Literature Review of Deep Learning Research

Ibrahim Ahmed Saleh1
Asmaa' H. AL_Bayati
Software Engineering Dept., College of Computer and Math., University of Mosul, Mosul – Iraq.
Omar Ibrahim Ahmed
Computer Systems Techniques Department, Northern Technical University, Mosul – Iraq.

Abstract

Deep learning is one of most research fields of artificial intelligence, which has been widely used from high-tech companies in the worlds such as Google and other high-tech companies have increased their exploitation in all fields of artificial intelligence. Deep learning is significant advance in neural networks to solve must of problem-relevant has big features, it has applied in the speech processing, online advertising, natural language processing, image recognition and so on. This paper summarizes the latest advanced in deep learning. First, review three basic types of deep learning, including multilayer perceptron's, convolutional neural networks, and recurrent neural networks. On this basis, further analysis of the emerging new convolutional neural network and recurrent neural network. Then this paper summarizes the deep learning in many fields of artificial intelligence, including speech processing, computer vision and natural language processing. Finally, discussed current problems of deep learning and given the corresponding possible solutions.

Keywords: Artificial intelligence; Convolutional Neural Network; Deep learning; Machine learning; Neural Networks; Recurrent neural network.

دراسة مراجعة لبحوث التعليم المعمق

ابراهيم احمد صالح

اسماء هادي ذنون

قسم هندسة البرمجيات، كلية الرياضيات والحاسبات، جامعة الموصل، موصل – العراق.

عمر ابراهيم احمد

قسم تقنيات الحاسبات، الجامعة التقنية الشمالية، موصل – العراق.

المستخلص

التعلم العميق هو أحد مجالات البحث في الذكاء الاصطناعي استخداماً في الوقت الحاضر، والذي تم استخدامه على نطاق واسع من شركات التكنولوجيا في العالم مثل Google وغيرها من شركات التكنولوجيا مما زاد من استغلالها في مجالات الذكاء الاصطناعي جميعها. يعد التعلم العميق تقدماً كبيراً في الشبكات العصبية لحل المشاكل ذات الصلة وله ميزات كبيرة، وقد تم تطبيقه في معالجة الكلام والإعلان عبر الإنترنت ومعالجة اللغة الطبيعية والتعرف على الصور وما إلى ذلك. تلخص هذه الورقة أحدث التطورات في التعلم العميق. أولاً، راجع ثلاثة أنواع أساسية من التعلم العميق، بما في ذلك الإدراك الحسي متعدد الطبقات، والشبكات العصبية التلافيفية، والشبكات العصبية المتكررة. على هذا الأساس، مزيد من التحليل للشبكة العصبية التلافيفية الجديدة الناشئة والشبكة العصبية المتكررة. في هذا البحث تم تلخيص خوارزميات التعلم العميق في معظم مجالات الذكاء الاصطناعي، بما في ذلك معالجة الكلام ورؤية الكمبيوتر ومعالجة اللغة الطبيعية. أخيراً، ناقش المشكلات الحالية للتعلم العميق وإعطاء الحلول الممكنة المقابلة

الكلمات المفتاحية: الذكاء الاصطناعي، الشبكة العصبية التلافيفية، تعلم عميق، تعلم الآلية، الشبكة العصبية، الشبكة العصبية المتكررة.

1. Introductions

Deep learning neural network technique is simulating resembles the human nervous system and the structure of the brain. It consists of processing units organized in input, hidden and output layers, this progress in Deep learning which led (in October 2016) the US government issued the "National Artificial Intelligence Research and Development Strategic Planning" document. The internet companies such as Google, Microphone, Facebook, Baidu, Tencent, Alibaba and other major Internet companies have also increased their input of industrial intelligence[1].

Industrial intelligence applications are gradually changing in various types of artificial intelligence, they are emerging all kinds of people human life. Deep learning is one of the key research areas of energy, applied to many fields of artificial intelligence such as speech processing, computer vision, natural language processing, etc.

In 1943, McCulloch and Pitts [1] proposed mathematical model of MP neurons type. In 1958, Rosenblatt[2] who was developed the first-generation single-layer perceptron neural network.

The first generation of neural network can distinguish basic shapes such as triangles and squares, these shapes make humans think it is possible to invent something can truly perceive, learn and remember intelligent machine. The first generation of neural networks put basic principles in the Artificial intelligence fields[2].

Dream, 1969, Minsky[3] published a monograph on perceptron: single-layer perception, their feature layer is fixed, but these nets didn’t ability to solve XOR problem.

In 1986, Hinton et.al.[4] proposed the second generation of neural network; the original single feature is fixed. It consists of multiple hidden layers, and the activation function is sigmoid function. The back propagation algorithm is trained to reduce error and obtain stable weight, the model can effectively solve the nonlinear classification problem.

In 1989, Cybenko and Hornik et.al.[5-6] proposed finite linear combinations to prove the universal approach theorem called "universal peroxidation", the network contented three-layer neural network approximates with arbitrary precision.

In the same year, Lemony et.al.[7] invented convolutional neural network used to recognize handwriting Zip code digits, the model took three days for training.

In 1991, the back-propagation algorithm was pointed out to have the problem of gradient disappearance. In the last ten years, various propose shallow machine learning models, also invented by Curtes and Vapnik[8], in 1995 support vector machine invented the research of neural network using two groups' permanent data learning and reversible data learning. In 2006, Hinton et al. discussed graph model in the brain and proposed auto coder to reduce data dimension[9], Also they proposed deep belief network use pre-training to suppress the gradient disappearance problem training method quickly trains[10]

Bengio et.al.[11] proved greedy algorithm directed belief networks one layer at a time, also applicable to unsupervised such as self-encoders supervised learning. In 2011, Glorot et.al.[12] proposed the rectifying neurons model of biological neurons with ReLU activation letter number, it can effectively suppress the problem of gradient disappearance.

Microsoft and Google[13-14] is used deep learning in speech recognition, it successively adopted deep learning to reduce error rate from 20% to 30%, which is largest performance in this field in 10 years.

