International Journal of Hybrid Information Technology
Vol.8, No.11 (2015), pp.113-120
http://dx.doi.org/10.14257/ijhit.2015.8.11.09

Research on Feature Extraction based on Deep Learning

Wu Pin, Yan Hongjie, Shang Weilie, Zhu Yonghua and Gao Honghao

School of Computer Engineering and Science, Shanghai University, 200444, Shanghai, China
Institute of Materials Genome, Shanghai University, Shanghai, China
Computing Center, Shanghai University, 200444, Shanghai, China
wupin@shu.edu.com.cn

Abstract

With the development of deep learning, it has achieved impressive results in feature extraction field. This paper drives research in feature extraction based on deep learning. First, this paper gives a brief introduction on the world's research status on deep learning and principle of Restricted Boltzmann machine (RBM). Then this paper conducts reducing experiment based on RBM for handwritten digits. According to the analysis based on the results of the experiments, this paper tries to get a proper dimension which handwritten digits reduced to achieve better performance. Finally, this paper finds that it reach the goal when handwritten digits is reduced to half dimensional raw digits. This is an important foundation of deep learning layering and offers help to researchers in feature extraction based on deep learning.

Keywords: *deep learning, feature extraction, RBM*

1. Introduction

With the development of multimedia information, multimedia data has become an important way to record and share our daily life. But the high-dimension data made the processing such as data recognition and retrieval too time consuming. Fortunately, feature extraction can solve the issue which high-dimension data facing, using little but enough data to represent raw data to make recognizing and retrieving quickly.

Feature extraction means that new features represent raw data with linear or non-linear transformations, and the dimensions of low-dimensional space are always the dimensions of sample data's features [1].

Traditional feature extraction technology extract features depends on many experiences. In this case, feature directly influences the whole system. Manual feature designing are a way that need people's wisdom and experience. But if it can make itself automatically learn basic feature from sample without supervision or control, computer will be more intelligent than before.

And this has become a hot area of research. The essence of deep learning is to learn more useful feature through building models with many hidden layers and extremely large amounts of training data, to improve the accuracy of classing and forecasting. The essential characteristic of deep learning is layering learning. Layering learning involves the question that how to choose the dimensions of every layer. For one RBM layer, this issue becomes that choosing proper dimensions raw data reducing to. If choosing proper dimensions, in deep learning model, it can have a good initialization to give publicity to bring a multiplier effect on feature extraction.

This paper introduces world research statuses of deep learning and studies choosing the dimensions of raw data reducing to with RBM algorithm. And we want to raise awareness of feature extraction from researchers.

ISSN: 1738-9968 IJHIT

This paper is structured as follows. In Section 2, it gives a brief introduction on the world's research status on deep learning. In Section 3, principle of RBM is introduced. In Section 4, reducing experiment based on RBM for handwritten digits is done and analysis of experimental results is summarized. Lastly, the conclusions are obtained in Section 5.

2. World Research Statuses

In 2006, Prof. Hinton at the University of Toronto and his student Salakhutdinov presented deep learning model [2], attracting many people's attention. DARPA plan of US defense first funded deep learning project, participants including Stanford University, New York University and NEC American College, in 2010. Li and Dong experts in voice recognition from Microsoft Research cooperate with Hinton, who is expert in deep learning. And it changed technical framework of voice recognition based on deep learning completely in 2011 [3].

In November 2012, Microsoft brought an automated simultaneous interpretation system into the public at recent event in Tianjin, supported by the key technique, DNN or called deep learning [4]. In 2012, *New York Times* disclosed Google Brain project [5], using computer to recognize a cat based on deep learning successfully. Prof. With 16000 CPU Cores, "deep neural networks" machine learning model made a great success in voice recognition and image recognition, leaded by Andrew Ng, expert in machine learning from Stanford University, and JeffDean, expert in large-scale computer system. Hinton and his two students used a deeper CNN to get the best result in ImageNet issue in 2012, this would be a progressive step [6].

In the rear of 2012, Baidu used deep learning to solve the issues in natural image OCR and face recognition. In 2013, Baidu founded Deep Learning Research which the first Deep Learning Research [7]. In April 2013, MIT Technology Review listed deep learning in the first of 2013 breakthrough technology [8].

In addition, Hinton presented deep belief network (DBN) [9], which is consists of RBMs [10]. RBM is a neural net, which random generated based on input data set to learning probability distributions.

