IOP Conf. Series: Materials Science and Engineering **407** (2018) 012114 doi:10.1088/1757-899X/407/1/012114

Deep Learning – Now and Next in Text Mining and Natural Language Processing

N I Widiastuti

Informatics Engineering Department, Universitas Komputer Indonesia,
Jl. Dipatiukur 112-116, Bandung, Indonesia

nelly.indriani@email.unikom.ac.id

Abstract. This study was conducted to find out what has not been discussed in last research in domain text mining and NLP using Deep Learning. In this literature review has covered more than 50 articles that can be accessed from a various portal of scientific articles. The focuses are conducted on important elements of a network and what influences them. The results of this study indicate that currently, more research discusses how data is presented. This is due to the assumption that input data is very important in performance of an algorithm. modification of network architecture and a combination of techniques are also attracted researchers. The next studies are many research points that can be done in text mining and NLP using Deep Learning. Especially on input features, learning methods, and the others issues in the domain of Text mining and NLP.

1. Introduction

Deep Learning is part of Artificial Neural Network. Deep Learning was introduced in 1986 by Rina Dechter [1] and google trends "says" that the algorithm continues to grow to start "booming" around 2014. Research conducted using deep learning has been done in many fields. They are medical imaging, Bioinformatics, speech recognition, including Text mining and Natural language processing (NLP). Text mining and NLP are two fields of researches that mutually exclusive. Both of these fields, processing text from unstructured documents, such as data that can not be classified or sorted numerically [2]. According to google trend, NLP is more interesting than text mining. Nonetheless, internet technology causes the number of documents containing natural languages increase sharply. Both text mining and NLP are required to handle this document surge.

A review of some research using Deep Learning was done by Marta et al. in an article that addresses issues on Deep Learning in Machine Translation [3]. In Machine Translation domain, they review from a standpoint of issues in statistical machine translation. Another review was done by Guoqiang et al in an article that discusses how to learn machine learning cases based on a data representation [4]. In the Deep Learning for Big Data article, Bilal Jan et al compared various techniques in Deep Learning. They tested various combinations of Deep Learning parameters and architectures [5].

More than 50 studies have been reviewed to get some ideas of what other researchers have done and what to do next. This literature review shows that there are currently many studies that modify the architecture, data representation, and combinations of techniques or incorporate all three with Deep Learning. However, none of these studies specifically address the main issues in Text mining and NLP domains. Such as context recognition or feature selection. Another thing that still needs to be examined is the appropriate network model for certain cases.

2. Method

The focus of this study is to review some literature that used Deep Learning to solved problems related to text mining and NLP. This review was conducted on over 50 studies undertook Deep Learning development. The literature used are sourced from Science Direct, Elsevier, IEEE, Arxiv and several other portals. Publication span between 2010 and 2018. All articles grouped by their focus. In this study, the articles divided into 3 groups namely, data representation, modify the network architecture, and research that developed learning method by adding certain algorithms, or by combining the three.

3. Results

Reviews were conducted on over 50 published studies from 2011 until early 2018. Discussions are based on main contributions each research undertakes. In this study, the point of view of discussion will be divided into architectural modification, how to represent data to be processed, and add other methods or algorithms respectively in sections 3.1, 3.2, and 3.3. All three of these points may be many incisions and are part of "Now" while "Next" is in part 4.

3.1. Architecture

In this section, the discussion focus on research that modifies neural network architecture. Modifications are generally made as a necessary adjustment to system requirements. Most of it is on the hidden layer, activation function or training method that affects the overall performance of the neural network. According to Andrew Ng in a lecture of Machine Learning, stated that the things that must be considered in Machines Learning are how to modify the network properly, choose a favorable optimization method and able to prevent the occurrence of overfitting. Networks can be modified based on activation functions such as Maxout and Rectified Linear Unit (ReLU). Network optimization in Deep Learning generally uses Adagrad, Adam, Adasecant etc. Finally, to prevent Deep Learning over fit using Dropout method.