In 2012, Hinton and his students using Markov models (HMM) for temporal variability and reduced Top5 error rate of the ImageNet[15] picture classification problem from 26% to

15%[16]. Enter the outbreak period. Dauphin et.al.[17] in 2014 and Choromanska et.al.[18] in 2015 respectively proved that the local minimum problem is usually not a serious problem, eliminating the local extreme value haze over the neural network.

Deep learning is actually part of machine learning; it has gone through two modes from shallow machine learning to deep learning[19]. Deep learning model are important differences from shallow machine learning models. Shallow machine learning module features need to be artificially extracted not use distributed representations, also model itself only classifies or predicts based on the features, according to the test, the quality of the artificially extracted features largely determines the quality of the entire system[20], Feature extraction requires professional domain knowledge, and feature extraction, features the project takes a lot of time. Deep learning is a kind of representation learning[21], which can enough to learn the higher-level abstract representation of the data, which can be automatically, extracted from the data Features[22-23]. And the hidden layer in deep learning is equivalent to the input feature Linear combination of, the weight between the hidden layer and the input layer is equivalent to the input features Weights in linear combination[24]. In addition, the model capabilities of deep learning will vary with it increases exponentially with depth[25].

1.1 Multilayer Perceptron

Multi-layer perceptron (MLP) [2] or called forward propagation network and deep feed forward; it is most basic deep learning network structure. MLP is composed of several layers; each layer contains several neurons for each layer is activation function. The multi-layer perceptron with radial basis function is called radial basis network. The forward propagation of the multilayer perceptron is shown in Figure (2):

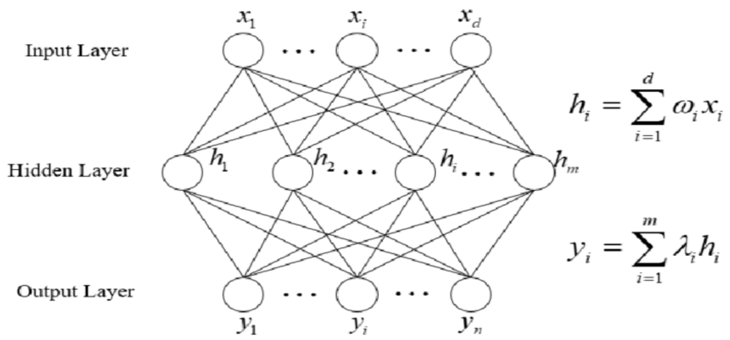

Figure (2)
The forward propagation of ML

The MLP forward propagation equation is shown in equations formula (1) and (2):

$$Z_i^{l+1} = \sum_j W_{ij}^l \; y_{ij}^l + b_i^l \qquad (1)$$

$$y_i^{l+1} = f\left(Z_i^{l+1}\right) \qquad (2)$$

Among them, y_i^j I s the output of j^{th} neuron in layer l, z_i^{l+1} is layer $l + 1$. The value of the i-th neuron before being activated by the activation function, W_{ij}^{l+1} is weights between of first layer j^{th} neuron and the i^{th} neuron in layer $l + 1$, bli Is offset, f (·) is a nonlinear activation function, common radial basis function, ReLU, PRELU, Tanh, Sigmoid, etc. If the function of mean square error is calculated as:

$$J = \frac{1}{2}\sum_i (y_i^l - y_i)^2 \qquad (3)$$

Among them, y_i^l Is the output of the i^{th} neuron in the last layer of the neural network, y_i is the true value of i-th neuron. The goal of neural network training is to minimize loss function; the optimization method usually uses the batch gradient descent method.

1.2 Convolutional Neural Network (CNN)

CNN[26] multilayer perceptron's usually mean fully connected networks, it suitable to process spatial domains in computer vision field. It can be used to process one-dimensional data product neural networks called "time-delay neural networks". The subject of CNN design ideas is inspired by visual neuroscience, it is mainly convolutional layer and pooling layer. Convolutional layer maintain to spatial continuity can extract the local features of image. While pooling layer use maximum pooling or average pooling, the pooling layer can reduce dimension of hidden layer and reduce the operations to other following layers volume to provide rotation invariance. The schematic diagram of convolution and pooling operations is shown in Figure 3:

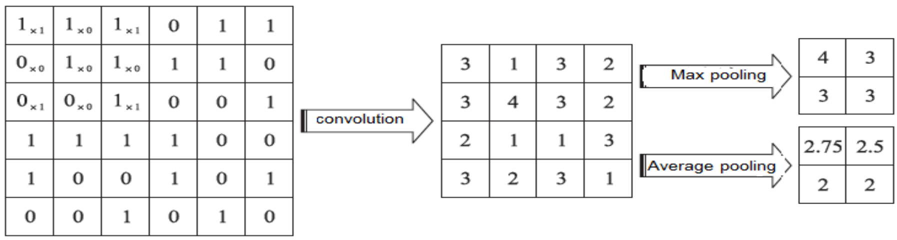

Figure (3)
Illustration Convolution and Pooling

As shown, the figure uses convolution kernel 3 × 3 and polling 2 × 2. The earliest convolutional neural network model was LeCun et.al.[26] in 1998 the proposed structure of LeNet-5 shown in Figure 4:

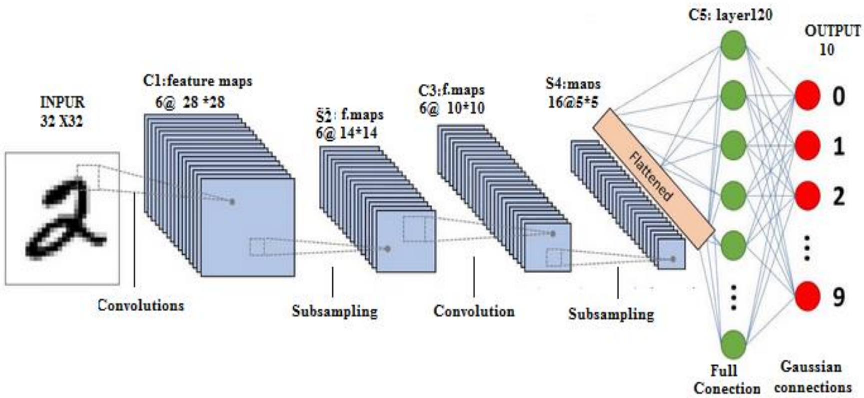

Figure (4)
The structure of LeNet -5

Input of MNIST diagram slice size is 32 × 32. After convolution operation from typical filter size 5 × 5. After pooling operation get image 28 × 28, then convolution and pooling get image 14 × 14. Finally get image 5 × 5 picture. The fully network connected layers of (120, 84, 10) neurons, the Soft-max function get probability of the numbers 0-9, and take the highest probability as the neural network prediction results. With convolution and pooling operations for higher network, the larger image is becoming smaller and smaller.