RBMs have been used as generative models of many different types of data for reducing the dimensionality [2], classification [11], collaborative filtering (CF), feature extraction [12] and topic models [13].

RBMs have two ways for training, including supervised learning and unsupervised learning, according to different tasks. RBM was named "Harmonium model" [14] by Paul Smolensky who invented it in 1985. It was developed by Boltzmann Machine (BM), which presented by Ackley and Hinton [15]. It became famous until Hinton and his college presented deep learning model in 2006. BM is essentially an energy model.

3. RBM

As mentioned above, DBN is consists of RBMs. And a simple RBM mode has 2 layers, including visible layer and hidden layer. Input raw data into visible layer with calculating, get the value of hidden layer. The RBM mode sketch is showed in the next page.

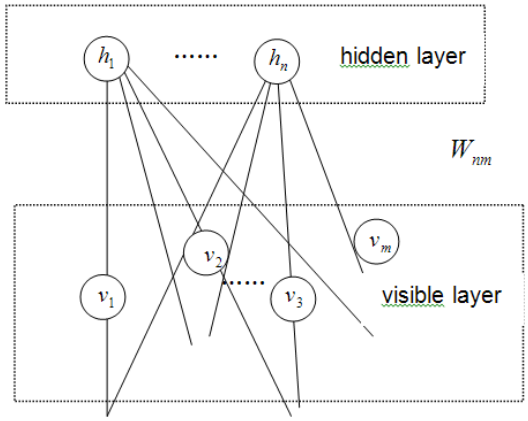

Figure 1. RBM mode sketch

In Figure 1, And this Figure also reveal that RBM's main function is used for dimension reduction.

RBM is a random network. Energy function is a good way to describe random networks. And RBM's energy function is showed below.

$$E(v,h) = -\sum_{i \in visible} a_i v_i - \sum_{j \in hidden} b_j h_j - \sum_{i,j} v_i h_j W_{ij} \quad (1)$$

In Function (1), h_j is the hidden variable. v_i is the visible variable. W_{ij} is the weight between hidden layer and visible layer. And a_i is the bias of visible layer, b_j is the bias of hidden layer.

Since we have the energy function of RBM, we can have the joint probability function between hidden layer h_j and visible layer v_i [16]:

$$p(v,h) = \frac{e^{-E(v,h)}}{\sum_{v,h} e^{-E(v,h)}} \quad (2)$$

As RBM is a bipartite graph, we can have condition probability function:

$$p(h_j = 1 \mid v) = \frac{1}{1 + e^{\sum_i v_i W_{ij} - a_i}} \quad (3)$$

$$p(v_i = 1 \mid h) = \frac{1}{1 + e^{-\sum_j W_{ij} h_j - b_j}} \quad (4)$$

4. Experiment

4.1. Experiment Environment and Conditions

Experiment environment:

Hardware environment includes a man computer with Inter Core 2 Duo CPI 2.89GHZ CPU and 8G Memory. Software includes Windows 7 sp1, Matlab 7.

Experiment conditions:

This experiment is based on the experiment of Hinton [17], and Data sets include The MNIST database of handwritten digits [18], which all in 28x28. And there are 60000 training patterns and 10000 test images.

4.2. Experiment Instruction

The ERROR function:

$$ERROR = \frac{1}{N}(v - h)^2 \quad (5)$$

In function (5), lower numbers equate to better performance, meaning the feature is close to pixels.

4.3. Experiment Process

An entire experiment process flowchart is showed in below.

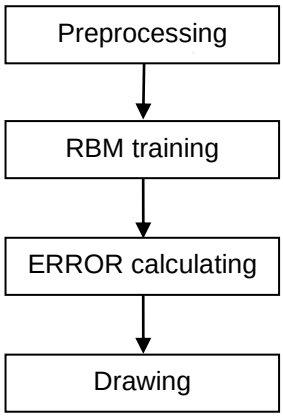

Figure 2. Entire Experiment Flowchart

In Figure 2, raw data is preprocessed to Matlab data firstly in the preprocess process, then using RBM to train for weight in the RBM training process. In the process of ERROR calculating, we calculate the differences between raw data and reconstruct data. After these, we draw the last 15 data from raw data and reconstructed data to compare in the Drawing process.

RBM training process is showed in the Figure below.

