International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X;pISSN:2321-2241 Volume: 6; Issue: 6;June -2017

Deep Learning Based Natural Language Processing Approach for Extracting Tacit Knowledge

R.J. Anandhi [1],Jeffrey Bakthakumar [2], Murali Manohar [3]

[1] Vice Principal, Professor & HOD(CSE)The Oxford College of Engineering, Bangaluru - 560068
[2,3] VIT University ,Vellore , Tamilnadu - 632014

Abstract : Deep learning is a category of machine learning that trains a computer to perform human-like tasks, such as speech recognition, pattern recognition or predictive analysis. Instead of organizing data to run through predefined equations, deep learning sets up basic parameters about the data and trains the computer to learn on its own by recognizing patterns using many layers of processing. The current interest in deep learning is because of cognitive computing, software applications that understand human input and can respond in humanlike form or output. Deep learning techniques have improved the ability to classify, recognize, detect and describe, or basically to understand and respond. Xbox, Skype, Google assistant and Apple's Siri®, Cortana are already employing deep learning technologies in their systems to recognize human speech and voice patterns. A specialization of text mining used to discover patterns in customer complaints, physician notes or news reports. A traditional approach to analytics is to use the data at hand to engineer features to derive new variables, then select an analytic model and finally estimate the parameters of that model, but completeness and correctness depend on the quality of the model and its features. The new approach with deep learning is to replace the formulation and specification of the model with hierarchical characterizations (or layers) that learn to recognize latent features of the data from the regularities in the layers. The idea with deep learning is to move from feature engineering to feature representation. In this paper, we have analysed two deep learning architectures, Convolutional neural network (CNN) and recurrent neural networks (RNN), and have tested for their appropriateness in natural language processing. This work is the sincere attempt to compare and to use them for tacit knowledge elicitation in software documents.

Keywords: Tacit knowledge, text mining, Neural networks, deep learning, knowledge elicitation.

1. Introduction

Tacit knowledge enablers and inhibitors include processes, structures and activities that have the potential of influencing knowledge flows and as a consequence have an effect in organisational knowledge stocks. The presence of knowledge enablers positively affects knowledge stocks and flows. However, their absence potentially acts as an inhibitor for such resources [1].Artificial intelligence functions in away, that itimitates the workings of the human brain in processing data and creating patterns for use in decision making. Deep learning is a subset of machine learning in artificial intelligence (AI) that has networks which are capable of learning unsupervised from data that is unstructured or unlabelled. It is also commonly referred as Deep Neural Learning or Deep Neural Network.One of the most common AI techniques used for processing Big Data is Machine Learning(ML). Machine learning deals with the **problems of extracting features from data** so as to solve many different predictive tasks.eg: forecasting, imputing missing data, detecting anomalies, ranking, classifying.Machine learning is a self-adaptive algorithm that gets improved analysis and patterns with experience or with new added data. If a digital payments company wanted to detect the occurrence of or potential for fraud in its system, it could employ machine learning tools for this purpose. Even some intelligent games tune to the capacity of the player and makes the games tougher accordingly. The computational algorithm built into a computer model will process all transactions happening on the digital platform, find patterns in the data set, and point out any anomaly detected by the pattern. AI is the broadest term, -applying to any technique that enables computers to mimic human intelligence, using logic, if-then rules, decision trees, and machine learning (including deep learning)[8].

But, machine learning is basically a subset of AI that includes abstruse statistical techniques that enable machines to improve at tasks with

experience. The category includes deep learning, which is nothing but the subset of machine learning composed of algorithms that permit software to train itself to perform tasks, like speech and image recognition, by exposing multi-layered neural networks to vast amounts of data.

