Image Classification using Adaptive Multi-Module

Wonil Kim¹, Chuleui Hong², Soonil Kwon¹, Changmin Lee¹, Junghyun Kim¹ and Hanku Lee³

¹ Department of Digital Contents, Sejong University, Seoul, Korea
² Department of Computer Science, Sangmyung University, Seoul, Korea
³ Department of Internet and Multimedia Engineering, Konkuk University, Korea

Abstract: For a classification using neural network, there exist many cases in which the distributions of classes are so complex that the classification with single network does not properly differentiate the given data into classes. This problem can be resolved if we employ multiple modules that can classify different data respectively. This paper proposes a new adaptive architecture for classification problems, and simulates its performance on image classification that is not easily classified using traditional neural network learning algorithms. Classification modules are added as learning proceeds dynamically, depending on the data. When a new module is introduced, it is trained for classification of the remaining data on which the currently existing module does not perform classification well. Simulation results show that the proposed Adaptive Multi-module Classification Network (AMCN) achieves accuracy improvement. It also outperforms single module classification case when they are compared in the condition of equal weight updates.

Keywords: Image Classification, Multi-module, Neural Network, MPEG-7 visual descriptors, Adaptive

1 Introduction

Classifying data with salient features is a relatively easy task for most traditional neural networks. However, many applications such as performance evaluation and image recognition are characterized by extensive mixture in some parts of the input space. In this case, employing modular approach can be a much better solution. In fact, modularity promotes effective learning in neural networks, but it is not easy to determine how to modularize a network for a new task. Most adaptive algorithms successively insert (into a network) nodes but not modules, since it is generally impossible to predetermine the purpose of the module [1]. Most of the traditional modular algorithms require prior knowledge, and the number of modules for a given task is decided in advance [2, 3].

The best strategy is to test and evaluate with given structure. Adaptive algorithms can employ several modules adaptively according to the characteristics of the data; each module performs a classification task without requiring excessive training effort. New classification modules are generated only when the performance of the current module is inadequate; thus, unlike many other Modular Neural Network (MNN) models in which having prior knowledge of the input space is essential, no information is required in an adaptive system regarding the number of modules needed for a given problem.

We propose an Adaptive Multi-module Classification Network (AMCN) algorithm, which successively adds small modules to the network. It does not need to have a prior knowledge about the classification task. The number of modules is determined adaptively according to the complexity and distribution of class. The proposed algorithm trains one module after another, therefore one module may be trained extensively when the classification is hard and other modules can be
trained using a small number of iterations when the classification is easy to learn. Each module is trained on a subset of the training data on which existing modules do not perform well, so that the total computational effort in training is much less than that of traditional modular networks.

We simulated this architecture in portrait image classification problem. The problem is to classify a given image into one of six classes depending on age groups. Simulation results show that the proposed AMCN achieves accuracy improvement. It also outperforms single module classification case when they are compared in the condition of equal weight updates.

This paper consists of as followings. In the next chapter, we discuss adaptive classification networks, after which multi-module classification approach is mentioned. The proposed Adaptive Multi-module Classification Network is introduced in Chapter 4, which is followed by the simulation part in Chapter 5. The simulation is performed using portrait image data, which classifies images into two or six classes. Chapter 6 concludes.

2 Adaptive Classification Approaches

2.1 Nature of Adaptive Networks

Space is one fundamental dimension of the learning process; time is the other. The spatiotemporal nature of learning is exemplified in many learning tasks. Moreover, animals have an inherent capacity to represent the temporal structure of experience. Such a representation makes it possible for an animal to adapt its behavior to the temporal structure of an event in its behavioral space.

When a neural network operates in a stationary environment, the network can be trained by a supervised learning technique. Even in this stationary environment, the performance of supervised learning can be improved if the structure of a network can be decided adaptively during the network training session.