NN provides a hierarchical representation of visual data. The CNN weight of each layer is real, it learns some components of image, the higher layer is more specific components. CNN will undergo original signal layer-by-layer, it processes whole sequentially identify parts[27], the second layer of CNN can recognize Corners, edges and colors; third layer indented more complex transgender; the fourth layer can have identified specific parts of dog's face, bird's legs, etc. fifth layer can recognize specific objects such as keyboards. For example, for face recognition, CNN is first Identify points, edges, colors, corners, then eyes, lips, nose, and then whole face.

CNN is easy to achieve and accelerate on FPGA and other hardware[28]; CNN shares weights in same convolutional layer, which are weights of convolution kernel. CNN Characteristics such as connection, weight sharing, and pooling operations reduce model parameters[28], it reduces complexity of network and also provides invariance of translation, distortion, rotation and scaling.

1.3 Recurrent Neural Network

RNN is a type of ANN which connected between nodes form a directed graph along a temporal sequence.it can use their internal memory to process the inputs of variable length sequences [4], are suitable for processing time series data, it used widely in speech processing and natural language processing. RNN and its expansion diagram are shown in Figure (5);

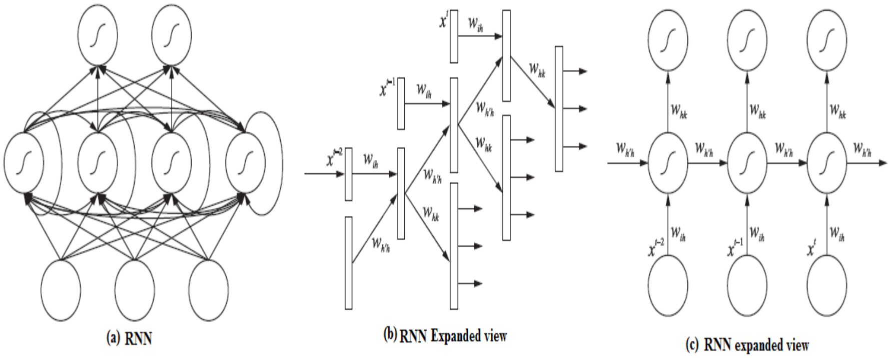

Figure (5)
RNN and its unfolding

The forward propagation formula of RNN is shown in equation (4-6):

$$z_h^l = \sum_{i=1}^{I} w_{ih}\, x_i^t \; + \sum_{h'=1}^{H} w_{h'h}\, a_h^{t-1} \qquad (4)$$

$$a_h^l = f_h\,(z_h^l) \qquad (5)$$

$$y_k^t = \sum_{h=1}^{h} w_{hk}\, a_h^t \qquad (6)$$

Among them, x_i^t is the i^th neuron of the input layer at time t, $a_{h'}^{t-1}$ is time t-1 The h' th neuron in the hidden layer, z_h^t is the h^{th} neuron in the hidden layer at time t, the value before the activation function, y_k^t is the k^{th} good in the output layer at time t. Jing Yuan, w_{ih}, $w_{'hh}$, and w_{hk} are input layer and hidden layer, hidden layer and hidden layer respectively. The weight between the hidden layer, hidden layer and output layer, $f_h(\cdot)$ is a nonlinear activation function.

In RNN hides layers, the output of hidden layer from previous moment as this moment The input of layer can use information of the past time, that is RNN shares weights at various times, which greatly reduces model parameters. However, the difficulty of RNN training is still relatively large, so Suskever et.al.[31].

2. Network structure improvement

2.1 Improvement of convolutional neural network

ImageNet [16] represent "ImageNet large scale visual recognition Competition, ILSVRC" has greatly promoted development for convolutional neural networks. The newly-invented convolutional neural network refreshed ImageNet performance from Krizhevsky[17] in 2012, ZFNet[27] in 2014, in 2014 VGGnet[28], and then GoogleLeNet[29] in 2015.The number of layers continues to increase and model capabilities continue to increase. The first exhibition of AlexNet Now with power of deep learning, ZFNet is a visual understanding of convolutional nerves as a result of network, VGGNet shows network depth can significantly improve deep learning. For the first time, GoogleLeNet broke model of convolutional layer pooling layer stacking, also reset successfully trained a neural network with a depth of 152 layers. The mainstream method of CNN applied to object detection is R-CNN[30] and later Improve FastR-CNN[31], and MaskR-CNN[32]. The improvement process is actually to replace a deep learning model instead of the shallow machine learning model literature[33] innovatively proposes idea of nesting networks; and spatial transformation network, the Improvement of effect models show as follow:

1) **AlexNet:** it called Convolutional Neural Network (CNN), designed by Alex Krizhevsky, it effectiveness of deep learning, invited first time by Hinton 2012 Participate in LSIVRC, it consists of five convolutional layers, a max-polling layer and droop layer, followed by three fully connected layers, final output layer has 1,000 Neurons. It enables corresponding to 1000 categories, are obtained after action of Soft-max function probability of each category. Ax-Net uses pan, flip, and intercept part of the picture to increase training data in other ways, use droop to prevent over fitting, use with a batch gradient descent method with momentum and weight decay to train the model. AlexNet Trained in parallel with two GPUs for 6 days, and used ReLU as the activation function [34].
2) **ZFNet:** ZFNet is support of ILSVRC 2013, the error rate is 11.2%, ZFNet can be considered as a fine-tuning of AlexNet, it has eight layers. Zeiler and Fergus use DE convolution network to visualize CNN understand role of each layer of CNN, visualization helps to find better than AlexNet If a better network structure ZFNet. ZFNet requires less training data, AlexNet uses 15 million pictures to train the model, while ZFNet only uses 1.3 million pictures. The first layer convolution kernel of AlexNet is 11 × 11, while ZFNet Is 7 × 7.

