Document Zone Content Classification for Technical Document Images Using Artificial Neural Networks and Support Vector Machines

Zaidah Ibrahim
Faculty of Comp & Math. Sci.
University Technology MARA
Malaysia
zaidah@tmsk.uitm.edu.my

Dino Isa, Rajprasad Rajkumar
School of Electrical Eng.
University of Nottingham
Malaysia Campus
Malaysia.
Dino.Isa@nottingham.edu.my
Rajprasad.Rajkumar@nottingham.edu.my

Graham Kendall
School of Computer Sci.
University of Nottingham
UK
gxk@cs.nott.ac.uk

Abstract

Artificial Neural Networks (ANN) are a classic pattern classifier and widely applicable to various problems and are relatively easy to use. Three of the most popular ANNs are Multilayer Perceptron (MLP) with Backpropagation learning algorithm, Self Organizing Map (SOM) and Recurrent Neural Network (RNN). Support Vector Machines (SVM) have gained great interest in the last few years in pattern recognition. Thus, this research compares the recognition performance of text and non-text images (text, table, figure and graph) from technical document images based on the pixel intensity of various zones between BPNN, SOM, RNN and SVM. Symmetrical and non-symmetrical zoning algorithms were compared as input. 400 different datasets have been tested and the experiments indicate that SVM classification is superior to the other three classifiers. The experiments also indicate that the combination of symmetrical and non-symmetrical zoning design is better than non-symmetrical or symmetrical zoning only.

Keywords : Backpropagation Neural Network, Self Organizing Map, Recurrent Neural Network, Support Vector Machine, Non-text Classification.

1. Introduction

A technical document usually consists of information that may be represented in different components such text, tables, graphs and figures. By segmenting and classifying these components separately, they could be input to search and retrieval systems. Research in automated document processing has been conducted for quite some time. Imade et. al. [1] conducted segmentation and classification of printed characters, handwritten characters, photographs and painted image regions based on the histograms of gradient vector directions and luminance levels. Rule-based classification based on physical block structures such as width, height and ratio width to height has been used by Shih et. al. [2] which classify the segmented blocks into text, horizontal/vertical lines, graphics and pictures. Etemad et. al. [3] classifies the segmented blocks into image, text and graphics based on moments of wavelets. Graphics, photographs and text were classified in [4] based on color, texture and shape.

This research applies zoning to extract features and act as input to four different classifiers, namely, a Multilayer Perceptron (MLP) MLP with back propagation learning algorithm, Self-Organizing Map (SOM), Recurrent Neural Network (RNN) and Support Vector Machine (SVM) as they have been applied to various classification applications. Zoning has been chosen since it seems to be one of the most effective approaches for feature extraction in pattern recognition [5]. Zoning is an example of statistical feature compared to structural features like strokes and their directions where an n x m grid is superimposed on the binary image and information for each n x m grid is calculated and is usually applied to alphanumeric character recognition [6], [7], [8].

The main objective of this research is to conduct a comparative study among BPNN, SOM, RNN and SVM to classify the text, tables, graphs and figures from technical document images based on zoning. The result of this classification, especially, figure classification, could be processed by the recognition module which is not discussed here. Figure 1 shows examples of the images used in this research that have been cropped manually from technical documents.

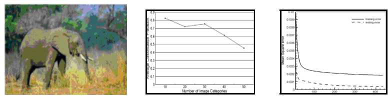


Figure 1 Sample data used for training and testing.

Artificial Neural Networks (ANN) are a classic pattern classifier and widely applicable to various problems and are relatively easy to use [9], [10]. In fact, current research in various applications is still being conducted using ANNs. The result in [11] indicates that BPNN outperforms Naïve Bayes and Nearest Neighbor algorithms in filtering the relevant information from digital documents. SOM has been applied in [12] to classify the bryophytes in the Tatra national park in Poland. A comparative study between BPNN and the Grow and Learn (GAL) algorithm has been performed in [13] and the experiment shows that both of these techniques provide similar result in classifying different types of heart sound. Authors in [14] compare BPNN with a Radial Basis Function Network (RBFN) in classifying defects on aluminum foil and the experiments indicate that BPNN classifies better than RBFN. River water quality classifications have been conducted in [15] using SOM, Cluster Analysis and Principal Component Analysis and the results show that SOM is better than the other two methods. RNN has been used in [16] for Arabic speech recognition while in [17] RNN has been used to detect the real extent of snow in mountainous regions. SVM has been used in [18] for hyperspectral image classification and [19] for classification of remote sensing classification.