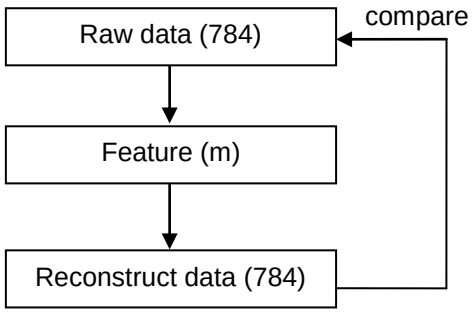

Figure 3. RBM Training Flowchart

Raw data is divided into n groups, each reducing to m using function (3), then use function (4) to get reconstruct data. Getting differences between raw data and reconstruct data to adjust weight. Repeat the previous steps.

4.4. Experiment Result

The reconstructed digits can reveal the performance of reducing process. The more similar raw handwritten digits and reconstructed digits shows the better the

reducing process performs. Raw handwritten digits and reconstructed digits are compared in the Figure below.

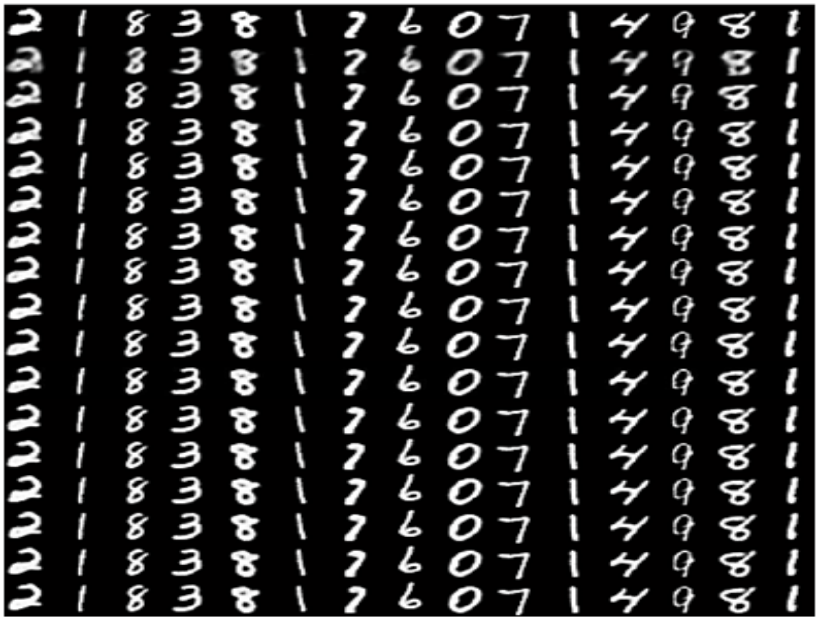

Figure 4. Handwritten Digits Figure

In Figure 4, first row are raw digits, second row are reconstructed digits from the data reduced to 49, and third row are reconstructed digits from the data reduced to 98 and so on. From Figure 4, we can find that the lower numbers of raw digits reduced to equate to more obscure, meaning worse performance, especially the second and third rows are extremely fuzzy. But when the number become big enough, the performance do not show obviously different.

There are two other criteria, the ERROR value and running time, for judging the performance of reducing. They are showed in table below.

Table 1. ERROR Value and Running Time

dimensions	Training ERROR	Teat ERROR	Running time
49	15.785	15.619	0:03:40
98	8.378	8.296	0:07:41
147	5.72	5.709	0:09:36
196	4.538	4.564	0:11:33
245	3.907	3.966	0:13:58
294	3.526	3.582	0:15:15
343	3.273	3.336	0:16:12
392	3.152	3.222	0:18:07
441	3.03	3.106	0:19:58
490	2.939	3.028	0:21:45
539	2.868	3.028	0:24:09
588	2.868	2.958	0:25:52
637	2.746	2.843	0:27:27
686	2.768	2.855	0:30:04
735	2.642	2.766	0:32:47

In Table 1, there are 60000 handwritten digits Figure for training and 10000 handwritten digits Figure for testing. Running time includes RBM training time, ERROR calculating time and drawing time.

Use test ERROR value to draw a graph for analysis which showed in the Figure below.