3) **VGGNet:** is produce from Simonyan et al. that development to AlexNet with successively increase the convolution layer, compare six different depth networks, and study the influence of network depth. The results show that deeper neural network is better effect. When it is increased to 16th and 19th layers, effect has improved significantly. The 19-layer network is called VGG-19. VGGnet strict Using a 3×3 convolution kernel; Using 2×2 max-polling, step size is 2. compared to ZFNet 7×7 Convolution kernel, the size of VGGNet convolution kernel is only 3×3, making model parameters fewer, and two consecutive convolutional layers make it have the effect of 7×7 convolution kernels, which Later people usually also use 3×3 convolution kernels. VGFNet - Caffe model is good effect when use image jitter to increase data training when classification image or object [33-34].
4) **GoogleNet:** is the winner of LSVRC2014, which has error rate is 6.7%, number of layers is 22. GoogleNet indicates to CNN with not stack convolutional layers and pooling layers in sequence. Google net using the Inception module, the convolution layer and pooling layer in the module are parallel, so is no need to choose whether to use convolutional layer or pooling layer. The total parameter amount of Google Net is only 1/12 of AlexNet. Good training when model classifies images, multiple deformed images have same probability and average of SoftMax.
5) **ResNet:** Deep Residual Network is produce from ILVSRC2015, its same network wins three categories: image classification, object positioning, and object detection. The error rate of image classification task is 3.57%, exceeding human error. The number of ResNet network layers reaches 152 or even 1,000 layers. Deep layer of ResNet network has suffered from disappearance of gradient, directly adding a linear communication between two or more layers path

constitutes a residual module to ensure that gradient can pass in linear path passes to bottom layer.

6) **R-CNN:** produce from Airsick et.al.[30]. Used to complete computer vision perceptual object detection task. Object detection goal convert all objects to improve training and testing speed while also increasing detection accuracy, this task is divided into two sub-tasks, first generated object frame, and other classified object. R-CNN adopts selective search method generate about 2000 boxes, using a trained CNN such as AlexNet used to feature extraction in each box, and then classification features into SVM perform and regression to get more accurate candidates Box.
7) **Fast R-CNN:** it combines three processes of CNN extraction, SVM classification, and regression in R-CNN form end-to-end overall model with improved speed and accuracy. Fast R - The input data of CNN is the whole picture and several boxes. **First,** several convolutional layers, **second** is pooling layers are used to process whole picture to obtain a feature map; use the region to process each box gets a fixed-size feature map.
8) **Faster R-CNN:** first uses convolutional layers and pooling whole image, it is processed by layer to obtain feature map. On this feature map, use “region of interest layer” to generate the box. That is, Faster-CNN replaces the method of generating boxes with deep learning modules. It changed from original generated on whole image to smaller features image. The model result is training speed is further accelerated.
9) **Mask R-CNN:** this model adds parallel branches of semantic segmentation to the basis of Faster R-CNN, also it adds segmentation tasks based on the original box generation, classification, and regression tasks, which can achieve object detection and semantic segmentation at the same time. The basic network of Mask R-CNN uses ResNeXt-101 and FPN (feature pyramid network)[35].

10) **(NIN) Network in Network:** It uses a reduced neural network such as a multi-layer perceptron to replace convolution kernel in CNN and forming a nested reduced structure in a neural network. the micro-network has been used for complex local modeling, the last fully connected layer in CNN can be replaced by global mean-pooling. This greatly reduces model parameters, prevents over fitting, and increases interpretability. NIN has 29 million parameters, which is 1/10 of AlexNE.

11) **(STNs) Spatial Transformer Networks:** it improves images accuracy by transforming the input instead of changing the network structure. STNs mainly contain spatial transformation modules, which are in turn composed of localization network, grid generator, and sampler. First, it uses the localized network to predict transformation of input image, and then grid generator and sampler transform on the image, the transformed images are put into CNN for classification. STNs are very robust and have spatial invariance such as translation, expansion, rotation, disturbance, and bending.

12) Other convolutional neural network improvements. In addition, there are other convolution gods After network improvement, including DE convolution network [36], stacked convolution auto-encoding[37], SRCNN[38], Over Feat[39], Floor- Net[40] …. etc.

2.2 Recurrent neural network improvement

Recurrent neural networks have problems of gradient disappearance or gradient explosion[41], and it cannot use long-time information in the past at long time, for example, when the activation function is Sigmoid, its derivative is a number less than 1, and multiple derivatives less than 1, This will cause the gradient to disappear. LSTM[42], a hierarchical RNN[43-44] are needles solution to this problem. RNN generally only one-dimensional time series data processing, etc. multi-dimensional RNN[45] was proposed to process image processing multidimensional Data. Also applied for natural language processing tasks[46], RNN has problems of complex training algorithm and large amount of calculation. Echo State Network[47] can achieve very high accuracy without repeatedly calculating gradients. Recurrent neural network lacks inference function, unable to complete require tasks reasoning, neural Turing machine[48] and memory network[49] to solve this problem by adding a memory module for the following:

1) Long-term short-term memory (LSTM): it replaced neurons with RNN unit. It contains Input gate, output gate, and legacy are added to the input, output, and forget the past information, respectively. LSM has two transmission states a cell state and hidden state, cell state is changes slowly with time, while hidden state changes at different times. LSM established a door mechanism to reach the input between old time input and new time input. The essence balance is to adjust the focus of memory according to the training goal and then enter line encoding. LSM can remember long-term memory; it can alleviate problem of gradient disappearance and gradient explosion in a longer sequence better performance than RNN.
2) Gated recurrent unit (GRU). GRU[50] is a variant type of LSTM, has only two gates" update gate and reset gate" with not output gate. The update gate determines how much information is kept in the past and how much information in input layer; the reset gate is similar to forget gate in LSTM. So, it always outputs the complete state. GRU uses fewer connections and fewer parameters in all places, so training is easier and faster.
3) Hierarchical RNN: it uses hierarchical structure for time dependence is a priori knowledge, long-term dependence can

be controlled by a variable. Hierarchical multi-scale RNN [44] capture all potential points sequentially by encoding time dependence in different time scales. Hierarchical RNN and hierarchical multi-scale RNN can alleviate problem of gradient disappearance.