A comparative study between SVM and FFNN has been conducted in [20] for fault detection and the results indicate that the performance of SVM is better than FFNN. Besides that, the training time is also less for SVM compared to FFNN. Comparable results were obtained in [21] between SVM and BPNN for credit rating analysis and in [22] for traffic speed prediction. Thus, the objective of this research is to investigate the potential of SVM in text and non-text zone classification compared to BPNN, RNN and SOM.

2. Zoning

Various zoning algorithms have been developed in the literature especially for character and digit recognition. Each zone of an image can be either symmetrical or non-symmetrical. A perceptual fuzzy-zoning algorithm has been developed in [23] where the image is divided into nine symmetrical zones but the boundaries of the zones are fuzzy. Then, the computation of the sum of the vector distance of all pixels from the origin for each cell is performed followed by the operation of vector distance normalization for each zone. Each classification has its own vector distance and can be shown in a feature vector graph. Authors in [24] use 8 x 8 symmetrical zones where the percentage of black pixels in each zone is calculated while in [25], the total number of black pixels is calculated separately for each line in the horizontal and vertical direction for each zone. Authors in [26] and [27] develop a non-symmetrical zoning, with very good results.

This research develops a combination of symmetrical and non-symmetrical perceptual zoning as shown in figure 2. Twenty different combinations have been tested where the total intensity of the black pixels in each zone is computed. Then, these values are entered into the ANNs and SVM to determine the optimal zoning design in classifying the images into text, tables, graphs and figures.

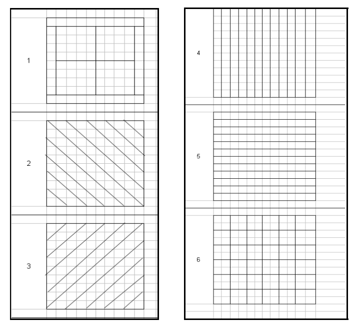


Figure 2 Different type of zoning design.

3. ANN and SVM

BPNN and RNN are two supervised neural networks which require training sets with input vectors and target vectors associated with the input vectors. The NN learner makes adjustments to the weight values to match as much as possible the desired target vector with the calculated target vector. A BPNN receives input signals and propagate these signals through all the layers to obtain the output of the NN. It is an example of a connectionist paradigm that relies on local computations to discover the information-processing capabilities of neural networks [28]. On the other hand, RNN has feedback connections from the output layer to the input layer. This feedback provides extra learning capability to the neural network.

Figure 3 illustrates the architecture of the BPNN and RNN used in this research. By referring to Figure 3, the dotted lines indicate the backward flow of the computation. A sigmoid function is used with one hidden layer. According to [28], the size of the hidden layer should be smaller than the size of the output layer. Thus, since the size of the output layer in this research is only four, the size of the hidden layer is less than or equal to four. Different sizes for the hidden layer and the number of epochs have been tested and the maximum classification rate is recorded.

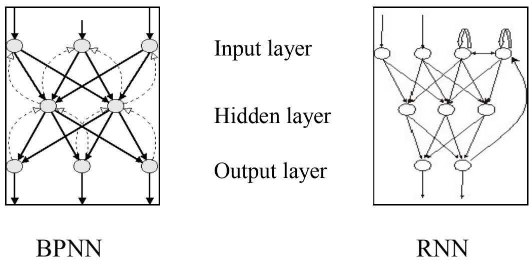


Figure 3 Architecture of BPNN and RNN

SOM is an unsupervised neural network that represents a feedforward structure with one computational layer with no target vectors required. It is based on competitive learning where the output nodes compete among themselves to be activated. It starts with input node initialization. For each input node, the discriminant function is computed which provide the basis for competition among the nodes. The winner of the competition is the node with the largest value of the discriminant function. The winning node determines the spatial location of the topological neighborhood of activated nodes which provides the basis for cooperation among the neighboring nodes. Then, the values of the discriminant function can be adjusted to produce better winning nodes [28]. Figure 4 illustrates the SOM topology.