In general, researchers modify the activation function by using ReLU [6, 7] or Maxout [8, 9]. Modifications on the Hidden Layer are performed using Restricted Boltzmann Machines (RBM) in the summarization system based on queries using RNN [10]. The CNN architecture was tailored to a number of words in the sentence. Thus the channel used is static and dynamic [11]. Kamran kowsari et al. has classified the sentence with the hierarchical architecture. At parent level, the sentence was classified into a more general class than the child level [12]. Local attention in a document is done by modifying the model of Neural Network Language Model developed by Bengio [13, 14].

In the NLP domain, the architecture was tailored to the purpose of the developed system. In systems that define Biomedical entities, the RNN architecture is adapted to NE [15]. Architecture Bilingual Compositional Sentence Model (BiCVM) [16] and Conditional Neural Language Models (CNLM) [17] combined in parsing semantics for question answering system [18]. RNN Encoder-Decoder for a translation engine developed though Kyunghyun Cho et al. Their research aims was training RNN to recognize phrases. Variables of varying length are trained so as having a fixed length (RNN encode). Then, re-trained into different (RNN decode) [19].

3.2. Data representative

All studies definitely need data for training and testing, but how data represented to have an effect on the results of the study. Unstructured data require preprocessing such as tokenizing, filtering or even stemming. In this section, a section that focuses on data representation.

IOP Conf. Series: Materials Science and Engineering **407** (2018) 012114 doi:10.1088/1757-899X/407/1/012114

The research topics in the text mining domains discussed in this section are analytics sentiments, automated summaries, retrieval systems, etc. Sentiment data such as Twitter or product reviews are most commonly used to classify text. This is because the sentence length is relatively short, but has a large number of data. The problem arises because in the data there is often a grammatical error. There are several processes performed to present the data used for example by filtering keywords [6, 20], feature extraction [21, 22], morphological analysis [23], word embedding [24, 25] and dimensional reduction [26].

Methods and objectives of data presentation are the main contributions of several studies. Some of these are an adaptation to large training data with a large number of targets. It is done by extracting features Stacked Denoising Autoencoder (SDA) and SVM [27], to facilitate the classification of sentiment [28], generates features using NTUSD, HowNet-VSA, NTUFSD and iMFinanceSD [29]. To overcome the meaning of the word "no" which is not necessarily negative means was done by arranging the text to form coordinates and results [30]. Assume interest as a verb and topic were considered a noun [31]. Each word in the source document was attached with POS, NER, TF and IDF [32]. Train data set by non-linear mapping [33]. Clustering on the NeuroSynth corpus by word and embedding documents [25].

In the NLP domain, Deep Learning was tested on cases such as an interpreter machine, an answer selection system, a conversation, or to build a corpus. It is necessary for NLP itself such as Name Entity tagger, POS tagger, or chunking for a language. LSTM is used to recognize conversations in Indonesian [34] and CNN to recognized multiclass in the data set [35]. Some data representation methods discussed in this section such as model questions with answers as vectors to connect the two[36], building a multilayer neural network that can be used for various tasks related to NLP such as POS tagging, chunking, NER and role labeling in semantics. Network flexibility is overcome without providing too many specific processes[37,38], and measure the interrelationship between sentences using WordNet for the knowledge base and corpus generated using Deep Learning vector space[39].

3.3. Technical embedding
The initial idea of Deep Learning development was learning to clarify what happens in the hidden layer. In traditional Neural Network, there is an unknown process occurring in the hidden layer. However, this is not yet fully feasible. In this section, some studies still use additional methods to help Deep Learning performance.