4) Bidirectional RNN: it processes same sequence forward and backward, which is equivalent two RNNs, and same output layer. There is no hidden layer connected between forward and backward to avoid information loop such as natural language processing tasks, with the following words above will impact the word, two-way RNN will be able to use contextual information word to simultaneously capture two-way RNN information before and after the two-way LSTM.[51-52], Two-way GRU[53] has also proven to be very effective.

5) Multi-dimensional RNN (MRNN): MRNN is suitable for processing time series and other one-dimensional data. it can process multi-dimensional data, including images, videos, medical imaging, etc. The idea of MRNN is to convert multi-dimensional data into one in a certain order dimensional data. Of course, this sorting order needs to maintain the spatial continuity of multidimensional data.

6) Echo State Network (ESN): t turns the hidden layer into a storage pool. The input will be echoed in storage pool as if there is an echo, so network is called echo state network. ESN input layer from hidden layer to hidden layer, weight is fixed after random initialization. When the model is trained, only update weights of hidden layer to output layer becomes a linear regression problem, so the training Fast training speed.

7) Neural Turing Machines: its essence used an external storage matrix to interact with RNN similar overall system; both are using neural network method to establish "input paper to output map of paper tape". The controller networks through

parallel read/ write head changes value of external storage matrix to read/ write to memory.

8) Memory Networks it mainly contains memory unit and four modules: input, generation, output and response. The memory unit is actually is an array; each element of the array holds a sentence. The network can answer questions that require complex reasoning and implement automatic question answering[54]. The Input text of is encoded into a feature vector by the input module, and the generation module uses this vector read and write operations to memory unit, and then update memory. The output module is weighted and combined according to the degree of problem relevance to obtain output vector, and finally "answer module” generates the answer to question based on the output vector encoding.

3. Deep learning Applications

3.1 Voice Processing

Deep learning is first made breakthrough progress in field of speech processing[55], whether it was on a standard small data set[56] or a large data set [17]. There are two main tasks of speech processing: speech recognition and speech synthesis. Deep learning is widely used in speech recognition[57], Google[58] launches end-to-end speech recognition system, Baidu[59] launches speech recognition system deep Speech 2.16, Microsoft[60] accurate speech recognition on daily conversation data. The rate reached 5.9%, reached human level for first time. Major companies also use deep learning to achieve speech synthesis, including Google[61] and Apple[62]. Google DEEPMIND[63] proposed a parallel Wave Net model for speech synthesis, and Ping[64] introduced a product-level real-time speech synthesis system, deepVoice3.

3.2 Computer Vision

Deep learning is widely used in various tasks of computer vision, including traffic sign detection and classification[65], face recognition[66], face detection[67], image classification[32], multi-scale transforms fusion image[68], Object detection[36], image semantic segmentation[69], real-time multi-person pose estimation[70], pedestrian detection[71], scene recognition[72], object tracking[45], end-to-end video classification[73], Human motion recognition in video[74], etc. In addition, there are some interesting applications, such as automatically coloring black and white photos[75], turning graffiti into art painting[76], transferring artistic styles[77], and removing mosaics from pictures[78]. Oxford University and Google Depend Mind[79] also jointly proposed Lip _Net to read lips with an accuracy rate of 93%, far exceeding average level of 52% of humans.

3.3 Natural language processing

NEC Labs America[80] was first applied deep learning to natural language processing. At present, when dealing with natural language, usually word2vec[81] converts words into word vectors, which can be used as the characteristics of words[4]. Deep learning techniques are widely used in various tasks in natural language processing, including part-of-speech tagging[62], dependency parsing[82], naming knowledge[83], semantic role tagging[84], and distributed representations using only letters To learn language models[85], use letter-level input to predict word-level output[85], Twitter sentiment analysis[86], Chinese micro blog sentiment analysis[87], article classification[88], machine translation[89], Reading comprehension[90], dialogue system[91], and so on.

3.4 Others

In bioinformatics, deep learning can be used to predict the activity of drug molecules[92], predict where human eye stays[93], and predict effects of noncoding DNA mutations on gene expression and disease[94]. The financial industry has accumulated a large amount of data, so deep learning also has many applications in finance, including financial market forecast[95], securities investment portfolio[96], insurance loss forecast[97], etc., and a number of financial technology entrepreneurs have emerged accordingly the company. In addition, the real-time power generation scheduling algorithm based on deep learning[98] can reduce the total pollutant emissions of the unit under the premise of meeting the real-time power generation tasks, and achieve the goal of energy saving and emission reduction.

5. Conclusion

This study of deep neural networks .it took a deeper push through into well-known architectures and training algorithms. It focused to their shortcomings, e.g., getting stuck in the local minima, over fitting and training time for large problem sets. The study examined several ways of state-of-the-art to overcome these challenges with investigated adaptive learning rates and effective animalization parameters to develop the accuracy of network. The surveyed and reviewed several recent papers, studied them and given their application and improve method training process.

References

[1] McCulloch W S, Bittss W H. “A logical calculus of the ideas immanent in nervous activity”. Bulletin Mathematical Biologic, 1943, 5 (4): 115-133.

[2] Rosenblatt, F., “The perceptron: A probabilistic model for information storage and organization in the brain “Psychology Review, 1958, 65 (6): 386-392.

[3] Minsky M.L, Papert S.A. “Perceptron: An introduction to computational geometry “Cambridge, USA: Mitpress, 1969.

