Classification of the Images of Gene Expression Patterns Using Neural Networks Based on Multi-valued Neurons

Igor Aizenberg¹, Ekaterina Myasnikova² and Maria Samsonova²

¹ Neural Networks Technologies (NNT) Ltd., 155 Bialik st., Ramat-Gan, 52523 Israel, igora@nnt-group.com; igora@netvision.net.il (IA)
² Institute of High Performance Computing and Data Bases, 118 Fontanka emb., St.Petersburg, 198005 Russia, myasnikova@fn.csa.ru (EM), samson@fn.csa.ru (MS),

Abstract. Multi-valued neurons (MVN) are the neural processing elements with complex-valued weights and high functionality. It is possible to implement an arbitrary mapping described by partial-defined multiple-valued function on the single MVN. The MVN-based neural networks are applied to temporal classification of images of gene expression patterns, obtained by confocal scanning microscopy.

1 Introduction

Ability of neural networks to accumulate knowledge about objects and processes using learning algorithms makes their application in pattern recognition very promising and attractive [1]. In particular, different kinds of neural networks are successfully used for solving the image recognition problem [2].

Neural networks based on multi-valued neurons have been introduced in [3] and then developed in [4-9]. A comprehensive observation of multi-valued neurons theory, their learning and applications is given in [6]. Multi-valued neural element (MVN) is based on the ideas of multiple-valued threshold logic [6]. Its main properties are ability to implement arbitrary mapping between inputs and output described by partially defined multiple-valued function, quickly converging learning algorithms based on simple linear learning rules and complex-valued internal arithmetic.

Several kinds of MVN-based neural networks have been proposed for solving the image recognition problems. Different models of associative memory have been considered in [3, 4, 6, 7, 8]. An approach to image recognition, which will be used here, has been introduced in [5] and then also considered and developed in [6, 9]. This approach is based on the following. Since it is always difficult to present the image description, which then could be used for the learning, in some formal way, a nice solution is objectification of the image presentation using some objective procedure. Jump from the image representation in spatial domain to the representation in frequency domain is a good way to this objectification. Nature of this data presentation is clear: since in frequency domain the signal energy is concentrated in a small number of the low frequency part spectral coefficients, it makes possible to use exactly these coefficients as objective description of the signal. Taking into account
that multi-valued neuron operates with the complex-valued data, it is natural to use Fourier transform basis for decorrelation of the images that have to be recognized.

In our study MVN-based neural networks are applied for classifying the objects, which vary continuously over time. A peculiarity of this type of data, which makes their classification especially difficult, is the impossibility to subdivide the dataset unambiguously into a certain number of discrete well-defined classes. Here we proceed from the preliminary classification, which was performed manually on the basis of visual inspection of objects and hence was somewhat arbitrary. A temporal class was operationally defined as a set of objects indistinguishable to a human observer, but that in turn suggested that assigning the borders of these classes was also quite arbitrary. Therefore we cannot ever say to what extent the recognition results reflect accuracy of the automatic classification, the hand classification, or both.

In this paper we consider the dataset of images of gene expression patterns, obtained by confocal scanning microscopy [10]. This is a new promising approach for acquisition of quantitative data on gene expression at the resolution of a single cell. Gene expression data are of crucial importance for elucidation of mechanisms of cell functioning, as well as for the early diagnosis of many diseases.

2 Description of the Data

We perform temporal classification of images of genes expression patterns controlling segmentation in the fruit fly *Drosophila*, which is a model organism for molecular biology studies. Like all other insects, the body of the *Drosophila* is made up of repeated units called segments. During the process of segment determination a fly embryo consists of a roughly prolate spheroid of about 5000 nuclei. Genes that act to determine segments are expressed in patterns that become more spatially refined over time. One can view each gene's expression pattern as a collection of “domains” (stripes), each of which is a region of expression containing one maximum (Fig 1). In the experiments gene expression was recorded by confocal scanning of embryos stained with fluorescence tagged antibodies. The obtained images were subjected to image segmentation procedure to obtain the data in terms of nuclear location [11] and then rescaled to remove a nonspecific background signal. In the processed image the nuclei are presented by single pixels with a fluorescence intensity proportional to the average value of gene expression in the respective nucleus.

Human observers classify the developmental stage of an embryo by careful study of its pattern, since each stripe possesses its own features at any stage of an embryo development. In such a way 809 embryos were subdivided into 8 temporal classes [12] (their representatives are shown in Fig.1). Each embryo was allocated to one of the temporal classes on the basis of thorough and extensive visual inspection of the expression pattern of the *eve* gene, which is highly dynamic. We selected embryos for scanning without regard for age, so we expect our dataset to be uniformly distributed in time. The 8 classes were approximately equally populated.

The evolution of the *eve* expression patterns during the segment determination are presented in Fig.1. Time classes 1, 2, and 3 do not have seven well-defined stripes and the number and location of stripes changes rapidly.