Distance and Dimensionality Feature, Continuous Bag of words (CBOW), and Variable Length (VL) signify implemented. Distance can measure by deducting the identical tokens word to feature vector. Distance invariably positive, it can deduct whatever from anyone. This measurement data can be used to train NN and make DL more specific. Distance, the weight of token, synset (synonyms) everything is a feature of the embedded token and all of the dimension are property and those are associated with processing speed, and accuracy. To obtain a feature dimension is an important feature of a token. CBOW is a feature based on dense vector feature extraction. NN get trained from inputted token individually with a unique feature. The weight of word executes this feature extends the important. A number of unique words are proffered to vector representation (One hot vector illustration) and it's a significant part of train DL [1]. Averaging embedded feature and weight averaging of an embedded feature is essentially allied to expand weight. DV has the primary tokens within the sequence. But DL inadequate to supervise to variable length. Consider an established measurement Windows Approach (WA) can be a simplistic explication of this limitation. WA strongly collaborate with POS but it not proper to operate with Semantic Role Labeling (SRL) [7].

C. Deep Neural Network

Determination of basis of deep learning Neural Network depends on the impression was operating to supervise to accomplish the purpose. Non-Linear Function, Output Transformation Function, Loss Function are numerously accessible on NLP. NN is approximator and it can approximate any non-zero liner function. Infinity number of positive and negative range and a bound output range is the property of a Non-Linear Function. In Outer Transformation Function the peripheral function of NN use as a transmutation function. To represent a Multi-class distribution, Outer Transformation is pretty much recommended and used extensively. How greatly the network encouraged from the genuine output is was intimate by Loss Function. It depends on the specification, character, and extent of tolerance. Ranking loss, categorical entropy loss, log loss are the current function of Loss Function [11]. Time Delay Neural Networks (TDNN) can be the best alternative when fashioning for long-distance dependencies. TDNN accumulated to local feature in the deeper layer and tiny local (more global) function inconsequently. TDNN perform a linear into tokens and it accomplishes on POS, NER and further complicated steps of NLP, like SLR [10]. Recurrent Neural Network (RNN) is techno scientifically on acknowledged handwriting. It presents dynamic ephemeral arbitrary by performing the primary element of the network. RNN connection into blends accompanies an addressed series. RNN further deal with variable length and present involvement by timestamps [10]. Other hand Convolution Neural Networks (CNN) obtain token to engrave feature by implementing convolution method and build an artificial network. CNN produces output from convolution to D Dimensional vector and transfers them to choose the most appropriate feature of presented features called pooling. Deciding the best feature by pooling it developed the ANN for text classification. CNN good for clustering but starving on learning sequential knowledge. When RNN train neural nets by backpropagation algorithm, vanishing gradient problem is occurring. To supervise this problem, Long-Short-Term-Memory (LSTM) is implemented. LSTM control log-logistic function and select the parameters [1].

D. Multi-Tasking Learning

Multitask learning (MTL) in Deep Neural Network is the method where NN perform several learning processes at the identical time to gain a complementary advantage. In NN, associated task particular feature is essential for another feature. In NLP, POS prediction feature activity is also accomplished for SRL and NER. Adjustment or upgrade to POS task also generalization for SRL with NER. In the deep layers of architected NN automatically learn. NN get trained on related task by according deep layers. Maximum the time, latest layer in the network is liability specified and according to the layer enhance the performance. In MTL tasks, cascading feature is the numerous dynamic way to accomplish the aspired output. Use task feature to another task is obvious to use an SLR and POS classifier and use the result to another feature to train a parser [11]. When several tasks labeled for in one dataset, shallow manner can be applied to the task jointly and one all task labels at the same time by a unique model. In shallow joint training can improve joint training on POS tagging and noun-phrase chunking task. Relation extraction, parsing, and NER can be jointly training in the statically parsing model to improve the achievement.

IV. Conclution

Natural Language Processing (NLP) formulate its demand with Deep Learning (DL) method. Artificial Neural Network (NN) and non-linear process, Natural Language processing tools enhance increased accurate and efficient and DL rule collaboration was introduced with the primary relationship. In the field of NLP, DL completely takeover Text Classification and Categorization, Named Entity Recognition (NER), Part-of-Speech Tagging (POST), Semantic Parsing and Question Answering, Paraphrase Detection, Language Generation and Multi-document Summarization, Machine Translation, Speech Recognition, Character Recognition, Spell Checking etc. NLP methods are modified when DL was associated. NLP process like POS, NER, Morphology, Syntactic Parsing, Coreference resolution are associated with Neural Network like; ANN, TDNN, RNN, CNN are discussed with the relation and in NLP process and tools of DL like; LSTM, MTL, DV, CBOW, VL, WA, SRL, Non-Linear Function maintain relationship with NLP process to collaborating basic relation.

REFERENCES

[1] J. Schmidhuber, "Deep learning in neural networks: An overview", Neural Networks, vol. 61, pp. 85-117, 2015.

[2] M. Hjouj, A. Alarabeyyat and I. Olab, "Rule Based Approach for Arabic Part of Speech Tagging and Name Entity Recognition", International Journal of Advanced Computer Science and Applications, vol. 7, no. 6, 2016.

[3] S. Fahad, "Design and Develop Semantic Textual Document Clustering Model", Journal of Computer Science and Information Technology, vol. 5, no. 2, 2017.

[4] J. Zheng, W. Chapman, R. Crowley and G. Savova, "Coreference resolution: A review of general methodologies and applications in the clinical domain", Journal of Biomedical Informatics, vol. 44, no. 6, pp. 1113-1122, 2011.

[5] A. Kaczmarek and M. Marcińczuk, "A preliminary study in zero anaphora coreference resolution for Polish", Cognitive Studies | Études cognitives, no. 17, 2017.

[6] W. Yafooz, S. Abidin, N. Omar and R. Halim, "Dynamic semantic textual document clustering using frequent terms and named entity", 2013 IEEE 3rd International Conference on System Engineering and Technology, 2013.

[7] H. Li, "Deep learning for natural language processing: advantages and challenges", National Science Review, vol. 5, no. 1, pp. 24-26, 2017.

[8] S. Fahad and W. Yafooz, "Review on Semantic Document Clustering", International Journal of Contemporary Computer Research, vol. 1, no. 1, pp. 14-30, 2017.

[9] O. Bonami and B. Sagot, "Computational methods for descriptive and theoretical morphology: a brief introduction", Morphology, vol. 27, no. 4, pp. 423-429, 2017.

[10] G. Goth, "Deep or shallow, NLP is breaking out", Communications of the ACM, vol. 59, no. 3, pp. 13-16, 2016.

[11] R. Gunderman, "Deep Questioning and Deep Learning", Academic Radiology, vol. 19, no. 4, pp. 489-490, 2012.