[4] Rumelhart, D. E., Hinton, G. E. & Williams, R. J “Learning representations by back propagating error”. Nature, 1986, 323 (6088): 533- 536.

[5] Cybenko G. “Approximation by superposition’s of a sigmoidal function “Mathematics of Control Signals and Systems, 1989, 2 (4): 303-314

[6] Hornik K., Stillcombe M., and Whit H. “Multi-layered feed forward networks are universal approximations “Neural Networks, 1989, 2 (5): 359-366.

[7] Lecony, Boser B., Denker J. S, et.al. ”Hand written digit recognition with a back propagation network“ Annual conference on Neural information Processing Systems,1989, 1 (4): 541-551.

[8] Cortes, C. and Vapnik, V. “Support-Vector Networks”. Machine Learning, 1995, 20(3), 273-297. http://dx.doi.org/10.1007/BF00994018.

[9] Hinton G E, Salakhutdinov R R. “Reducing the dimensionality of data with neural networks”. Science, 2006, 313(5786): 504–507.

[10] Hinton G. E., Osend S., and E.W” A fast learning algorithm for deep belief nets “Neural Computation, 2006, 18 (7): 1527-1554.

[11] Bengio Y., Lamblin P., Popovic D., et al.” Greedy layer wise training of deep network” Annual Conference on

Neural information Processing System. Cambridge, USA: MIT press, 2006: 153-160.

[12] Glorot X, Bordes A, Bengio Y. "Deep sparse rectifier neural networks[C]//International Conference on Artificial Intelligence and Statistics". Piscataway, NJ, USA: IEEE, 2011: 315-323

[13] Hinton G. E., Ding L., Y. D, et al." Deep Neural Network for acoustic modeling in speech recognition ". IEEE Signal Processing Magazine, 2012, 29 (6): 82-97.

[14] Dahl G.E., Yu D., Ding L., et al." Context depended pre-trained deep neural network for large Vocabulary Speech Recognition "IEEE Transaction on Audio Speech & Language Processing, 2011, 20 (1): 30-42.

[15] Ding J., Tang W., Socher R., et al. "ImageNet: A large–scale hierarchical image database" IEEE Conference on Computer Vision and Pattern Recognition Piscat away NJ USA: IEEE, 2009: 248-255.

[16] Krizhesky A., Sutschever I., Hinton G. E. Image Net classification with deep convolutional neural networks [C] // Annual Conference on Neural Information Foreign Promotions Systems. Cambridge, USA: MitPress, 2012: 107-110.

[17] Daphin Y.N., Pascanu R., Gullhere C., et al. Identifying and Attacking the Saddle Point Problem in High Dimensional non- Convex Optimization" Annual Conference on Neural Information Processing Systems Cambridge, USA: MitPress, 2014: 2933-2941.

[18] Choromanska A., Henaff M., Mathieu M., et al. The loss Surfaces of multilayer networks" International Conference on Artificial intelligence l. gene and Statistics. Piscataway, NJ, USA: IEEE, 2015: 1992-204.

[19] Yu Kai, Jia Lei, Chen Yuqiang, et.al. ”Deep learning yesterday, today and tomorrow“ Computer Research and Development, 2013, 50 (9): 1799-1804.
[20] Bengio Y, Delalleau O, Roux N L. “The curse of highly variable functions for local kernel machines“ Annual Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2006: 107-114
[21] Benggio Y., Courier A., Vincent P., “Representation Learning: A review and New Perspective“ IEEE Transactions on Pattern Analysis and Machine Intelligence 2013, 35 (8): 1798-1828.
[22] LeCun Y., Bengio Y., Hinton G.” Deep learning” Nature, 2015, 521 (7553): 436-4444.
[23] Goodfellow I., Benggio Y., Courville A. “Deep learning [M]. Cambridge, USA: MitPress, 2016.
[24] Bengio Y. “Leaning deep architectural for AI “Foundation & Trends in Machine Learning, 2009, 2 (1): 1-127.
[25] Montfar G.F., Pascan R., Cho K., et al. “On the Number of Leaner Regions of Deep Neural Network“ Annual Conference on Neural Information System, Cambridge, USA: MitPress, 2014: 2924-2932..
[26] LeCun Y., Boot L., Benggio Y., et al.,” Gradient-based learning applied to document recognition“ Proceedings IEE, 1998, 86 (11): 2278-2324.
[27] Zeiler M. D. and Fergus R. “Visualizing and Under Standing Convolution Networks“ European Conference on Computer Vision Cham, Switz Errand: Springer International Publishing AG, 2014: 818-833.
[28] Diceco R., Lacey G., Vasiljevic J., et al. “Caffeinated FPGA frame work for convolution neural network” International Conference Field-Programmable Technology Piscataway, NJ, USA: IEEE, 2017: 265-268.

[29] Szegedy C., Liu W., Jiay, et al.," Going deep Learning with Convolutions" IEEE Conference of Computer vesion and pattern Recognition, Piscat-Away, NJ, USA: IEEE, 2015: 1-9.

[30] Girsick R., Donahue J., Darrel T., et al." Rich Feature tree Hierarchies for Accurate Object Detection and Semantics Segmentation" IEEE conference computer vision and pattern recognition. Piscataway, NJ, USA: IEEE, 2014: 580-587.

[31] Girsick R. "FastR-CNN "IEEE Conference of Computer vision and pattern Recognition, Piscataway, NJ, USA: IEEE, 2015: 1440-1448.

[32] He K., Gkioxari G., Dollrpr P., et.al.,"MaskR-CNN" International Conference of computer vesion Piscataway, NJ, USA: IEEE, 2017: 280-1988.

[33] Lin M., Chen Q., Yan S.," Network in Network " Eprint Arxiv, 2013.

[34] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton," ImageNet Classification with Deep Convolutional Neural Networks", http://code.google.com/p/cuda-convnet/.

[35] Lin T.Y., Dolrpr P., Girsick R., et al.," Feature Pyramid Networs for Object Detection" Eprint Arxiv, 2016.