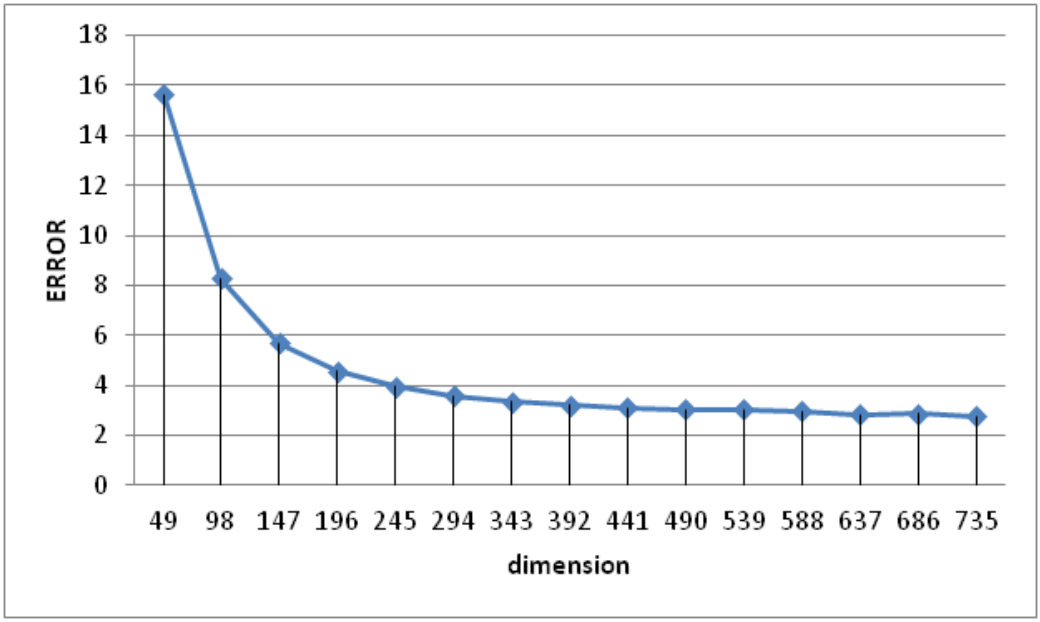

Figure 5. Test ERROR Graph

From Figure 5, we can see that the higher numbers raw data reduced to equate to lower number of test ERROR, meaning better performance. After the number raw data reduced to bigger than 392, curve of test ERROR become gentle.

Running time is showed in the Figure below.

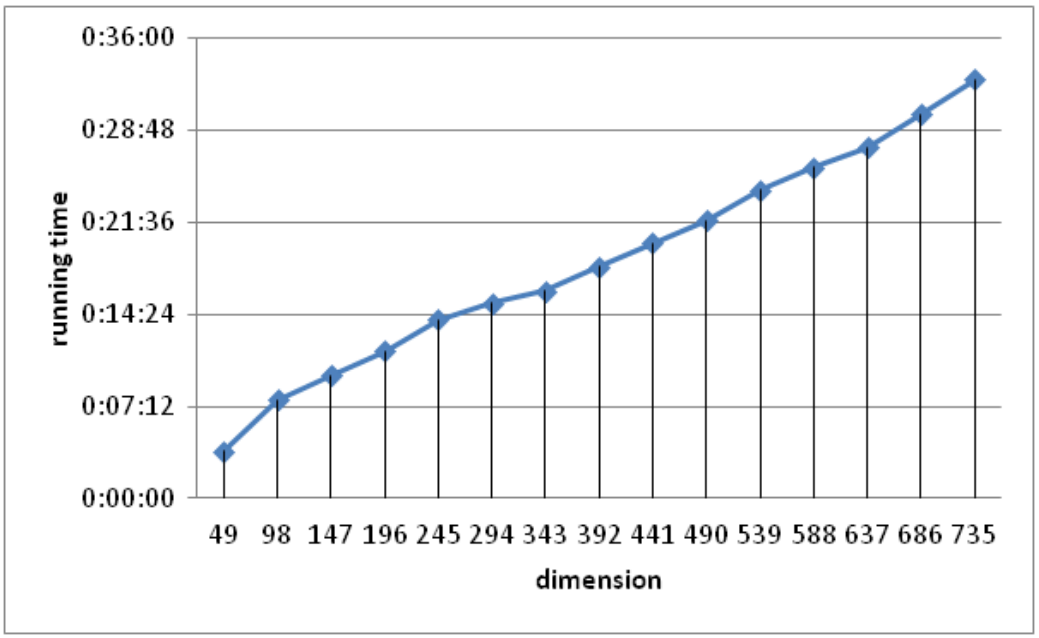

Figure 6. Running Time Graph

From Figure 6, we can see the running time almost linear increases. The higher number raw data reduced to equate to longer running time.