Deep learning algorithms seek to exploit the unknown structure in the input distribution in order to discover good representations, often at multiple levels, with higher-level learned features defined in terms of lower-level features. The objective is to make these higher-level representations more abstract, with their individual features more invariant to most of the variations that are typically present in the training distribution, while collectively preserving as much as possible of the information in the input. [2].The interests in deep learning is gaining more importance in recent years because of various reasons. Foremost is the availability and accumulation of large datasets, and also feature extraction is not easy in machine learning for larger datasets. The advancement in technology has resulted in steep hardware advances along with decrease in prices. This paired along with innovative techniques have resulted in rapid automation most of the process, thereby resulting in huge accumulation of data. The varied applications in deep learning includes classification of millions of labelled images as in google, machine translation ,speech recognition and synthesis, as seen to be in major use in voice controlled assistance, automatic image caption generation, self-driving cars and intelligent and innovative game playing.

2. Deep Learning Algorithms

Deep Learning methods are a modern update to Artificial Neural Networks that exploit abundant cheap computation.They are concerned with building much larger and more complex neural networks and, many methods are concerned with semi-supervised learning problems where large datasets contain un structed data or very little labelled data.The most popular deep learning algorithms are Deep Boltzmann Machine (DBM), Deep Belief Networks (DBN),Convolutional Neural Network (CNN) and Stacked Auto-Encoders. Apart from these traditional methods, there are other algorithms, like reinforcement learning, evolutionary algorithms commonly discussed in computational intelligence, Graphical Models and techniques used in Computer Vision. Generally, these methods can be classified under feature selection algorithms, yet are evaluated using the algorithm's accuracy. Performance metrics basically aims at minimizing a loss function thereby implying a faster learning rate.

Implementing a machine learning algorithm will give us a deep and practical appreciation for how the algorithm works. This knowledge can also help oneself to internalize the mathematical description of the algorithm by thinking of the vectors and matrices as arrays and the computational intuitions for the transformations on those structures. Tensor Flow is a Python library for fast numerical computing created and released by Google. Tensor Flow is intended to be used to develop deep learning models.

The performance of the deep learning models can be improved based on data, or with algorithms or with fine tuning the parameters of the algorithms. Normally to tunea neural network algorithm, to improve performance, one can depend on factors like weight initialization, learning rate, activation functions, re-evaluate the network topology, monitor batches and epochs and early termination or converging the process. A major breakthrough has been achieved in performance with the advent of ensembles. Ensembles basically combine the predictions from multiple models. After algorithm tuning, this is the next big area for improvement suggested by either combining models, or combining views. Or by stacking.

If one desires to explore the usage of deep learning for any of the above discussed application, the strengths and weakness of deep learning also must be clearly documented. The foremost noteworthy strengths is that deep learning is the current state-of-the-art for certain domains, such as computer vision and speech recognition. Deep neural networks perform very well on image, audio, and text data, and they can be easily updated with new data using batch propagation. Their architectures (i.e. number and structure of layers) can be adapted to many types of problems, and their hidden layers reduce the need for feature engineering. These algorithms also have their own pitfalls, that one should carefully trends upon. Deep learning algorithms are usually not suitable as general-purpose algorithms because they require a very large amount of data. In fact, they are usually outperformed by tree

ensembles for classical machine learning problems. In addition, they are computationally intensive to train, and they require much more expertise to tune (i.e. set the architecture and hyper parameters).

3. Natural language Processing with Deep Learning

Deep neural networks (DNN) have revolutionized the field of natural language processing (NLP). Convolutional neural network (CNN) and recurrent neural network (RNN), the two main types of DNN architectures, are widely explored to handle various NLP tasks. CNN is supposed to be good at extracting position-invariant features and RNN at modelling units in sequence. The state of the art on many NLP tasks often switches due to the battle between CNNs and RNNs.

The RNN model learns itself from the data how to represent memory. While shallow feedforward neural networks (those with just one hidden layer) can only cluster similar words, recurrent neural network which can be considered as a deep architecture, can perform clustering of similar histories. This allows for instance efficient representation of patterns with variable length.One key factor that limits the scalability of RNNLMs is the computation of the normalization term in the output layer. This has a significant impact on both training and testing, especially when a large output vocabulary is used, in particular, in full output based RNNLMs. One technique that can be used to improve the testing speed is introducing the variance of the normalization term into the conventional cross entropy based objective function. By minimizing the variance of the normalization term during training, the normalization term at the output layer can be ignored during testing time thus gaining significant improvements in speed. However, the explicit computation of this normalization term is still required in training and it does not improve training speed.