[36]Zeiler MD, Krishnan D, Taylor GW, et al. " Deconvolutional networks "//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2010: 2528-2535.

[37] Masci J, Meier U, Cireşan D, et al. "Stacked convolutional auto-encoders for hierarchical feature extraction "International Conference on Artificial Neural Networks. Berlin, Germany: Springer, 2011: 52-59. https://link.springer.com/chapter/10.1007%2F978-3-642-21735-7_7

[38] Dong C., Loy C.C., He.K., et al., "Learning Deep Convolutional Network for image Super-resolution"

European conference on Computer Vision. : Springer International Publishing AG, 2014: 184-1999.
[39] Sermanet P., Eagle D., Zing X., et al. "OverFeat : Integrated Recognition Localization and Detection Using Convolutional Networks " International Conference on Learning Representation . Piscataway, NJ, USA: IEEE, 2014. http://www. oalib.com/paper/4042258
[40] Dosovskiy A., Fisher P., Ig E., et al. "FloatNet: Leaning optical flow with Convolution Network" International conference on Computer Vision Control . Piscataway, NJ, USA: IEEE, 2015: 2758-2766.
[41] Benggio Y., Simard P., Frasconi P., "Learning Long –term Dependences with Gradient Descent is difficult" IEEE Transactions Neural Net- work, 1994, 5 (2): 157-1.
[42] Hochreiter S. and Schmidhuber J.,"Long Short-Term Memory" Neural Computation, 1997, 9 (8): 1735-1780.
[43] Hi S.E and Bengio Y. "Hierarchical Recurrent Neural Network for Long Term Dependence" Annual Conference on Neural Information Processing Cambridge, USA: MIT Press, 1995: 493-499.
[44] Chung J., Ahn S. and Bengio Y., " hierarchical multi-scale recurrent Neural Network " Eprint Arxiv, 2016.
[45] Graves A., Fernández S, Schmidhuber J. Multi-dimensional recurrent neural networks[C]//International Conference on Artificial Neural Networks. Berlin, Germany: Springer, 2007:549-558. https://link.springer.com/chapter /10.1007/978-3-540-74690-4_56
[46] Schuster M., and Palival K. K., "Bidirectional Recurrent Neural Network "IEEE Transactions Signal Procedures, 2002, 45 (11): 2673- 2681.
[47] Jaeger H., "The "echo stat" approach to analysis and Training Recurrent Neural Network with an erraturn note"

Bonn, Germany: German National Research Center for Information Technology . GMD Technical Report 2001.
[48] Graves A., Wayne G.,and Danihelka I., " Neural Turing machines " Eprint Arxiv, 2014.
[49] Weston J., Chopra S., and Bordes A., "Memory Networks" Eprint Arxiv, 2014.
[50] Chung J., Gullechre C., Cho K., et.al., "Empirical Evaluation of Gated Recurrent Neural Networks Sequence Modeling" Eprint Arxiv, 2014.
[51] Graves A., and Schmidhuber J. "Frame Wise Phoneme classification with Bidirectional "view of the basics of the western body of the bibliographical LSTM and Other Neural Network Architecture" Neural Networks the official of international Neural Network Society, 2005, 18 (5/6): 602-610.
[52] Thoreau T., Reczko M.," Bidirectional long short- term memory networks for predicting the Sub cellular localization of eukaryotic protein" IEEE, ACM Transactions Community Communications Biology & Bioinformatics, 2007, 4 (3): 441-446.
[53] Vuketic V, Rymond C, Gravier G.," A Step Beyond Local Observations with a Dialog await bidirectional GRU Network for Spoken Language " Conference in international Speech Special Community Association". Piscataway, USA: IEEE, 2016: 3241-3244.
[54] Weston J., Bordes A., Chopra S., et al. Toward ai-complete Question Answering: A set of Prerequisite toy task ". Eprint Arxiv, 2015.
[55] Vinyals O, Tosheva A, Binggio S, et al.,"Show and tell: A Neural Image Caption Generator "IEEE Conference Compute Vision and Pad-Tern Recognition. Piscataway, NJ, USA: IEEE, 2015: 3156-3164.

[56] Mohamed A.R., Dahl G.E., and Hinton G.," Acoustic Modeling Using Deep Belief Network "IEEE Transaction Audi, Speech, and Language Processing, 2012, 20 (1): 14-22.

[57] Graves A., Mohamed A.R., and Hinton G., "Speech recognition with Deep Recurrent Neural Network" IEEE Integrating Conference on Speech and SP. Piscataway, USA: IEEE, 2013: 6645-6649.

[58] Chiu C.C., Sinaith T.N., Wu Y., et al., "State –of- the- art Speech Recognition with Sequence to Sequence Model", Eprint Arxiv, 2017.

[59] Amadei D, Anantarayan S., Anubai R., et al., "Deep Speech 2: End-To-End Speech Recognition English and mandarin "International conference Machine Leaning. New York, NJ, USA: ACM, 2016: 173-182.

[60] Xiong W., Droppo J., Hang X., et al, " Achieving human parity conversational speech Recognition " Eprint Arxiv, 2016. The 275-281.

[61] Ze H, Senor A, and Schuster M.," Statistical parametric Speech synthesis using Deep Neural Network " IEEE International Conference A- Caustics,, Speech and SP. Piscataway, USA: IEEE, 2013: 7962-796.

[62] Capes T, Colors P, Conkie A, et al. Siri on –device Deep Learning Guided Unit Selection Text to Speech System " Conference of International Speech Communication Association. Piscataway, USA: IEEE, 2017: 4011-4015.

[63] Oord A.V, Li Y., Babuschkin I., et al.," Parallel WaveNet: Fast high-fidelity Speech Synthesis " Eprint Arxiv, 2017.

[64] Ping W, Ping K, Gibiansky A, et al. Deep voice3: 2000-speaker neural text-to-Speech". EprintArxiv, 2017.

[65] Zhu Z., Liang D., Zhang S., et al." Traffic Sign Detection and Classification in the Wild "IEEE Conference Compute

Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2016: 2110-2118.