A network will undergo training sessions that modify the synaptic weights to match the environment. After the learning process, the network will capture the underlying statistical structure of the environment. Thus a learning system will rely on its memory to recall and exploit past experience. This process itself can be accomplished adaptively. Frequently, we do not have any information about the environment in advance, thus making the learning task hard to accomplish in a satisfactory manner. Instead of using a fixed structure, the algorithms described in this paper vary the number of network during training, adapting the whole network structure according to the given environment.

2.2 Adaptive Classification Neural Networks

Frean proposed an algorithm for developing a multilayer hierarchical network [4]. In this method, the network develops sub networks for misclassified samples. At each step, subnets are added that operate exclusively on the misclassified samples. The number of layers in the resulting network is at most logarithmic in the number of training samples. But, this approach is linear in a sense that only subnet is added one at a time, whereas the proposed approach differentiates the modules when it settles, not features. Therefore multiple classes with the same hidden feature may be classified into one class.

Sirat and Nadal also proposed a tree-structured network, but the nodes in the network resemble decision nodes in a decision tree instead of directly modifying the output of a higher level node [5]. In this case, one node examines only the training samples for which the parent node output is 1, while a sibling node is used to make the best decision for the training samples for which the parent node output is 0. The decision node discussed above is for explicit field, not features, possibly missing important hidden features.

3 Multi-module Classification Approach

3.1 Spatial and Temporal Crosstalk

Jacobs mentioned that there are two kinds of problems which motivate us to use a modular architecture [6]. One is spatial crosstalk which occurs when output units of a network provides conflicting error information to a hidden unit. He noted that this occurs when the back-propagation algorithm is applied to a single network containing a hidden unit that projects to two or more output units. Plaut and Hinton noted that if the network architecture has a separate module for each output unit, then it is immune to spatial crosstalk [7].

Another problem is that a unit might receive inconsistent training information at different times, which creates temporal crosstalk. For example, when a network is trained for one function, the network is specialized to that function. When the same network is trained for second (different) function which may be applicable in a different region of the data space, the network's performance for the first function will tend to be degraded.
According to Sutton, back-propagation tends to modify the weights of hidden neurons that have already developed useful properties [8]. Consequently, after being trained on the second task, the network is no longer able to perform the first task.

There are many more reasons why spatial and temporal crosstalk should be eliminated. The proposed adaptive modular networks are designed to tackle these problems, and successfully cope with both spatial and temporal crosstalk problems. It adds modules adaptively one by one, hence eliminates temporal crosstalk. Each module will perform the classification on different set of data hence eliminates spatial crosstalk.

The main problem in this case is how to select the proper modules or can it be done automatically. In this approach the module selection is done automatically according to the performance of the module on the given input vector.

### 3.2 Modular Classification Neural Networks

MNN have been studied in biological neural computations because of the modular nature of the human brain. Such an approach has been used successfully for behavior control in robotics, phoneme classification, character recognition, task decomposition, and piecewise control strategy. A modular network performs task decomposition in the sense that it learns to partition a task into two or more functionally independent tasks and allocates distinct network modules to learn each task. This property enhances the efficiency of learning a particular task. If there is a natural way of decomposing data into a set of simpler spaces where the classification task is easy, then a modular architecture should be able to learn the simpler classification faster than a single network can learn the original data. Another merit of modular networks is their generalizing ability. Also modular architectures are capable of developing more suitable representations that those developed by non-modular networks.

Jadhav presented a new approach for cardiac arrhythmia disease classification [9]. The proposed method uses MNN model. The experimental results showed that more than 82.22% testing classification accuracy may be obtained. Chang and Fu proposed a self-organizing map (SOM) neural network that selects more appropriate centers for radial basis function-based (RBF) network [10]. The experimental results show that the training time is faster than the traditional RBF neural network and improve the classification rate. Later, he also used SOM and learning vector quantization (LVQ) in which it can obtain more appropriate centers for the RBF neural network [11]. Principal component analysis (PCA) is applied to reduce the dimension of features.