Convolutional Neural Networks is one of the most widely used model of deep learning. Convolutional neural networks (CNNs), originally invented in computer vision, has recently attracted much attention in natural language processing (NLP) on problems such as sequence labelling, semantic parsing and search query retrieval In particular, recent work on CNN-based sentence modelling has achieved excellent, often state-of-the-art, results on various classification tasks such as sentiment, subjectivity, and question-type classification[4].

Convolution is a mathematical operation between two signals, where two signals are combined into new form of third signal. It is the most used technique in Digital Signal Processing. The most important characteristic of given signal become very obvious in the new representation of given signal. It also could reconstruct the input signal from output signal when the impulse response of the system is given by deconvolution. Convolution is important because it relates to input signal, the output signal and the impulse response.

Convolutional neural networks are a type of artificial neural networks that use convolution methodology to extract the features from the input data to increase the number of features. The basic steps are as follows: Considering the input data set, we decide and select N number of convolution functions, also known as filters, where N = maximum number of features we want. After the selection of filter functions we convolute the input with the filters. The output from the convolution is a D-Dimensional vector, which is then sent for pooling. Pooling is a way of selecting most relevant features from the set of given features. Mostly for text analytics we select the features with max or average values. Once the best features are selected, the normal architecture of ANN is followed.

CNN demonstrates a powerful ability for sentence modelling. Based on this successful sentence encoding ability, CNN is also can be served as powerful technique for Machine Translation tasks due to the continuous representation from word embedding[9]. The first CNN based machine translation model is proposed with the name of Recurrent Continuous Translation Models (RCTM). RCTM map the source language to a probability distribution over the sentence in the target language without losing generality. In this model, they use CNN as an encoding method and use Recurrent Neural Network for decoding or generating the target sentence.RCTM is one simple extension from sentence modelling to machine translation without using word alignment.

Success stories of Deep Learning !!

There are many useful and most commonly used applications which demonstrates the power of deep learning.Xbox is a video gamingbrand created and owned by Microsoft. It represents a series of video game consoles developed by Microsoft, with three consoles. Apple Siri,(speech interpretation and Recognition interface), is an intelligent personal assistant, part of Apple Inc.'s iOS, watchOS, macOS, and tvOS operating systems. The assistant uses voice queries and a natural language user interface to answer questions, make recommendations, and perform actions by delegating requests to a set of Internet services. The software adapts to users' individual language usages, searches, and preferences, with continuing use. Returned results are individualized. Its speech recognition engine is provided by Nuance Communications, and Siri uses advanced machine learning technologies to function. Siri supports a wide range of user commands, including performing phone actions, checking basic information, scheduling events and reminders, handling device settings, searching the Internet, navigating areas, finding information on entertainment, and is able to engage with iOS-integrated apps. Google Assistant is considered as an upgrade or an extension of Google Now - designed to be personal - as well as an expansion of Google's existing "OK Google" voice controls.Google Now feature smartly pulls out relevant information for the user[7]. It knows where one works, and it knows meeting locations and travel plans, the sports teams you like, and what interests you. This data is presented in cards and through reminders on the Android device.

Google Assistant will be able to keep track of the conversation, determine context, and audibly respond with the right information. The Assistant is voice-enabled artificial intelligence (AI) software that bundles machine learning, the Google Knowledge Graph, and voice and image recognition natural language processing (NLP) to build a “personal Google for each and every user”.Google’s AI research group, DeepMind, unveiled a new speech synthesis technique called WaveNet. Using individual sounds instead of complete syllables or words, WaveNet relied on a ‘computationally expensive’ process to generate complex, realistic-sounding audio. This is the voice powering the Assistant.