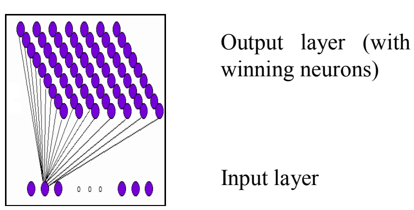


Figure 4 SOM topology

SVM is a supervised feedforward network that constructs a hyperplane as the decision surface where the margin of separation between positive and negative examples is maximized. It is an approximate implementation of the method of structural risk minimization [28]. The learning algorithm used in this research is the radial-basis function because it is the most common learning algorithm for SVM where the inner-product kernel between a "support vector" and the vector is drawn from the input space [28]. Figure 5 illustrates the architecture of SVM.

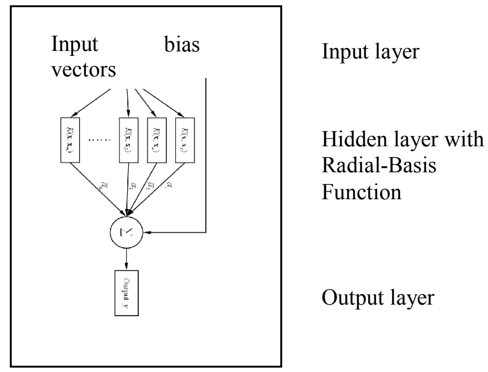


Figure 5 Architecture of SVM

4. Experimental results and conclusion

100 datasets from each category have been cropped from technical document images which totals up to 400 datasets, and 50% of the datasets have been used for training while the other 50% have been used for testing.

Figure 6 illustrates the zone content classification result for the text, tables, figures and graphs while Table 1 summarizes the best classification rate for all classifiers. By looking at Figure 6 we can see that SVM classification is better for all zoning designs with the best classification rate of 92.5%. SOM appears to classify the worst compared with the other two ANNs while RNN classifies better than BPNN. By looking at Table 1, we can see that the combination of the symmetrical (design 6 in Figure 2) and non-symmetrical zoning design (other designs in Figure 2) would be a better input feature compared to only using the symmetrical or non-symmetrical zoning design for this research.

Our future research will investigate extracting other features for zone classification and integrate SVM with RNN since both of these classifiers perform better than the other two classifiers. It is hoped that by integrating both the SVM and RNN, the classification result could be enhanced.

Table 1 Best classification rate for each classifier.

Classifier	Best Classification Rate	Zoning Design
BPNN	78	Design 2 & 6
SOM	75	Design 1
RNN	80.5	Design 2 & 4
SVM	**92.5**	Design 3 & 6

5. References

[1] Imade, S.; Tatsuta, S. and Wada, T. "Segmentation and Classification for Mixed Text/Image Documents Using Neural Network", Proc. International Conference on Document Analysis & Recognition (ICDAR1993), pp 930-934.

[2] Shih, F.Y. and Chan, S.S. "Adaptive Document Block Segmentation and Classification", IEEE Transactions on Systems, Man, and Cybernetics, vol. 26, no. 5, October 1996, pp. 797-802.

[3] Etemad, K,; Doerman, D.S. and Chellappa, R. "Multi-scale Segmentation of Unstructured Document Pages Using Soft Decision Integration", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 1, pp. 92096, Jan. 1997.

[4] Schettini, R.; Brambilla, C.; Ciocca, G.; Valsasna, A. And De Ponti, M. "A Hierarchical Classification Strategy For Digital Documents", Pattern Recognition 35 (2002), pp. 1759-1769.

[5] Impedovo, S.; Lucchese, M.G. and Pirlo, G. "Optimal Zoning Design by Genetic Algorithms", IEEE Trans. On Systems, Man, and Cybernetics, vol, 36, no. 5, September 2006.

