

Chapter 82

Intelligent Classifiers Fusion for Enhancing Recognition of Genes and Protein Pattern of Hereditary Diseases

Parthasarathy Subhasini

*Avinashilingam Deemed University for Women,
Thadagam Post, India*

Dieter Roller

Universität Stuttgart, Germany

Bernadetta Kwintiana Ane

Universität Stuttgart, Germany

Marimuthu Krishnaveni

*Avinashilingam Deemed University for Women,
Thadagam Post, India*

ABSTRACT

Most the objective of intelligent systems is to create a model, which given a minimum amount of input data or information, is able to produce reliable recognition rates and correct decisions. In the application, when an individual classifier has reached its limit and, at the same time, it is hard to develop a better one, the solution might only be to combine the existing well performing classifiers. Combination of multiple classifier decisions is a powerful method for increasing classification rates in difficult pattern recognition problems. To achieve better recognition rates, it has been found that in many applications, it is better to fuse multiple relatively simple classifiers than to build a single sophisticated classifier. Such classifiers fusion seems to be worth applying in terms of uncertainty reduction. Different individual classifiers performing on different data would produce different errors. Assuming that all individual methods perform well, intelligent combination of multiple experts would reduce overall classification error and as consequence increase correct outputs. To date, content interpretation still remains as a highly complex task which requires many features to be fused. However, the fusion mechanism can be done at different levels of the classification. The fusion process can be carried out on three levels of abstraction closely connected with the flow of the classification process, i.e. data level fusion, feature level fusion, and classifier fusion. The work presented in this chapter focuses on the fusion of classifier outputs for intelligent models.

INTRODUCTION

To date the computational methods and computer-based systems deployed in science, engineering, industry, business and many other aspects of life has become powerful tools for problem solving, particularly when human beings have to cope with such variety of data at such a rate that precludes human analysis. Prime examples are found in medical and bioinformatics sciences. Currently, with the advancement of soft-computing methods and parallel computing systems it is possible to identify some diseases in humans by sequencing and recognition of genes and proteins pattern.

Generally speaking, genetic disorder or hereditary diseases is a result of mutations. There are two primary pathways to genetic defects. First, genetic disorders caused by the abnormal number of chromosomes, e.g., in Down syndrome there are three instead of two “number 21” chromosomes, therefore a total of 47. Second, triplet expansion repeat mutations caused by modification of gene expression or gain of function respectively, e.g., fragile X-syndrome and Huntington’s diseases (Mehta, 2007). Defective genes are often inherited from the parents. Often, this happens unexpectedly when two healthy carriers of a defective recessive gene reproduce. In other cases, it can also happen when the defective gene is dominant. Currently around 4,000 genetic disorders are known, with more being discovered.

Four types of genetic disorders are known. First, single gene disorder occurs as the result of a single mutated gene due to genomic imprinting and uniparental disomy that might include Mendelian disorders (e.g., Autosomal, X-linked and Y-linked) and non-Mendelian disorders (e.g., mitochondrial inheritance). Second, multi-factorial and polygenic disorders occur likely associated with the effects of multiple genes in combination with lifestyle and environmental factors. Third, disorders with variable modes of transmission, called heredity malformations, are congenital malformations which might be familial and genetic

or might be acquired by exposure to teratogenic agents in the uterus. Fourth, cytogenetic disorder exists due to alterations in the number or structure of the chromosomes and might cause autosomal disorders and sex chromosome disorders. Presently, heart disease, dermatitis, and cancer are known as genetic diseases that likely to occur due to multi-factorial disorders. Although in most cases complex disorders often cluster in families, but they do not have a clear-cut pattern of inheritance. This fact makes it difficult to determine a person’s risk of inheriting or passing on these disorders. This chapter would like to describe and further discuss the discovery of heart disease, dermatitis, and cancer through recognition of genes and protein patterns.

Pattern recognition is an integral part of machine vision and image processing (Duda, 2000; Fu, 1982, Gonzales, 1978; Pavlidis, 1977; Perner, 1996). The objective in pattern recognition is to recognize objects in the scene from a set of measurements of physical objects (Acharya, 2005). Each object is a pattern and the measured values are the features of the pattern. A set of similar objects possessing more or less identical features are said to belong to a certain pattern class. Presently, there are many classification techniques can be used for the recognition of patterns. These techniques work in which the classification of an unknown pattern is decided based on some deterministic or statistical or even fuzzy set theoretic principle.

The classification methods are mainly of two categories, i.e., supervised learning and unsupervised learning. Then, the supervised classification algorithms can further be classified into parametric classifiers and non-parametric classifiers. In parametric supervised classification, the classifier is trained with a large set of labeled training pattern samples in order to estimate the statistical parameters of each class of patterns such as arithmetic-mean, standard deviation, variance, etc. The term “labeled pattern samples” means the set of patterns whose class memberships are known in advance. Here the input feature vectors obtained during the

27 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/intelligent-classifiers-fusion-enhancing-recognition/76135

Related Content

Advances in Biosensors for In Vitro Diagnostics

David P. Klemmer (2010). *Biocomputation and Biomedical Informatics: Case Studies and Applications* (pp. 178-186).

www.irma-international.org/chapter/advances-biosensors-vitro-diagnostics/39611

Cancer Gene Expression Data Analysis Using Rough Based Symmetrical Clustering

Anasua Sarkar and Ujjwal Maulik (2013). *Bioinformatics: Concepts, Methodologies, Tools, and Applications* (pp. 1626-1641).

www.irma-international.org/chapter/cancer-gene-expression-data-analysis/76138

Detection of Breast Cancer by the Identification of Circulating Tumor Cells Using Association Rule Mining

Jananee S. and Nedunchelian R. (2016). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 17-25).

www.irma-international.org/article/detection-of-breast-cancer-by-the-identification-of-circulating-tumor-cells-using-association-rule-mining/171416

State-of-the-Art Review on the Models, Techniques, and Datasets to Diagnose COVID-19 Disease

Vetrithangam D., Naresh Kumar Pegada, Himabindu R., Arunadevi B. and Ramesh Kumar A. (2024). *Research Anthology on Bioinformatics, Genomics, and Computational Biology* (pp. 1427-1452).

www.irma-international.org/chapter/state-art-review-models-techniques/342583

Performance Analysis of Microarray Data Classification using Machine Learning Techniques

Subhendu Kumar Pani, Bikram Kesari Rath and Ajay Kumar Mishra (2015). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 43-54).

www.irma-international.org/article/performance-analysis-of-microarray-data-classification-using-machine-learning-techniques/167709