5. Conclusion

The results of the experiment indicate that lower number raw data reduced to equate to worse reconstructed digits, bigger ERROR value and shorter running time. In the contrary, the higher number result in smaller ERROR value and longer running time, but not better reconstructed digits.

In another word, get a proper number raw data reduced to could get the best performance. What is more, raw data reduce to $1/2$, reconstructed digits, ERROR value and running time can have a perfect performance.

Acknowledgments

This paper is supported by the research grant (No.14DZ2261200) from Shanghai Government and Shanghai Institute of Materials Genome, Foundation of Science and Technology Commission of Shanghai Municipality (14590500500), Young University Teachers Training Plan of Shanghai Municipality (ZZSD13008) and Natural Science Foundation of Shanghai (15ZR1415200) firstly, which supports for our research. In addition, we would like to show our deepest gratitude to Dr. Hinton for offering code and datasets. Then, we want to express our great gratitude to all the people who helped us during the writing of this paper. Finally, we should thank Shanghai University for the providing a good academic environment.

References

[1] P. Feng, Research on Feature Extraction and Feature Selection, (in Chinese), **(2011)**.
[2] E. G. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks", **(2006)**.
[3] G. E. Dahl, D. Yu, L. Deng and A. Acero, Context-Dependent Pre-Trained Deep Neural Networks for Large Vocabulary Speech Recognition, **(2012)**.
[4] J. Markoff, Scientists see promise in deep-learning programs, **(2012)**.
[5] J. Markoff, How many computers to identify a cat?, **(2012)**.
[6] A. Krizhevsky, I. Sutskever and G. E. Hinton, Imagenet classification with deep convolutional neural networks, **(2012)**.
[7] Y. Kai, J. Lei, C. Qiangyu and X. Wei, Deep Learning: Yesterday, Today and Tomorrow, (in Chinese), **(2013)**.
[8] 10 Breakthrough Technologies, **(2013)**.
[9] R. Sarikaya, G. E. Hinton and A. Deoras, Application of Deep Belief Networks for Natural Language Understanding, **(2014)**.
[10] A. Fischer and C. Igel, Training restricted Boltzmann machines: An introduction, **(2014)**.
[11] H. Larochelle and Y. Bengio, Classification using discriminative restricted Boltzmann machines, **(2008)**.
[12] A. Coates, A. Y. Ng and H. Lee, An analysis of single-layer networks in unsupervised feature learning, **(2011)**.
[13] G. E. Hinton and R. R. Salakhutdinov, Replicated softmax: an undirected topic model, **(2009)**.
[14] P. Smolensky, Information processing in dynamical systems: Foundations of harmony theory, **(1986)**.
[15] D. H. Ackley, G. E. Hinton and T. J. Sejnowski, A learning algorithm for boltzmann machines, **(1985)**.
[16] G. E. Hinton, A practical guide to training restricted Boltzmann machines, **(2010)**.
[17] "Hinton's Matlab code", is available at http://www.cs.toronto.edu/~hinton/MatlabForSciencePaper.html.
[18] "The MNIST data set", is available at http://yann.lecun.com/exdb/mnist/index.html.

Authors

Pin Wu, vice Professor, She received the B.S. and Ph.D. at Nanjing University of Science and Technology in 1998 and 2003. She had worked in Zhejiang University as a postdoctor for two years, and had worked in Michigan State University as a senior visiting scholar for one year. She is working in the School of Computer Engineering and Science of Shanghai University now. Her current research interests are focusing on high performance computing (HPC), computational fluid dynamics (CFD) and image processing. Over the past ten years, she has published over 30 technical papers in the related fields. She will keep on the research work addressing Cross-disciplinary of computer and mechanics.

Yonghua Zhu, vice Professor, He received the B.S. at Xi'an Jiaotong University in 1990, the M.S. at Tongji University in 1993 and Ph.D. at Shanghai University. He is working in the School of Computer Engineering and Science of Shanghai University. His current research interests are focusing on high performance computing, network computing, and interconnected network design, Communication and Information Engineering, Intelligent Controlling. Over the past ten years, he has published over 20 technical papers in the related fields. He will keep on the research work addressing Cross-disciplinary of communication, computer and automation.

Honghao Gao, received the Ph.D degree in the School of Computer Engineering and Science of Shanghai University, Shanghai, China, in 2012. His research interests include Web service and model checking.