Technical embedding in this section is to add another algorithm that has been widely used. The techniques used in some research were greedy algorithm [40], Random Term Frequency and Stochastic Auto-Encoder version [41], Word Graph representation [42], Greedy algorithm and Integer Linear Programming [43], linear machine learning [44] and DCNN for modeling sentence by adding N-gram and Bag of Word [45]. Some studies even used a group of machine learning algorithms at a time to solve the problem, for example, CNN and Gated RNN were used for sentiment classification[46], combine LDA to model the topic of a sentiment with LSTM and CRF[47], RNN with LSTM, Naïve Bayes with Support Vector Machine(NB-SVM) also word2vec and bag-of words[48], distributed Maximally Collapsing Metric Learning (tMCML)[49] which maps the probability function of training data[50].

Similar to what has been done in text mining, various techniques are also added in NLP. Research that adds other techniques to identify POS Tag in English and Portuguese by using unsupervised methods to train words, and supervised for characters [51], SVM and Tf-Idf [52], combine kernel in Deep architecture [53] or CNN and word embedding [54]. A recursive technique was done to parse segments of some image and sentence features, to produce a complete view and sentence [55].

4. Discussion
Architecture in Deep Learning has indeed become the focus of most basic research. However, most studies undertake further development of the results obtained in case development [19, 31]. Few researchers specifically address the effect of modifications on the hidden layer on the expected features. Other than that

it was expected to realize Deep Learning which is completely free from feature engineering [56]. Architectures can also be developed based on the data set [12, 38]. Deep Learning parameter test was also a concern [30].

In addition to enhancing other algorithms, the process of Deep learning can be improved by utilizing available training data. In addition, the process of Deep learning can be improved by utilizing the available training data. Several studies have led to handling of data set. Large data set were trained because they required alignment using Bootstrap [18, 25]. Data set can also be combined with other data from several related subjects [34, 53] or in polarized data [20]. The structure of a sentence in the NLP domain may also affect the technique to be used [13, 23, 28, and 36].

Deep Learning's initial idea was to use the hidden layer as a feature descriptor. Nonetheless, adding other techniques to better interests was still considered legitimate. The use of word extraction methods and information in sentences such as POS tags, NER, or chunk [35, 51]. The relationship between the document and the cluster image becomes an advanced document classification work based on the image [25]. Some optimization methods [26] or sentence features are added to have improved network performance [21, 41].

Modifications on the hidden layer, learning methods or learning parameters should not make the overall complexity of the system higher. The simplicity of the process in Deep Learning should take precedence. The issue of features selection in text mining has not been noticed. In further research, the relationship between extraction and features selection with hidden layers were still needed. A lot of big questions and high expectations on in-depth learning to be non-specific or Deep Learning Generalization still untouched. Whether Deep Learning was able to recognize the context of sentences or documents.

5. Conclusion

We have submitted a brief initial discussion on some research in Text Mining and NLP that use Deep Learning at this time. So, there was still many research gaps to develop Deep learning especially in text mining and NLP. How Deep Learning understand issues such as the context of a sentence, the effect of features on performance and architecture were still unanswered issues. Finally, this short review was expecting something to be used for basic research that new to Deep Learning or for those who want to know the direction of developing Deep Learning especially in the Text Mining and NLP domains.

Acknowledgements

Thanks to the UNIKOM research institute and community service that has funded this research in internal research scheme 2018.

References

[1] Dechter R 2015 "Learning While Searching in Constraint," *Conference Proceedings of the 5th National Conference on Artificial Intelligence AAAI* **86** 178–183.

[2] Weiss S M, Indurkhya N, Zhang T and Damerau F J 2005 *Text Mining Predictive Methods for Analyzing Unstructure Information* (United states of America: Springer).

[3] Costa-jussà M R, Allauzen A, Barrault L, Cho K and Schwenk H 2017 "Introduction to the special issue on deep learning approaches for machine translation," *Comput. Speech Lang* **46** 367–373.

[4] Zhong G, Wang L, Ling X and Dong J 2017 "ScienceDirect An overview on data representation learning : From traditional feature learning to recent deep learning," *J. Financ. Data Sci.* **2** 265-278.

[5] Jan B, Farman H, Khan M, Imran M, Islam I U, Ahmad A, Ali S and Jeon G 2017 "Deep learning in big data Analytics: A comparative study," *Comput. Electr. Eng.* **1** 1–13.

