Neurocomputing 125 (2014) 217–228

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A multi-objective evolutionary algorithm-based ensemble optimizer for feature selection and classification with neural network models

Choo Jun Tan [a], Chee Peng Lim [b,*], Yu–N Cheah [a]

[a] School of Computer Sciences, Universiti Sains Malaysia, Malaysia
[b] Centre for Intelligent Systems Research, Deakin University, Australia

ARTICLE INFO

Article history:
Received 25 July 2012
Received in revised form
20 December 2012
Accepted 27 December 2012
Available online 11 July 2013

Keywords:
Multi-objective optimization
Evolutionary algorithm
Feature selection
Neural network classifiers
Ensemble models

ABSTRACT

In this paper, we propose a new multi-objective evolutionary algorithm-based ensemble optimizer coupled with neural network models for undertaking feature selection and classification problems. Specifically, the Modified micro Genetic Algorithm (MmGA) is used to form the ensemble optimizer. The aim of the MmGA-based ensemble optimizer is two-fold, i.e. to select a small number of input features for classification and to improve the classification performances of neural network models. To evaluate the effectiveness of the proposed system, a number of benchmark problems are first used, and the results are compared with those from other methods. The applicability of the proposed system to a human motion detection and classification task is then evaluated. The outcome positively demonstrates that the proposed MmGA-based ensemble optimizer is able to improve the classification performances of neural network models with a smaller number of input features.

1. Introduction

Machine learning techniques require an effective feature representation for both training and knowledge acquisition [1]. Reports in the literature have shown the possibility of extracting useful patterns, e.g. discovering patterns from Mackey–Glass time series [2]; analyzing medical data sets [3,2] from the UCI machine learning repository [4] using hybrid optimization models and neural networks. In general, the main objective of feature selection is to identify and remove irrelevant and/or redundant features, and to retain only the important ones. A number of feature selection methods are available, which include dimensionality reduction and extraction of data features [5]. In this respect, the challenge to construct an accurate classifier with the smallest possible number of features has been addressed [6]. A feature selection strategy in conjunction with decision stumps, sample compression, and Bayesian-based learning settings has been deployed as part of the classifier design [6]. Besides that, evolutionary algorithm-based methods for simultaneous optimization of feature selection and feature extraction with parameter tuning in classification tasks using microarray data sets [7], and thyroid-related data sets [8] have also been reported.

On the other hand, ensemble models have been shown to be useful for improving performances of machine learning techniques. In this respect, an ensemble weather prediction model, i.e. the DEMETER project [9], has shown the ability to improve the deterministic forecast of future atmospheric states, and to produce useful probabilistic information for weather-sensitive parties. In [10], the Local Occupancy Pattern (LOP) features with an Actionlet Ensemble Model (AEM) have been proposed for human action recognition with depth cameras. The concept of Actionlet is based on a conjunction of features for a subset of the human kinematic joints or a structure of the features, and the Actionlet Ensemble represents a linear combination of the Actionlets with discriminative weights learned from a multiple kernel learning method. The LOP feature is associated with each three-dimensional joint, and is treated as the depth of appearance. In one of the test data sets, i.e. MSRDailyActivity3D (a data set with sixteen human activities captured by a Kinect device in a living room), the AEM has shown improvement in classification accuracy as compared with those from various methods, i.e. Dynamic Temporal Warping, the methods associated with the LOP features only and with the joint position features only, and the Support Vector Machine (SVM) with the Fourier Temporal Pyramid features. Besides that, the AEM shows better tolerance to the intra-class variations and better robustness to noise and errors in the depth maps and joint positions in two other data sets, i.e. MSR-Action3D and CMU MoCap data sets. In [11], the Locally Weighted Ensemble (LWE) [12] method has been used to combine the prediction outputs of various transfer learning models, i.e. the Bag of Visual-Words (BoVW) model and the Bag of Bilingual-Words (BoBW) model, to address the problem of recognizing an unknown action from

* Corresponding author. Tel.: +613 5227 3307.
E-mail address: chee.lim@deakin.edu.au (C.P. Lim).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.12.057

an unseen (target) view using training data taken from other (source) views. The LWE results are 5–10% better than those from the global weighting method as well as other methods using a human IXMAS multi-view action data set. In another study, the problem of articulated human pose estimation in videos has been addressed with an ensemble of tractable models using a human VideoPose2.0 data set [13]. The proposed method has shown better performance as compared with other results in single-frame parsing based on a simple max-marginal combination algorithm.

Similar to [14], feature selection is carried out using an ensemble of Multi-Objective Evolutionary Algorithm (MOEA)-based optimizers in this study. Specifically, we propose an MOEA-based optimization model to generate a set of solutions for neural network classifiers. The main objectives of our work include

- To track and reinforce the behaviour of an MOEA-based optimizer, i.e. a Modified micro Genetic Algorithm (MmGA) [15], towards the pareto front. This is achieved via the Apportionment of Credit (AoC) and Reinforcement Learning (RL) schemes [16].
- To undertake Multi-objective Optimization Problems (MOPs) that aim to minimize the number of features and to maximize a number of performance indicators. This is achieved via an ensemble of MmGA-based optimizers, which is designed using the classifier fusion [17] concept and the voting-based elite-selection scheme [18].

