



Multiple Self Organising Maps (mSOMs) for simultaneous classification and prediction: Illustrated by spoilage in apples using volatile organic profiles

Siong Fong Sim ^{a,*}, Virág Sági-Kiss ^b

^a Universiti Malaysia Sarawak, Faculty of Resource Science and Technology, 94300 Kota Samarahan, Sarawak, Malaysia

^b Corvinus University of Budapest, Department of Applied Chemistry, H-1118 Budapest, Villányi út 29-33, Hungary

ARTICLE INFO

Article history:

Received 11 May 2011

Received in revised form 12 July 2011

Accepted 2 August 2011

Available online 7 August 2011

Keywords:

Volatile organic compounds

Apple spoilage

Multiple SOMs

Growing SOMs

Simultaneous classification and prediction

ABSTRACT

Self Organising Maps (SOMs) have been modified and extended for various applications. This paper employed multiple SOMs (mSOMs) in supervised manner for multitasking involving simultaneous classification and prediction providing more information on a sample. This was demonstrated on the GC–MS dataset of apple spoilage where the volatile organic compounds (VOCs) of two groups of apples, control and inoculated, were monitored over 2 to 10 days. Multiple SOMs were used to determine whether a sample is a control or an inoculated apple and at the same time the spoilage day is predicted, i.e. how many days an apple has been left with or without inoculation. Multiple SOMs are different from traditional supervised SOMs; in mSOMs, samples are divided into classes for training on different maps. This approach of SOMs does not require optimisation of the scaling value however, it is important to make sure that the clustering qualities of mSOMs are comparable. Growing Self Organising Maps (GSOMs) was incorporated to automatically determine the suitable map size with comparable clustering qualities based on the mean quantization error. The findings demonstrated that mSOMs can be potentially applied for simultaneous analysis allowing more information to be retrieved on a sample reducing the overall computational time of an analysis.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Self Organising Maps (SOMs) is one of the most popular neural network methods. It was introduced as an unsupervised learning method however over the years, various extensions and modifications have been proposed to enhance the applicability of SOMs to a broader range of applications [1–5] for example, supervised SOMs are employed for pattern recognition i.e., classification [6,7], prediction [8] and variable selection [9,10]. In this paper, we demonstrate the application of SOMs for a more demanding task involving simultaneous classification and prediction. Simultaneous classification and prediction has been found to be very useful in marketing research; individuals are simultaneously classified and predicted according to segments and segment levels so that responses to a new product design can be predicted and simulations can be performed at the segment level [11]. Other simultaneous analyses such as classification and feature selection have also been published; these analyses are fundamentally based on the machine learning method [12–15].

As the size of datasets continues to grow, methods for simultaneous analyses have grown in popularity as this would provide more information on a sample and reduce the overall computational time. Multilevel Simultaneous Component Analysis has been one of the most commonly used approaches in enhancing the interpretability of multivariate data [16]. In this paper, we introduce multiple SOMs (mSOMs) in a supervised manner for simultaneous classification and prediction. This is demonstrated on the GC–MS dataset of apple spoilage where the volatile organic compounds (VOCs) of two groups of apples, control and inoculated, were monitored over 2 to 10 days [17]. Multiple SOMs are used to determine whether a sample is a control or an inoculated apple and at the same time the spoilage day is predicted, i.e. how many days an apple has been left with or without inoculation. Typically, the approaches employed for classification and prediction of apple spoilage are based on Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Soft Independent Modelling of Class Analogy (SIMCA), Partial Least Squares (PLS) and Artificial Neural Network (ANN) [18–21]. In classification, it involves prediction of categorical (discrete) labels whilst prediction models continuous valued functions [22]. Classification using SOMs traditionally includes a set of variables representing the class membership information for the training of the map [7]; for prediction however, a vector describing the interested property is incorporated [23]. In this paper, SOMs are

* Corresponding author. Tel.: +60 82 582995.

E-mail address: sfsim@frst.unimas.my (S.F. Sim).