Santos proposed the use of an entropic clustering algorithm as a way of performing task decomposition [12]. They presented experiments on several real world classification problems. Zhang investigate the effectiveness of a financial time-series forecasting strategy which exploits the multi resolution property of the wavelet transform [13]. In transformed space, each individual wavelet series is modeled by a separate multilayer perceptron (MLP). Ernst described an approach that enhanced the detection, tracking and fine analysis (classification of gender and facial expression) of faces [14]. Their benchmark results are given on standard and publicly available data sets.

Aminian proposed system that has the ability to identify faulty components or modules in an analog circuit by analyzing its impulse response [15]. In this approach, the circuit is divided into modules, which, in turn, are divided into smaller sub-modules successively. Lasch and Diem reported the applicability of an improved method of image segmentation of infrared micro spectroscopic data from histological specimens [16]. They found that the unsupervised methods of clustering, specifically agglomerative hierarchical clustering (AHC) were helpful in the initial phases of model generation. Stahl and Bramer introduce Prism algorithm and investigate its scaling behavior [17]. They describe their work to overcome limitations by developing a framework to parallel algorithms of the Prism family and similar algorithms.

But one major problem in modular networks is that we do not have any prior knowledge about the input space, hence the required number of modules for a task has to be guessed, and is often incorrect. The approach proposed in this paper explores an adaptive approach for module number optimization. It does not require any prior knowledge about the number of modules. Instead, the number of classification modules will be adaptively determined during the training process. The new module will be generated by weight update, in which it uses the data that the classification performs very well, hence reduces training time and error rates in relatively short time, and consequently improves overall network performance rate.
4 The Proposed Adaptive Multi-module Classification Network

The main idea of the Adaptive Multi-module Classification Network is that if training data are misclassified, then they are not used in the current training phase, instead they are used in training one of next modules later. Initially, each module is trained on a subset of the training data on which existing modules do not perform classification well. After which all the data will be classified rightly. The proposed AMCN algorithm is as follows in the box.

<AMCN Algorithms.>

**Step 1**: Let the module learn the classification for a while, say the weight change rate is below threshold. Then check the performance of classification for each data.

**Step 2**: If the data performs the right classification then include in the current classification module, otherwise ignore.

**Step 3**: Using the correct data, calculate the centroid and radius of the reference vector.

**Step 4**: Train the classification module using these correctly classified data until the performance rate exceeds the given threshold performance rate.

**Step 5**: Complete one classification module with the data centroid and radius. Remove these data from the training set and if there remains more than predefined ratio of data then do go to step 1 and continue. Otherwise stop and training task is over.

Therefore the module generation is adaptive according to the nature of training data. The centroid of each module is generated according to the means of corresponding data for each module. In the testing phase, one module will be selected according to the distance between the centroid of module and the given input data. A module corresponding with the closest centroid value to the given input will perform the classification task.

The main features of the AMCN algorithm are as follows: (1) The neural network consists of several classification modules, added adaptively. (2) Each module is very specialized, and hence can be trained very quickly. (3) Each module is associated with a subset of data which the classification task is easy.

Initially, each module is trained on a subset of the training data on which existing modules do not perform classification well. After which all the data will be classified rightly.

The algorithm consists of repeated introduction of new modules, until all the data are classified. When a new module is introduced, it is trained on the remaining input data on which currently existing modules do not perform well enough. In general, data corresponding to different classes (to be trained using different modules) may not be clearly separable, and classification is not informative for the classification task. The structure of AMAN is depicted in “Figure 1”.

5 AMCN Simulation on Portrait Image Classification

5.1 Portrait Image Recognition

We applied AMCN to portrait image classification problem. Portrait image recognition technology has been researched for various application domains. It is used by security applications which recognizes facial image as identification, by image classification applications, and by machine vision applications.

Das and Loui described their investigation on a system for making automatic album of consumer photographs, which uses face-based information such as age and gender extracted from images [18]. Age and gender classifier is the core component of the system and based on individual facial feature measurements using AdaBoost algorithm. In this system, faces are clustered by similarity to produce
clusters which represents frequently occurring individuals in the image set.