[6] Trier, Q.D., Jain, A.K. and Taxt, T. "Feature Extraction Methods for character Recognition – A Survey", Pattern Recognition, Vol. 29, No. 4, pp.641-662, 1996.

[7] Blumenstein, M., Liu, X.Y. and Verma, B. "A Modified Direction Feature For Cursive Character Recognition", IEEE International Joint Conference on Neural Networks, Vol. 4, pp.2983-2987, 2004.

[8] Lecce, V.D., Dimauro, G., Guerriero, A., Impedovo, S., Pirlo, G. And Salzo, A. "Zoning Design for Hand-Written Numerical Recognition", Proc. Of the 7th International Workshop on Frontiers in Handwriting Recognition, 2000.

[9] Egmont-Petersen, M., de Ridder, D. and Handels, H. "Image Processing with Neural Networks – A Review", Pattern Recognition 35, pp. 2279-2301, 2002.

[10] Marinai, S.; Gori, M. and Soda, G. "Artificial Neural Networks for Document Analysis and Recognition", IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 27, no. 1, January 2005.

[11] Manevitz, L. and Yousef, M. "One-Class Document Classification via Neural Networks", Neurocomputing 70, pp. 1466-1481, 2007.

[12] Samecka-Cymerman, A., Stankiewicz, A., Kolon, K. And Kempers, A.J. "Self-Organizing Feature Map (Neural Networks) as a Tool in Classification of the Relations Between Chemical Composition of Aquatic Bryophytes and Types of Streambeds in the Tatra National Park in Poland", Chemosphere 67, pp. 954-960, 2007.

[13] Gupta, C.N., Palaniappan, R., Swaminathan, S. and Krishnan, S.M. "Neural Network Classification of Homomorphic Segmented Heart Sounds", Applied Soft Computing 7, pp. 286-297, 2007.

[14] Lin, S.W., Chou, S.Y. and Chen, S.C. "Irregular Shapes Classification by Back-Propagation Neural Networks", International Journal on Adv. Manuf. Technology, 34, pp. 1164-1172, 2007.

[15] Astel, A., Tsakovski, S., Barbieri, P. And Simeonov, V. "Comparison of Self-Organizing Maps Classification Approach with Cluster And Principal Component Analysis for Large Environmental Data Sets", Journal on Water Research , 41, pp. 4566-4578, 2007.

[16] El Choubassi, M.M.; El Khoury, H.E.; Alagha, c.E.J.; Skaf, J.A. and Al-Alaoui, M.A. "Arabic Speech Recognition Using Recurrent Neural Networks", Proc. of the 3rd IEEE International Symposium on signal Processing and Information Technology, pp 543-547, 2003.

[17] Simpson, J.J. and McIntire, T.J. "A Recurrent Neural Network Classifier for Improved Retrievals of Areal Extent of Snow Cover", IEEE Trans. On Geoscience and Remote Sensing, vol. 39, no. 10, October 2001.

[18] Guo, B.; Gunn, S.R.; Damper, R.I. and Nelson, J.D.b. "Customizing Kernel Functions for SVM-based Hyperspectral Image Classification", IEE Trans. On Image Processing, vol. 17, no. 4, April 2008.

[19] Ommen, T.; Misra, D.; Twarakavi, N.K.C.; Prakash, A.; Sahoo, B. and Bandopadhyay, "An Objective Analysis of Support Vector Machine Based Classification for Remote Sensing", Math Geosci (2008) 40: 409-424.

[20] Samanta, B.; Al-Balushi, K.R. and Al-Araimi, "Artificial Neural Networks and Support Vector Machines with Genetic Algorithm for Bearing Fault Detection", Engineering Applications of Artificial Intelligence 16 (2003), pp. 657-665.

[21] Huang, Z.; Chen, H.; Hsu, C.; Chen, W.; and Wu, S. "Credit Rating analysis with Support Vector Machines and

Neural Networks : A Market Comparative Study", Decision Support Systems 37 (2004), pp. 543-558.