Cortana is an intelligent personal assistant created by Microsoft for windows. Cortana can set reminders, recognize natural voice without the requirement for keyboard input, and answer questions using information from the Bing search engine. (e.g., current weather and traffic conditions, sports scores, biographies). Searches will only be made with Microsoft Bing search engine and all links will open with Microsoft Edge, except when a screen reader such as Narrator is being used, where the links will open in Internet Explorer.The natural language processing capabilities of Cortana are derived from Tellme Networks (bought by Microsoft in 2007) and are coupled with a Semantic search database called Satori.

4. Conclusion

An abstract, yet essential process in knowing how the experts work and how their expertise can be captured revolves around how a knowledge developer captures the expert’s views and experiences. It also involves in converting that expertise into a coded program. The main challenge in the whole process is mainly because even the expert might not be aware of the latent knowledge in him or her. The next bottleneck is the need for voluntary and eager involvement of the employee, in the firm’s quest for his tacit knowledge, showing no reluctance in sharing the same. Our main problem definition includes in how to effectively organize software project lessons learnt and devise some practices to convert those tacit to explicit knowledge for usage in maintaining and developing future software projects[5]. We have found that the systematic way of gathering tacit knowledge reduces the challenges in later codification process. Significant research is in progress in the field of Tacit Knowledge management, but very few researchers are keen to extract knowledge from either logs or predefined templates, questionnaires available or from recorded sections of any project management tool log books[7]. The optimal utilization of the employee’s knowledge and capacities, are essential for small-scale business success. The small-scale companies are noted for their fast and streamlined decision making process as it has less bureaucracy and lower fixed costs. Due to these processes, the response time is very fast to the requirements of first-hand clienteles. Hence when it comes to product development, these features with their advantages make them well suited to incorporate novelty. It is very crucial to maintain the tacit knowledge in small scale IT Company due to the limited number of employees and also the attrition rate is more. The Tacit Knowledge derived from experience and insight, are subjective, cognitive and experiential learning; and also, highly personalized and difficult to formalize, hence also referred as sticky knowledge[3]. As the concept of tacit knowledge refers to individual, it has to be acquired without formal symbols or words. All these qualities makes the extraction of tacit knowledge using deep knowledge a more viable option and we feel that

more work done in this area will result in more fruitful application to the software industry.

5. References

1. Garcia-Perez, A. and Mitra, A. "Tacit Knowledge Elicitation and Measurement in Research Organisations: a Methodological Approach." *The Electronic Journal of Knowledge Management* Volume 5 Issue 4, pp. 373 - 386, available online at www.ejkm.com

2. *Yoshua Bengio,* "Deep Learning of Representations for Unsupervised and Transfer Learning." *Proceedings of ICML Workshop on Unsupervised and Transfer Learning, PMLR 27:17-36, 2012.*

3. Gendreau, Olivier, and Pierre N. Robillard. "Knowledge acquisition activity in software development." *Advances in Information Systems and Technologies.* Springer Berlin Heidelberg, 2013. 1-10.

4. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011). Naturallanguage processing (almost) from scratch. volume 12, pages 2493–2537.

5. M. Ali Babar, I. I. Gorton, and R. Jeffery, "Capturing and Using Software Architecture Knowledge for Architecture-Based Software Development," *Fifth Int. Conf. Qual. Softw.*, pp. 169–176, 2005.

6. Davenport, T. H. 1997. Information ecology: Mastering the information and knowledge environment. New York:Oxford University Press. 255 pp.

7. Davenport, T., DeLong, D.W. & Beers, M.C. 1998. Successful knowledge management projects.Sloan Management Review, 39(2), pp. 43-57.

8. Moss, L.T. & Atre, S. 2003. Business intelligence roadmap: The complete project lifecycle for decision-support applications. Boston, MA: Addison-Wesley. 543 pp.

9. Boureau, Y.-L., Ponce, J., and Lecun, Y. (2010). A theoretical analysis of feature pooling in visualrecognition. In 27TH INTERNATIONAL CONFERENCE ON MACHINE LEARNING,HAIFA, ISRAEL