[6] Ramadhani A M 2017 "Twitter Sentiment Analysis using Deep Learning Methods," *7th International Annual Engineering Seminar (InAES)* (Yogyakarta).

IOP Conf. Series: Materials Science and Engineering **407** (2018) 012114 doi:10.1088/1757-899X/407/1/012114

[7] Allamanis M and Sutton C 2016 A "Convolutional Attention Network for Extreme Summarization of Source Code," *Int. Conf. on Machine Learning* vol **48** (New York, NY) 2091–2100.
[8] Goodfellow I J and Courville A 2014 "An empirical investigation of catastrophic forgeting in gradientbased neural network," *Proc. of Int. Conf. on Learning Representations.*
[9] Luong M-T, Pham H and Manning C D 2015 "Effective Approaches to Attention-based Neural Machine Translation," *Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing* (Lisbon: Portugal) 1412-1421.
[10] Liu Y, Zhong S, Li W, Hom H and Kong H 2012 "Query-Oriented Multi-Document Summarization via Unsupervised Deep Learning," *Proc. Twenty-Sixth AAAI Conf. Artif. Intell.* 1699–1705.
[11] Kim Y 2011 "Convolutional Neural Networks for Sentence Classification," *Conf. on Empirical Methods in Natural Language Processing* (Doha: Qatar) 1746–1751.
[12] Kowsari K, Brown D E, Heidarysafa M, Meimandi K J, Gerber M S and Barnes L E 2017 "HDLTex : Hierarchical Deep Learning for Text Classification," *16th IEEE Int. Conf. on machine learning and aplication* 364–371.
[13] Bengio Y, Vincent P, Ducharme R and Jauvin C 2003 "A Neural Probabilistic Language Model," *J. Mach. Learn. Res. 3* **3** 1137–1155.
[14] Rush A M and Weston J 2015 "A Neural Attention Model for Sentence Summarization," *Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing* (Lisbon, Portugal: Association for Computational Linguistics) 379–389.
[15] Jiang Z, Li L, Huang D and Jin L 2015 "Training Word Embeddings for Deep Learning in Biomedical Text Mining Tasks," *IEEE Int. Conf. on Bioinformatics and Biomedicine* 625–628.
[16] Hermann K M and Blunsom P 2014 "Multilingual Models for Compositional Distributed Semantics," *Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics* (Baltimore, Maryland, USA) 58–68.
[17] Mnih A and Hinton G 2007 "Three New Graphical Models for Statistical Language Modelling," *Proc. of the 24th Int. Conf. on Machine Learning* 641-648.
[18] Grefenstette E, Blunsom P, Freitas N De and Hermann K M 2013 "A Deep Architecture for Semantic Parsing," *Proc. of the ACL 2014 Workshop on Semantic Parsing* (Baltimore, Maryland USA) 22-27.
[19] Cho K, Merrienboer B van, Gulcehre C, Bahdanau D, Bougares F, Schwenk H and Yoshua Bengio 2014 "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Doha, Qatar) 1724–1734.
[20] Zhang Z, He Q, Gao J and Ni M 2018 "A deep learning approach for detecting traffic accidents from social media data," *Transp. Res. Part C* **86** 580–596.
[21] Singh S P, Kumar A, Mangal A and Singhal S 2016 "Bilingual Automatic Text Summarization Using Unsupervised Deep Learning," *Int. Conf. on Electrical, Electronics, and Optimization Techniques (ICEEOT)* 1195–1200.
[22] R J V C I, Jiang B, Yang J, Lv Z, Tian K, Meng Q and Yan Y 2017 "Internet cross-media retrieval based on deep learning q," *J. Vis. Commun. Image Represent.* **48** 356–366.
[23] Tsumoto S, Kimura T, Iwata H, Hirano S, Tsumoto S, Kimura T, Iwata H and Hirano S 2017 "Mining Text for Diagnosis Disease," *5th Int. Conf. on Information Technology and Quantitative Management, ITQM* vol 122 (Elsevier B.V.) 1133–1140.
[24] Huang D 2017 "Using Deep Learning To Recognize Biomedical Entities," *2017 12th Int. Conf. on Intelligent Systems and Knowledge Engineering (ISKE)* 8–11.
[25] Monti R, Lorenz R, Leech R, Anagnostopoulos C and Montana G 2016 "Text-mining the NeuroSynth corpus using Deep Boltzmann Machines," *Int. Workshop on Pattern Recognition in Neuroimaging*

