Computers and Electronics in Agriculture 102 (2014) 85–97

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Predicting shellfish farm closures using time series classification for aquaculture decision support

Md. Sumon Shahriar *, Ashfaqur Rahman, John McCulloch

Intelligent Sensing and Systems Laboratory (ISSL), Commonwealth Scientific and Industrial Research Organization (CSIRO), Information and Communication Technology (ICT) Centre, Hobart, Australia

ARTICLE INFO

Article history:
Received 15 July 2013
Received in revised form 10 January 2014
Accepted 14 January 2014

Keywords:
Time series classification
Prediction
Machine learning
Aquaculture decision support

ABSTRACT

Closing a shellfish farm due to pollutants usually after high rainfall and hence high river flow is an important activity for health authorities and aquaculture industries. Towards this problem, a novel application of time series classification to predict shellfish farm closure for aquaculture decision support is investigated in this research. We exploit feature extraction methods to identify characteristics of both univariate and multivariate time series to predict closing or re-opening of shellfish farms. For univariate time series of rainfall, auto-correlation function and piecewise aggregate approximation feature extraction methods are used. In multivariate time series of rainfall and river flow, we consider features derived using cross-correlation and principal component analysis functions. Experimental studies show that time series without any feature extraction methods give poor accuracy of predicting closure. Feature extraction from rainfall time series using piecewise aggregate approximation and auto-correlation functions improve up to 30% accuracy of prediction over no feature extraction when a support vector machine based classifier is applied. Features extracted from rainfall and river flow using cross-correlation and principal component analysis functions also improve accuracy up to 25% over no feature extraction when a support vector machine technique is used. Overall, statistical features using auto-correlation and cross-correlation functions achieve promising results for univariate and multivariate time series respectively using a support vector machine classifier.

1. Introduction

Data mining and machine learning techniques can play important roles in decision support systems. (Lavrac and Bohanec, 2003). Examples of supervised learning being used to assist decision making can be found in many domains, e.g. health (Khan et al., 2012), finance (Matsatsinis, 2002), industry (Chi et al., 2009). In the shellfish aquaculture industry, decision support systems play a role in significant decision making around farm closures to manage the risk to public health.

In Tasmania, the Tasmanian Shellfish Quality Assurance Program (TSQAP) is responsible for closing farms to protect public health (D'Este et al., 2012). One of the frequent reasons for closing farms is due to high levels of microbial contamination (thermotolerant coliforms) in the water. These pollutants are closely related to environmental process of rainfall run-off from land use. Microbe levels cannot be directly measured so the TSQAP use a complex set of proxies, guidelines and expert knowledge to determine timing of closing and re-opening farms due to possible microbial contamination. Typically they examine recent rainfall and river flow data when making decisions.

The industry bases a significant portion of its wet processes on the tides, and as such, operations can occur any time day or night. Prediction of closure decisions, even a few hours in advance, will benefit the industry significantly by reducing unnecessary labour and stock handling. Even though management decisions are guided by rules around rainfall and river flow, expert system based predictors of these decisions are poor (D'Este et al., 2012). This paper investigates data-driven approaches to this prediction problem. We aim to develop a data-driven system that can provide a useful "one day ahead" prediction of decisions that the TSQAP is likely to make.

Fig. 1 outlines our approach to providing decision support to assist in the prediction of TSQAP decisions.

Firstly we consider a time series of the previous two-weeks of daily rainfall and river flow, and relate that to the following day's closure status. The length of the time series (2-weeks) is based on the domain knowledge from the TSQAP. We then reduce the dimensionality of this time series data through the use of one of several feature extraction processes such as auto-correlation function (ACF) and cross-correlation function (CCF) which are explained later in this section. We then train several different

* Corresponding author. Tel.: +61 362325548; fax: +61 362325229.
E-mail addresses: mdsumon.shahriar@csiro.au (M.S. Shahriar), ashfaqur.rahman@csiro.au (A. Rahman), John.McCulloch@csiro.au (J. McCulloch).

0168-1699/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.compag.2014.01.011

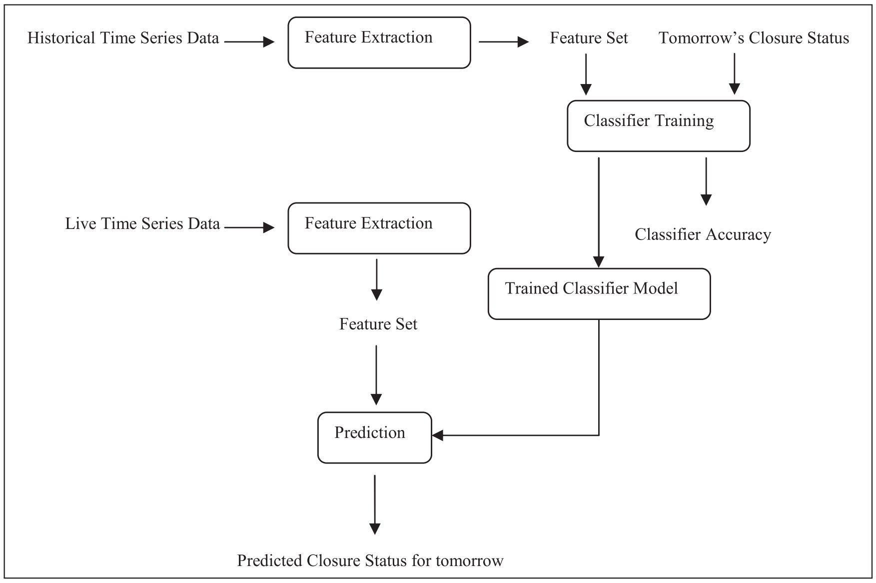

Fig. 1. Prediction model of shellfish farm closure using time series classification.