[22] Vanajakshi, L. and Rilett, L.R. "A Comparison of the Performance of Artificial Neural Networks and Support Vector Machines for the Prediction of Traffic Speed", IEEE Intelligent Vehicles Symposium, Italy, 2004.

[23] Lajish, V.L. "Handwritten Character Recognition Using Perceptual Fuzzy-Zoning and Class Modular Neural Networks", IEEE 2008.

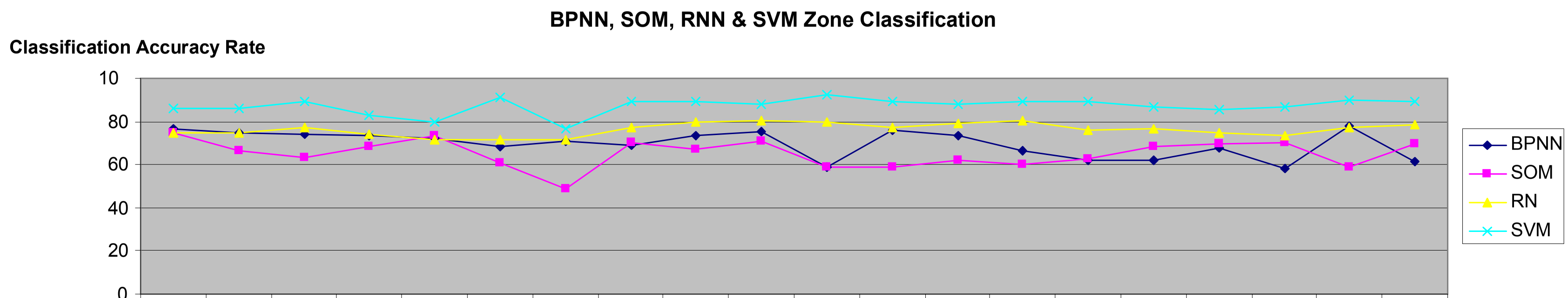
[24] Ross, P.M. and Pasero, E. "Artificial Neural Networks for Real Time Reader Devices", Proc. Of International Joint Conference on Neural Networks, Orlando, Florida, USA, August 12-17, 2007.

[25] Khawaja, A., Tingzhi, S., Memon, N.M. and Rajpar, a. "Recognition of Printed Chinese Characters by Using Neural Network", IEEE 2006.

[26] Freitas, C., Oliveira, L., Aires, S., and Bortolozzi, F. "Zoning and Metaclasses for Character Recognition", SAC'07, March, 11-15, 2007.

[27] Aires, S.B.K.; Freitas, C.O.A.; Bortolozzi, F. and Sabourn, R. "Perceptual Zoning for Handwritten Character Recognition", Advances in Pattern Recognition, 1998.

[28] Haykin, S. Neural Network : A Comprehensive Foundation, 2nd. Ed., Prentice Hall, 1999.



	8DifZone	SlantRight	SlantLeft	12Column	12Ro	36SameZone	SlantLeftSlan	SlantLeft8DifZ	SlantLeft12C	SlantLeft12R	SlantLeft36S	SlantRight36S	36SameZone	36SameZone	36SameZone	SlantRight8Dif	SlantRight12C	SlantRight12R	SlantRight36S	SlantLeftSlan
BPNN	76.5	74.5	74	73.5	72	68.5	70.75	69	73.5	75.5	59	76	73.5	66.5	62	62	68	58.5	78	61.5
SOM	75	66.5	63.5	68.5	73.5	60.5	48.5	70.5	67	71	59	59	62	60	62.5	68.5	69.5	70	59	69.5
JRNN	74.5	74.5	77	74	71.5	71.5	71.75	77.5	79.5	80.5	79.5	77.5	79	80.5	76	76.5	75	73.5	77.5	78.5
SVM	86	86	89	83	79.5	91	76.75	89	89	88	92.5	89	88	89.5	89	86.5	85.5	87	90	89.5

Zone Type

Figure 6 Testing result for BPNN, SOM, RNN and SVM for zone content classification.