The quality of MOEA solutions can be measured using several criteria [19]. It is believed that a good coverage of the trade-off surface from a pareto optimal set needs to strike a balance among convergence, diversity and spread [20] of the solutions. These properties can be measured using various performance indicators [21]. In our previous work [15], the MmGA model has been proposed for improving the convergence rate towards the pareto front (**pf**) in undertaking MOPs. The MmGA model incorporates an NSGA-II inspired elitism strategy, a pareto dominance sorting concept, and an extended population formation procedure into the original micro Genetic Algorithm (mGA) [22]. In this paper, we further extend the work in [15] by assembling a number of MmGA optimizers in an ensemble structure. Our work contributes towards formulating an ensemble optimization model for undertaking feature selection and classification problems, and with the following advantages:

- Efficiency: a model that is able to construct an accurate classifier using a smaller number of features.
- Flexibility: an ensemble structure that enables the constituent optimizers to form a team without having to have a prior configuration, and this is achieved by using a feedback environment with the AoC and RL schemes.

A number of benchmark data sets from the UCI machine learning repository and a real human motion detection and classification data set have been used to evaluate the proposed approach. The data samples are converted to the WEKA (Waikato Environment for Knowledge Analysis [23]) acceptable formats before invoking a number of classifiers in WEKA, i.e. (i) J48 for the UCI data sets; (ii) the Multi-Layer Perceptron (MLP) network [24] and Radial Basis Function (RBF) [24] neural network classifiers for the human motion detection and classification data set. A voting-based elite-selection scheme is integrated with the MmGA-based ensemble, and the resulting model is used in tandem with the classifiers in WEKA. The 10-fold cross-validation method is used to produce the results from each classifier. The bootstrap method [25] is used to quantify the experimental results, as well as to ascertain the stability of the proposed model.

The rest of this paper is organized as follows. In Section 2 a discussion on related work is presented. The details of our proposed model are presented in Section 3. In Section 4, two experimental studies using UCI data sets and a human motion detection and classification data set are presented, with the results discussed and analyzed. Conclusions and suggestions for further work are presented in Section 5.

2. Related work

The relationships between human motion and health are complex, as they are influenced by various factors. Indeed, the complex nature of physical activities and the difficulties in measuring physical activities largely govern the issues relating human motion and health [26,27]. To recognition human motions, the accelerometer, e.g. in mobile devices, has been widely used, as reported in the literature, e.g. in [28–30]. In [28], the SVM-based classifier was used to recognize seven different activities with a total of 48.2 h of data collected from seven subjects. In [29], the C4.5 decision tree, Naive-Bayes (NB), k-nearest neighbor (kNN), and SVM were used to recognize five activities from eight subjects. The activities included stationary, walking, running, bicycling, and driving. According to the findings in [29], some cycling events were misclassified as walking and driving activities. A more recent work [30] used a device equipped with an IC (Integrated Circuit)-based accelerometer and a factorized quaternion approach to determine the arm limb orientations and arm motions. It was conducted using a wireless sensor network with a tri-axial accelerometer attached to the human arm.

On the other hand, wearable sensors, which include accelerometers, have been used for classifying various human activities, e.g. level walking, walking upstairs and downstairs, jogging, running, hopping with the left or right leg, as well as jumping [31–34]. The accelerometer was denoted as the most information-rich and most accurate sensor for activity recognition using the kNN classifier [31]. In [32], a series of experiments using a number of classifiers, i.e. Instance-Based Learning (IBL), C4.5, and NB, to detect human physical activities using five small bi-axial accelerometers tagged at different parts of the human body was conducted. The results showed that the input features (from multiple accelerometers) could be used to effectively discriminate and recognize different human physical activities. In another work [33], C4.5, kNN, NB, and the Bayesian-Net classifiers were used to evaluate multiple feature sets and sampling rates obtained from multiple wearable sensors placed at different body locations. The aim was to classify six different physical activities in real time. The decision tree and MLP classifiers were used in [34] for classification of seven different human activities with the accelerometer signals obtained from wearable sensors.

Similar to other investigations in the literature, we use an accelerometer embedded in a smartphone for human motion detection and classification in this study. But, we focus on two activities (walking and running) with a larger pool (i.e. a total of 57) of subjects. Each subject has been asked to perform the two activities (walking and running) with a fixed frequency, i.e. 100 steps for each activity. We then use the MmGA-based ensemble to optimize the performances as well as the number of input features needed for recognizing and classifying these activities using neural network models. Our main contribution is to demonstrate the effectiveness of the MmGA-based ensemble optimizer to work in tandem with neural network classifiers in undertaking MOPs.

The use of an ensemble of classifiers has been reported in the literature. In general, the idea is that a number of classifiers