Automatic facial action detection system, based on changes in contours of facial components and face profile contours, was proposed by Pantic and Rothkrantz [19]. They described an automated system to recognize facial gestures in static, frontal and profile view color face images. A multi-detector approach to facial feature localization was utilized to spatially sample the profile contour, and the contours of the facial components such as the eyes and the mouth. From the extracted features, the system recognizes facial muscle actions occurring alone or in combination using rule-based reasoning. Age and gender estimation is the one of key issues to classify facial images. Hayashi et al. presented age estimation technique for image processing, in which they extracted wrinkle from given image for age and gender estimation [20]. They also extracted a skin region by using color image, and then made a histogram from extracted skin image to emphasize wrinkles.

Liu and Wechsler presented their novel Gabor-Fisher Classifier (GFC) for face recognition [21]. The method was applied to the Enhanced Fisher linear discriminant Model (EFM) to an augmented Gabor feature vector derived from the Gabor wavelet representation of face image. They also presented comparative experimental studies of various facial recognition schemes, the Eigenfaces method, the Fisherfaces method, the EFM method, and the combination of Gabor and Fisherfaces method. Gutta and Wechsler previously presented a face recognition system using hybrid classifier in reference [22].

Lighting condition and complex background are negative factors to recognize human face images. Hsu et al. propose a face detection algorithm for color images based on a novel lighting compensation technique and a non linear color transformation [23]. The method detects skin regions over the entire image and then generates face candidates based on spatial arrangement of these skin patches. The algorithm constructs eye, mouth, and boundary maps for verifying each face candidate. However, there are not many studies on face recognition schemes which use adaptive modular approaches yet.

5.2 Portrait Image Simulation

In this subsection, the portrait image classification system using the proposed AMAN algorithm will be presented. The input values are extracted using MPEG-7 visual descriptors. The classification network uses back-propagation to update weight values.

The features of training images are extracted in XML format using the MPEG-7 XM program. This feature information in XML format is parsed in the next step and is normalized into values between 0 and 1 with respect to values generated by each descriptor. These normalized values are used as inputs for the neural network classifier. The original values are converted into normal values, and followed by the class information. The class information, which is attached to the feature value, is the orthogonal vector value. For example, a category one gender image is represented as (1 0), whereas category two gender is (0 1).

The classification module employs neural network. The neural network classifier learns a relation of the feature values and a corresponding class by modifying the weight values between nodes. We use the back-propagation algorithm to train the network. It consists of input layer, output layer, and multiple hidden layers. The number of input nodes depends on a dimension of each descriptor, whereas the number of output nodes two or six depending on the categories (gender, age group). In a testing process, similar to the training process, the system extracts features from query images using MPEG-7 descriptors and classifies the images using the neural network that is generated by the training process.

For the input values extracted using MPEG-7 visual descriptors, we choose Edge Histogram and Homogeneous Texture descriptors for testing, since for age classification edge and texture are important factors. The “Figure 2” shows the overview of this simulation process. We performed gender and age group classifications. The age group is divided as follows; baby, toddler, teen, young adult, middle age, and seniors.
For gender recognition simulation, 400 images were used for training (200 for each image case), and 140 for testing (70 for each image case). For age group simulation, the same 600 images were used for training (100 for each age category), and another 240 for testing (40 for each image case). In both cases two descriptors were employed for feature values, Homogeneous Texture and Edge Histogram. In all classification cases of simulations, the traditional one-module classification networks were equipped with two hidden layers with 50 nodes each and were trained 100,000 iterations. For the proposed AMCN, the total iteration number for all the generated modules was reduced in half and the structure stayed intact. That means half of the weight updates needed for the same or better results. For both gender and age classifications, AMCN produced multiple numbers of modules depending on the parameters. When the data region size for corresponding module is smaller, the number of produced module becomes larger.