(PRNI), 2016 (Trento, Italy: IEEE).
[26] Sheng X, Road S, Road Y, Hospital S C and Road F 2016 “A Novel Text Mining Algorithm based on Deep Neural Network,” *Int. Conf. on Inventive Computation Technologies (ICICT),* 2–7.
[27] Glorot X 2011 “Domain Adaptation for Large-Scale Sentiment Classification : A Deep Learning Approach,” *Proc. of the 28 th Int. Conf. on Machine Learning* (Bellevue, WA, USA) 513-520.
[28] Uysal A K and Murphey Y L 2017 “Sentiment classification : Feature selection based approaches versus deep learning,” *IEEE Int. Conf. on Computer and Information Technology* 23–23.
[29] Day M and Lee C 2016 “Deep Learning for Financial Sentiment Analysis on Finance News Providers,” *IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining* (San Francisco, CA, USA) 1127–1134.
[30] Wandabwa H, Naeem M A and Mirza F 2017 “Document Level Semantic Comprehension of Noisy Text Streams via Convolutional Neural Networks,” *Proc. of the 2017 IEEE 21st Int. Conf. on Computer Supported Cooperative Work in Design* 475–479.
[31] Lu T, Hou S, Chen Z, Cui L and Zhang L 2015 “An Intention-Topic Model Based on Verbs Clustering and Short Texts Topic Mining,” *2015 IEEE Int. Conf. on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing* 837-842.
[32] Nallapati R, Xiang B and Santos C dos 2016 “Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond,” *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL) 2016* (Berlin, Germany: Association for Computational Linguistics) 280–290.
[33] Min R, Stanley D A and Bonner A 2009 “A Deep Non-Linear Feature Mapping for Large- Margin kNN Classification,” *2015 IEEE Int. Conf. on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing* 357-366.
[34] Chowanda A and Chowanda A D 2017 “Recurrent Neural Network to Deep Learn Conversation in Indonesian,” *Procedia Computer Science* **116** (Bali, Indonesia: Elsevier B.V.) 579–586.
[35] Dong X, Qian L, Guan Y, Huang L, Yang J and Yu Q 2016 “A Multiclass Classification Method Based on Deep Learning for Named Entity Recognition in Electronic Medical Records,” *Scientific Data Summit (NYSDS)* (New York: IEEE) **2** 59-65.
[36] You L, Moritz K, Phil H and Stephen B 2014 *Deep Learning for Answer Sentence Selection arXiv* (*Preprint* arXiv:1412.1632).
[37] Collobert R, Weston J, Karlen M, Bottou L, Kavukcuoglu K and Kuksa P 2011 “Natural Language Processing (almost) from Scratch,” *J. Mach. Learn. Res.* **12** 2493-2537.
[38] Zheng X, Chen H and Xu T 2013 “Deep Learning for Chinese Word Segmentation and POS Tagging,” *Proc. of the 2013 Confe. on Empirical Methods in Natural Language Processing* (Seattle, Washington, USA,) 647–657.
[39] Banea C, Chen D, Arbor A and Cardie C 2014 “SimCompass : Using Deep Learning Word Embeddings to Assess Cross-level Similarity,” *Proc. of the 8th Int. Workshop on Semantic Evaluation (SemEval 2014)* (Dublin, Ireland) 560–565.
[40] Nallapati R, Zhai F and Zhou B 2016 “SummaRuNNer : A Recurrent Neural Network Based Sequence Model for Extractive Summarization of Documents,” *Proc. of the Thirty-First AAAI Conf. on Artificial Intelligence (AAAI-17)* 3075–3081.
[41] Yousefi-azar M and Hamey L 2016 “Text Summarization Using Unsupervised Deep Learning,” *Expert Syst. with Appl. An Int. J.* **68** 93–105.
[42] Zhang Y, Er M J and Pratama M 2016 “Extractive Document Summarization Based on Convolutional Neural Networks,” *Industrial Electronics Society , IECON 2016 - 42nd Annual Conf. of the IEEE*