[66] Hu G., Yang Y., Yi D., et al., “When Face Recognition Meets with Deep Learning: An Evaluation of Convolution Neural Network for Face Recognition “International conference of Computer Vision. Piscataway, NJ, USA: IEEE, 2015: 142-150.

[67] Yang S, Luo P, Loy C.C., et al.,” Fakeness-Net: Face Detection through Deep Facial Part Responses” “IEEE Trans on Pattern Analysis and Machine Intelligence, 2017, PP (99): 1-14.

[68] Lin S.Z., and Han. Z., “Images Fusion based on deep stack convolutional neural network” Chines journal of computers, 2017, 40 (11): 2506- 2518.

[69] Long J, Shellham E., and Darrel T., “Fully convolution network for semantic segmentation “IEEE Conference of Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2015: 3431-3440.

[70] Cao Z., Simon T., Wei S.E., et al.,” Real time multi-person 2D pose estimation using part affinity fields “, Eprint Arxiv, 2016.

[71] Tian Y., Lu P., Wang X., et al. “Deep learning strong pedestrian detection” International Conference of Computer Vision. Piscataway, NJ, USA: IEEE, 2015: 1904-1912.

[72] Zhou B., Lapriza A., Xiao J., et al.,” Leaning deep feature for sense recognition using places database “ Annual Concentration Environment Neural Information Processing Systems. Cambridge, USA: Mit Press, 2014: 487-495.

[73] Fernando B, and Gould S., “Learning end-to-end video classification with rank pooling “International Concentration on Machine Learning. New York, NJ, USA: ACM, 2016: 1187-1196.

[74] Lan Z., Zhu Y., Hauptmann A.G., et al.," Deep local video feature for action recognition "IEEE Conference Compute Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2017: 1219-1225.

[75] Ching Z., Yang Q., and Shen B., "Deep colorization "International conference of Computer Vision. Piscataway, NJ, USA: IEEE, 2015: 415-423.

[76] Champandard A.J.," Semantic style transform and training two-bit doodle into fine work "Eprint Arxiv, 2016.

[77] Gates L.A., Ecker A.S., Bethge M., "Image Style transfer using convolution neural network "IEEE Conference Compute Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2016: 2414-2423.

[78] Dahl R., Norouzi M., Schlens J.," Pixel recursive super resolution". Eprint Arxiv, 2017.

[79] Asal Y.M., Shillingfor B., Whitson S., et al.," LipNet: End-to-end sentence –level lip reading "sentence-level-liability [J]. Eprint Arxiv, 2016.

[80] Colbert R, Weston J, Boot L, et al.,' Natural language processing (almost) scratch" Journal of Machine Learning Research, 2011, 12 (1): 2493-2537.

[81] Mikolov T, Susketver I, Chen K, et al. Distributed representations of word and phrases and their compositional "Annual Conference on Neural Information Promotions Synchronization Systems. Cambridge, USA: Mit Press, 2013: 3111-3119.

[82] Weiss D, Alberti C, Collins M, et al. Structural tracing in the form of the trajectory of the railway-basica ty-based basis [J]. Eprint Arxiv, 2015.

[83] Lample G, Ballesteros M, Subramanian S, et al. "Neural architecture for named entity recognition "Eprint Arxiv, 2016.

[84] He L, Lee K, Lewis M, et al. “Deep semantic role labeling: What work and what’s next” Annual Meeting info Association Community- tactical Linguistics. New York, NJ, USA: ACM, 2017: 473-473.

[85] Kim Y, Jernite Y, Sontag D, et al. Character- aware neural language models “AAAI Conference Artificial Intelligence Actual Integration. Keystone, USA: AAAI, 2016: 2741-2747.

[86] Saveryn A. and Moschitti A. “Twitter sentiment analysis with deep convolution neural network “International Conference on Research Development Information Retrieval. New York, NJ, USA: ACM, 2015: 959-962.

[87] He Y. X., Niu F.F., et al. A deep learning model enhanced with emotion semantics for micro blog sentiment analysis”. Journal of Computers, 2017, 40 (4): 773-790.

[88] Joulin A, Grave E., Bojanowski P, et al. “Bag of tricks for efficient text classification” EprintArxiv, 2016.

[89] Gehrig J, Auli M, Granger D, et al., “Convolutional sequence to sequence learning” EprintArxiv, 2017.

[90] Herman K.M., Kosyky T., Greenset E, et al.” Teaching machines to read and comprehend” Annual Conference Neural Information Processing Systems. Cambridge, USA: MITPress, 2015: 1693-1701.

[91] Zhou X., Tang D., Wu H., et al. multi-view response essential for human-community conversions [C] // Conference Empirical Methods in Natural Language Processing. New York, NJ, USA: ACM, 2016: 372-381.

[92] Ma J, Sheridan R.P., Liaw A., et al.,” Deep neural nets as a method for quantitative structure “[J]. Journal of Chemical in formation & modeling, 2015, 55 (2): 263-274.

[93] Liu N., Han J., Zang D., et al. Predicting eye fixations using convolution neural network “IEEE Conference Compute

Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2015: 362-370.
[94] Leung M. K., Xiong H.Y., Lee L. J, et al. Deep learning of the tissue-regulated splicing code”- Bioinformatics, 2014, 30 (12): i121- i129.
[95] Dixon M, Klabj D, Bang J H. Classification based financial markets prediction using deep neural networks [J]. Eprint Arxiv, 2016.
[96] Heaton J. B., Polson N. G., and Winter J. H. Deep pointing format: Deep pointing format “Apple Studios Mode Modsin Business & Industry, 2016, 33 (1): 3-12.
[97] Zhang R., Li W., Tan W., et al. Deep and shallow model for insurance churn predication service “international Conference on Services computing. Piscataway, N J, USA: IEEE, 2017:346-353.
[98] Ke Y.Y, Yng X.Z, Xiong Y., et al.,”Power Generation dispatching for environmental protection based on recursive neural network and ant colony optimization algorithm“, Information and Control, 2017, 46 (4): 415-421.