International Journal of Innovative
Computing, Information and Control
Volume **7**, Number **12**, December **2011**

 ISSN 1349-4198
pp. 7051–7066

HANDLING CLASSIFICATION PROBLEMS WITH IMPERFECT LABELS USING AN EVIDENCE-BASED NEURAL NETWORK ENSEMBLE

Mahdi Tabassian[1,2], Reza Ghaderi[1] and Reza Ebrahimpour[2,3]

[1]Faculty of Electrical and Computer Engineering
Babol University of Technology
Babol, Iran
r_ghaderi@nit.ac.ir

[2]School of Cognitive Sciences
Institute for Research in Fundamental Sciences (IPM)
Niavaran, P.O. Box 19395-5746, Tehran, Iran
{ tabasian; ebrahimpour }@ipm.ir

[3]Brain and Intelligent Systems Research Lab
Department of Electrical and Computer Engineering
Shahid Rajaee Teacher Training University
Lavizan, P.O. Box 16785-163, Tehran, Iran

Received June 2010; revised October 2010

Abstract. *In this paper, a method for handling imperfect labels using belief functions has been presented. By extracting different types of features from data, the proposed method takes advantage of information redundancy and complementariness between sources. The initial label of each training sample is ignored and based on its closeness to prototypes of the main classes, it is then reassigned to one class or any subset of the pre-defined classes. Multilayer perceptron (MLP) neural network is used as base classifier and its outputs are interpreted as basic belief assignment (BBA) and in this way, partial knowledge about the class of a test pattern is encoded. The BBAs are then discounted based on the reliability of the base classifiers in identifying validation samples and are pooled using Dempster's rule of combination. Experiments with artificial and real data demonstrate that, by considering the ambiguity in labels of the learning data, the proposed method can outperform single and ensemble classifiers that solve the classification problem using data with initial imperfect labels.*
Keywords: Data with imperfect labels, Dempster-Shafer theory, Classifier combination, Neural network

1. **Introduction.** Combination of multiple classifiers is a popular and suitable approach to deal with supervised classification problems [1-4]. This popularity comes from the ability of an ensemble system in handling main challenges of the supervised learning. Classification problems with heavily overlapping class distributions, nonlinear class boundaries, small sample size and imperfect class labels are some examples of difficulties that a supervised classification framework may confront. Among them, the issue of uncertainty in labels of the learning data can dramatically degrade the classification performance. It is because of the importance of the labeled data in the learning procedure where a classifier tries to appropriately fit to the training data and make decision about the class of a test sample based on the characteristics of the learned data. When the class labels have been imperfectly assigned to the learning samples, the optimal performance of a classifier on