

Chapter 42

A Hybrid Kernel Extreme Learning Machine and Improved Cat Swarm Optimization for Microarray Medical Data Classification

S. Chakravarty

Orissa Engineering College, India

R. Bisoi

Siksha 'O' Anusandhan University, India

P. K. Dash

Siksha 'O' Anusandhan University, India

ABSTRACT

This paper presents the pattern classification of the binary microarray gene expression based medical data using extreme learning machine (ELM) and its variants like on-line sequential ELM (OSELM) and kernel based extreme learning machine (KELM). In the KELM category two variants namely the wavelet based kernel (WKELM) extreme learning machine and radial basis kernel extreme learning machine (RKELM) along with support vector machine (SVMRBF) and support vector machine polynomial (SVMPoly) are used to classify microarray medical datasets. Further to reduce the high dimensionality of Microarray medical datasets giving rise to high number of gene expression and small sample sizes, a modified evolutionary cat swarm optimization (MCSO) technique is adopted. The efficiency of the proposed algorithm is verified using a set of performance metrics for four binary medical datasets belonging to breast cancer, prostate cancer, colon tumor, and leukemia, respectively.

DOI: 10.4018/978-1-7998-1204-3.ch042

1. INTRODUCTION

Microarray analysis and classification play very important role for early diagnosis and treatment of disease like cancer. It shows the highest rate of morbidity and mortality in both economically developed and developing countries (Kanovas, 2006) where two hundred types of cancers in human beings have been recorded using microarray technology (Ferley et al., 2012). Molecular analysis of four common types of cancers belonging to lungs, female breast, bowel and prostate (Siegel et al., 2015) reveals that their gene expression profiles and high-density DNA microarray measures (Li et al., 2011; Lee & Lee, 2003) are different and therefore they can be used to diagnose them with significant accuracy. Early detection of these cancers can prolong human life and therefore it is pertinent to classify and recognize these types of cancers mentioned above as early as possible (Li et al., 2011).

Microarray technology that produces big datasets with gene expression values for thousands of genes (6000~60000) in a cell mixture (Guyon & Andre, 2006) are economically unviable where the number of samples (n) is much larger than the number of features (p) (Bolon-canedo et al., 2014). Thus for reducing the dimensionality problem efficient feature extraction and optimal feature selection of the medical datasets are required. To overcome this problem, microarray medical datasets need dimension reduction (Bolon-Canedo et al., 2014; Saeys et al., 2007; Tan, 2006; Liu & Hiroshi, 2007). Several feature extraction techniques like the principal component analysis (PCA) (Subasi & Ismain, 2010; Lopez et al., 2009; Karg et al., 2009), kernel principal component analysis (KPCA), linear discriminate analysis (LDA) (Subasi & Ismain, 2010; Lopez et al., 2009) and canonical correlation analysis (CCA) (Vert et al., 2002) have been used widely by several researchers. However, from the extracted features an optimal feature selection approach is essential to improve the classification accuracy and its better interpretation. Feature selection methods proposed in the literature include Fast Correlation Based Filter (FCBF) (Senliol et al., 2008), Relief algorithm (Zhang et al., 2008), Support Vector Machine recursive feature elimination technique (Duan et al., 2005), sequential forward (SFS) selection and sequential backward elimination (SBE) (Peng et al., 2005). Amongst several approaches for optimal feature selection SFS and SBE outperform all others and exhibit ease of implementation and low computational complexity. However, both SFS and SBE suffer from nesting and feature deletion problems (Dunne et al., 2002), respectively, and, therefore, evolutionary techniques are used. Various evolutionary optimization techniques for feature selection include genetic algorithm (Huang et al., 2006), simulated annealing (Lin et al., 2008), ant colony optimization (Kanan et al., 2008), particle swarm optimization (Unler & Alper, 2010; Fong et al., 2013; Lie et al, 2011), differential evolution (Khusaba et al., 2011; Al-Ani et al, 2013), bacterial foraging optimization (Jakhar et al, 2011), harmony search (Dias et al., 2012), cuckoo search (Rodrigius et al., 2013), firefly (Banati et al., 2011), bat algorithm (Nakamura et al., 2012) and cat swarm optimization (Lin et al., 2009), etc.

Dimensionality reduction helps in the classification of microarray medical datasets by improving its accuracy. The important role of medical data classifier is to provide the explanation and justification for the accurate prediction of the disease (Li et al., 2011). Many traditional classifiers like KNN (Guo et al., 2003), naïve-bayes (NB) (McCallum et al., 1998), decision tree (Vlahou, et al., 2003), random forest (Diaz-Uriarte et al., 2006), ID3 (Elias, 2003), C4.5 (Quinlan, 2014) and various neural network based classifiers like multi layer perceptron (MLP) (Chaudhuri & Bhattacharya, 2000), RBFNN (Subasini et al., 2009), FLANN (Misra & Dehuri, 2007), SVM (Chu & Lipo, 2005; Cortes et al., 1995; Furey et al, 2000; Hsu et al, 2003) are found in the literature. Amongst all the classifiers, ANN and its variants are extensively used by researchers to classify medical datasets (Delen et al., 2005). The success

34 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/a-hybrid-kernel-extreme-learning-machine-and-improved-cat-swarm-optimization-for-microarray-medical-data-classification/243144

Related Content

Big Data and Cloud Computing: A Review of Supply Chain Capabilities and Challenges

Marcus Tanqueand Harry J. Foxwell (2018). *Exploring the Convergence of Big Data and the Internet of Things* (pp. 1-28).

www.irma-international.org/chapter/big-data-and-cloud-computing/187888

Study and Analysis of Delay Factors of Delhi Metro Using Data Sciences and Social Media: Automatic Delay Prediction System for Delhi Metro

Arun Solankiand Ela Kumar (2018). *Innovative Applications of Big Data in the Railway Industry* (pp. 209-223).

www.irma-international.org/chapter/study-and-analysis-of-delay-factors-of-delhi-metro-using-data-sciences-and-social-media/191747

A Novel Approach for Tenuous Community Detection in Social Networks

Muhammad Asif, Hassan Razaand Muhammad Imran Manzoor (2022). *International Journal of Data Analytics* (pp. 1-12).

www.irma-international.org/article/a-novel-approach-for-tenuous-community-detection-in-social-networks/297518

Towards Clinical and Operational Efficiency through Healthcare Process Analytics

Vassiliki Koufi, Flora Malamateniouand George Vassilacopoulos (2016). *International Journal of Big Data and Analytics in Healthcare* (pp. 1-17).

www.irma-international.org/article/towards-clinical-and-operational-efficiency-through-healthcare-process-analytics/171401

Depression Detection in Online Social Media Users Using Natural Language Processing Techniques

Haseeb Ahmad, Faiza Nasir, C. M. Nadeem Faisaland Shahbaz Ahmad (2022). *Handbook of Research on Opinion Mining and Text Analytics on Literary Works and Social Media* (pp. 323-347).

www.irma-international.org/chapter/depression-detection-in-online-social-media-users-using-natural-language-processing-techniques/298878