classifiers using these features and the associated closure status in order to select the best classifier model/feature extraction method combination based on accuracy. The run-time system then feeds streaming time series data into the feature extraction function and then applies the trained classifier to the result to form a prediction of tomorrow's TSQAP closure status decision.

In the model, we find that feature extraction of the time series prior to classification using extracted features is important for prediction of the closure status. Specifically, the accuracy of the classifiers depend on the feature extraction methods.

The feature extraction methods for univariate and multivariate time series are different. We now present examples to illustrate the methods.

Fig. 2 shows the time series of rainfall and river flow that leads to TSQAP closing an aquaculture farm at Montagu, Tasmania, Australia. Fig. 3 shows the time series of rainfall and river flow that leads to TSQAP's re-opening decision of the same farm.

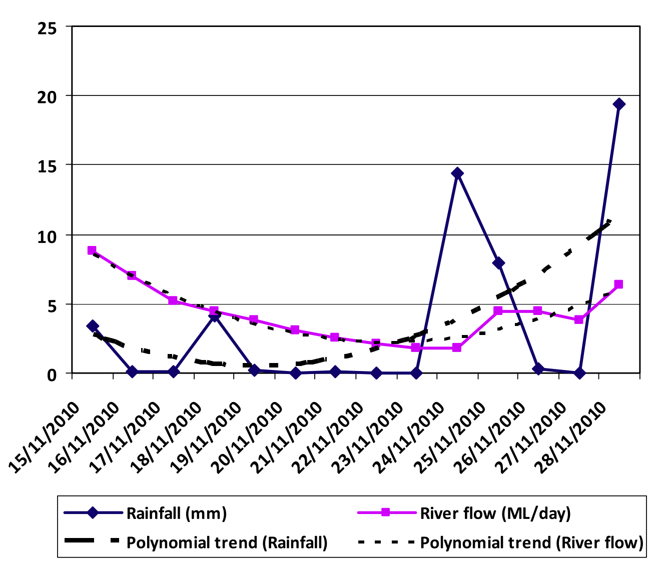

Fig. 2. Time series of rainfall and river flow for two weeks that leads to a decision to close a farm (i.e. *Closed*) on 29/11/2010 at Montagu, Tasmania, Australia.

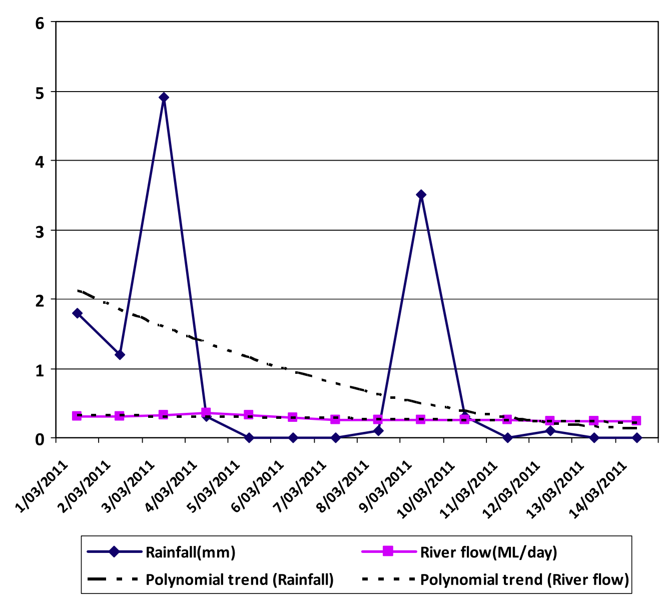

Fig. 3. Time series of rainfall and river flow for two weeks that leads to a decision to re-open a farm (i.e. *Opened*) on 15/03/2011 at Montagu, Tasmania, Australia.

From Fig. 2, we observe that a high rainfall event, and the subsequent high river flow, led to a farm closure. On the contrary, in Fig. 3, low rainfall, and subsequent low river flow, led to re-opening the farm. This observation is consistent with the TSQAP's processes of examining rainfall and river flow. The TSQAP's guidelines vary from site to site and may be based on rainfall and river flow, combined or separately. In our univariate approach we consider rainfall as the time series, and in our multivariate approach we consider rainfall and river flow. Rainfall does affect river flow, however the relationship is complex (Nayak et al., 2012). The relationship between rainfall and riverflow can further be complicated due to other factors such as the geology of a location and discovering this complex relationship is beyond the scope of our research. Thus, our aim is to see what features of rainfall and what features of combined rainfall and river flow can discriminate in predicting closure and re-opening of farms.