The gender classification presents good result both in Homogeneous Texture and Edge Histogram. In all Homogeneous Texture and Edge Histogram descriptor simulations, both one-module classification network and AMCN successfully classify portrait images over 80% of accuracy in average. In case of AMCN, the performance is better than the one module approach by 5% to 7%. Since it is stabilized in half iterations, the number of computation is much lower than the traditional methods. In both cases, male classification rates are better than female classification rates.

The age group classification shows relatively lower results than the gender image classification since it deals with six categories of images. For both one module and AMCN cases, it ranges from 70% for baby to 50% for teenager. It is reasoned that the age bear very distinct image features. Generally, the system performs good result in Edge Histogram descriptor and produces worse result in Homogeneous Texture descriptor. It justifies that human facial status degrades depending on age as Edge of an image is a salient feature for wrinkles. Same as in the previous gender classification, the proposed AMCN outperforms one module network in correct rate of classification. As in the previous case, we trained half of the iteration used in the traditional method and outperformed the results. These results seem very promising and can be applied to various image processing domains. It can be applied and implemented as the main part of image search engine or image collection engine.

For a large image data base, it is very useful tool for image retrieval system.

6 Conclusion and Future Works

In this paper, we propose a novel approach in adaptive classification network. The network adds classification modules according to the performance of the current module. Therefore the number of modules are not predetermined and decided according to the nature of problem. Simulation results show that the proposed Adaptive Multi-module Classification Network (AMCN) achieves accuracy improvement. It also outperforms single module classification case when they are compared in the condition of equal weight updates.

Acknowledgements

This research was supported by the MKE(Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency)“(NIPA-2012-H0301-12-3006)

References


[8] R. S. Sutton, Two problems with back-propagation and


Wonil Kim received the B.E in Metal Engineering from Hanyang University, Seoul, Korea in 1982. He worked for Korean Air from 1981 to 1985 as System designer and programmer. He received the B.S., M.S in Computer Science from Southern Illinois University, U.S.A. in 1988, 1990 respectively. He received Ph.D. in Computer and Information Science from Syracuse University, U.S.A. in 2000. From 2000 to 2001, he worked for Bhasha INC, U.S.A. as technical and research staff. He was with Ajou University, Suwon, Korea from 2002 to 2003. Since 2003, he has been with Department of Digital Content, College of Electronics and Information, Sejong University, Seoul, Korea. His research interests include artificial intelligence, multimedia contents, bioinformatics, adaptive systems, and computer security.

Chuleui Hong received B.S. degree in Hanyang University, Seoul, Korea in 1985, and his M.S. degree and Ph.D. degree in Computer Science at New Jersey Institute of Technology, USA and University of Missouri-Rolla, USA in 1989 and 1992, respectively. He was a senior researcher in Electronic and Telecommunications Research Institute, Korea, from 1992 to 1997. He joined Sangmyung University as a faculty member in 1997 and is currently professor of Computer Science Department.
at Sangmyung University, Seoul, Korea. His research interests include parallel and distributed system, optimization algorithm, multimedia application, and intelligent agent.

Soonil Kwon received the PhD degree from University of Southern California. He is currently an assistant professor in Sejong University. His research interests are in the areas of HCI, especially on speech/audio interface.

Changmin Lee received the B.E in Digital Contents from Sejong University, Seoul, Korea. He is currently the master’s course student of Digital Contents at Sejong University and working with Professor Wonil Kim in intelligent systems lab at Sejong University. His main research interests include artificial intelligence, intelligent agent.

Junghyun Kim received the B.E in Computer Engineering from Sejong University. He is currently the master’s course student of Computer Engineering at Sejong University, Korea and working with Professor Wonil Kim in intelligent systems lab at Sejong University. His main research interests include e-Learning system, artificial intelligence, mobile computing and multimedia contents.

Hanku Lee is the director of the Social Media Cloud Computing Research Center and an associate professor of the division of Internet and Multimedia Engineering at Konkuk University, Seoul, Korea. He received his Ph.D. degree in computer science at the Florida State University, USA. His recent research interests are in cloud computing, distributed real-time systems, distributed computing, and compilers.