(Florence, Italy: IEEE) 918–922.
[43] Cao Z, Wei F, Dong L, Li S and Zhou M 2015 "Ranking with Recursive Neural Networks and Its Application to Multi-Document Summarization," *Proc. of the Twenty-Ninth AAAI Conf. on Artificial Intelligence* 2153–2159.
[44] Araque O, Corcuera-platas I, Sánchez-rada J F and Iglesias C A 2017 "Enhancing deep learning sentiment analysis with ensemble techniques in social applications," *Expert Syst. Appl.* **77** 236–246.
[45] Kalchbrenner N, Grefenstette E and Blunsom P 2014 "A Convolutional Neural Network for Modelling Sentences," *In Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics* (Baltimore, Maryland, USA) 655-665.
[46] Tang D, Qin B and Liu T 2015 "Document Modeling with Gated Recurrent Neural Network for Sentiment Classification," *Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing* (Lisbon, Portugal) 1422–1432.
[47] Jansson P and Liu S 2017 "Topic modelling enriched LSTM models for the detection of novel and emerging named entities from social media," *2017 IEEE Int. Conf. on Big Data (Big Data)* (Boston, MA, USA: IEEE) 4329-4336.
[48] Liu G, Xu X, Deng B, Chen S and Li L 2016 "A Hybrid Method for Bilingual Text Sentiment Classification Based on Deep Learning," *2016 IEEE SNPD* (Shanghai, China: IEEE) 93-98.
[49] Min R and Bonner A 2007 "Deep Supervised t-Distributed Embedding," *Proc. of the 27th Int. Conf. on Machine Learning* (Haifa) 791-798.
[50] Xu L, Jiang C and Ren Y 2015 "Deep Learning in Exploring Semantic Relatedness for Microblog Dimensionality Reduction," *2015 IEEE Global Conf. on Signal and Information Processing (GlobalSIP)* 98–102.
[51] C´ ıcero Nogueira dos Santos and Zadrozny B 2014 "Learning Character-level Representations for Part-of-Speech Tagging," *Proc. of the 31 st Int. Conf. on Machine Learning* **32** (Beijing,China) 1818-1826.
[52] Ranzato M A and Szummer M 2008 "Semi-supervised Learning of Compact Document Representations with Deep Networks," *Int. Conf. on Machine Learning* (Helsinki, Finland) 792–799.
[53] Filice S, Castellucci G, Basili R and Tor R 2017 "Deep Learning in Semantic Kernel Spaces," *Proc. of the 55th Annual Meeting of the Association for Computational Linguistics* (Vancouver, Canada) 345–354.
[54] Xue Q, Liu L, Chen W and Chuah M C 2017 "Automatic Generation and Recommendation for API Mashups," *Int. Conf. on Machine Learning and Applications* (Cancun, Mexico: IEEE) 119–124.
[55] Lin C C, Ng A Y and Manning C D 2011 "Parsing Natural Scenes and Natural Language," *Proc. of the 28 th Int. Conf. on Machine Learning* (Bellevue, WA, USA) 129-136.
[56] Luu T M, Phan R and Davey R 2017 "A Multilevel NER Framework for Automatic Clinical Name Entity Recognition," *IEEE Int. Conf. on Data Mining Workshop* 1134-1143.