Future Generation Computer Systems 33 (2014) 81–89

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Zebra mussels' behaviour detection, extraction and classification using wavelets and kernel methods

Piotr Przymus [a,*], Krzysztof Rykaczewski [a], Ryszard Wiśniewski [b]

[a] Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Chopina 12/18, 87-100 Toruń, Poland

[b] Laboratory of Applied Hydrobiology, Nicolaus Copernicus University, Gagarina 9, 87-100 Toruń, Poland

HIGHLIGHTS

- We have developed a fully automated method for extraction and analysis of the behaviour of Dreissena polymorpha.
- We have evaluated usefulness of the feature set used for classification.
- We have proposed a framework for the classification of control and stress conditions for the purpose of the risk analysis.

ARTICLE INFO

Article history:
Received 15 June 2012
Received in revised form
11 April 2013
Accepted 13 April 2013
Available online 22 April 2013

Keywords:
Wavelet
Zebra mussel
Feature extraction
Classification
BEWS

ABSTRACT

This paper concerns the detection, feature extraction and classification of behaviours of *Dreissena polymorpha*. A new algorithm based on wavelets and kernel methods that detects relevant events in the collected data is presented. This algorithm allows us to extract elementary events from the behaviour of a living organism. Moreover, we propose an efficient framework for automatic classification to separate the control and stressful conditions.

1. Introduction

One of the most important problems connected with health and environmental protection is the monitoring of water pollution. However, many existing systems do not offer continuous monitoring and examine only a narrow range of substances. Therefore, systems based on living organisms (i.e. Biological Early Warning Systems, BEWS) have become increasingly popular recently [1–3]. Building such a system is a complex task, which requires a selection of an appropriate bioindicator for the monitored environment, preparation of a measurement system that provides data for further processing, analysis and classification methods.

It is well-known that aquatic organisms are sensitive to changes in concentrations of various substances (such as *xenobiotics* and other toxic compounds) in water, therefore can be successfully used as bioindicators. Among the most frequently used are: *cladocerans* [4,5], *amphipods* [6], *mussels* [7–9], the larvae of aquatic insects *(Chironomidae)* [10] and fish [11]. Especially mussels, like *Mytilus* or *Dreissena*, as sessile bivalves, are very suitable for long-term, *in situ* water quality monitoring.

There are several methods for measuring the response of mussels to stress factors. One option is to measure the frequency of shell closing–opening states, through gluing wires to both halves of the shell and connecting them through an interface to a computer [12,13]. The number of closed mussels in a treated group, in comparison to control one, is a measure of stress response. Another option is to use a magnetic coil (or a Hall sensor) on one valve, and a magnet on the other [14,15]. The value of the amplified signal is proportional to the distance (gape) between the two valves. Various factors are taken into consideration as a response to stress of a tested group, such as the difference of the average value of the gape in comparison to control mussels [15], increased activity or increased time in which the shells are closed or are open [14].

These solutions, however, neglect more complex behaviour of mussels, such as changes in elementary movements. The sequence

* Corresponding author. Tel.: +48 56 611 3463; fax: +48 56 611 2987.
E-mail addresses: eror@mat.umk.pl, piotr.przymus@gmail.com (P. Przymus), mozgun@mat.umk.pl (K. Rykaczewski), wisniew@biol.uni.torun.pl (R. Wiśniewski).

0167-739X/$ – see front matter
http://dx.doi.org/10.1016/j.future.2013.04.015

of elementary events, i.e. the extent of gape change and the time of the first return to the initial gape value can form specific patterns for various natural or stress caused activity rhythms [16,17]. In some cases, the appearance of different patterns in elementary behaviour is the only difference compared to the control group showing appearance of hazardous substances [18].

Therefore, a fully automated method of analysis for the extraction of the behaviour of *Dreissena polymorpha* mussels is proposed and a framework (see Fig. 1) for risk analysis and classification of control and stress conditions based on elementary behaviour is presented. We discuss here our extraction algorithm based on wavelets and kernel methods. Then we show how to build a collection of observations based on a combination of our event extraction algorithm and the method proposed in [19] for EEG signals as well. Moreover, an assessment of the adequacy of the set of features used for classification is made. We carry out risk analysis and classification experiments using different classifiers (*k*-NN, SRDA, FDA) and the sets of features. Finally, we present and discuss the results.

Based on our extraction algorithm, a set of tools for automatic extraction and analysis of elementary events is proposed. We investigate the usefulness of the proposed extraction method as a tool to support laboratory work, which is an important improvement since this work was mostly done manually before.

The paper is partly an extension of [20]. It is organised as follows. The motivation of this work is presented in Section 2. Section 3 presents the data collection used in this work. Section 4 presents the theory of wavelets necessary in Section 5, where behaviour extraction algorithm is described. The evaluation of the effectiveness of the behaviour extraction algorithm is given in Section 6. The framework for automatic classification of control and stress conditions is discussed in Section 7. Finally, Section 8, summarise the results of the paper.

2. Motivation

Biomonitoring based on the behaviour of mussels, and particularly of the species *Dreissena polymorpha*, has been repeatedly tested in the laboratory and in natural conditions, and its effectiveness has been confirmed in many studies [12,14,15,13]. Many of the biomonitoring systems use *Dreissena polymorpha* as a bioindicator, but each of them operates on different principles. The system described in [12,13] monitors the activity of mussels by checking whether they are open or closed. A measure indicating the level of risk is the percentage of closed mussels in comparison to the percentage of closed mussels in the control group. In systems based on [14,15], the distance between valves is measured using Hall sensor and a magnet. Risk analysis is based on the following factors: the average level of valve distance, mussels activity (number of opening and closing events), and the average time during which the mussels are closed/open.

However, the existing systems do not exploit the full potential of mussels as a bioindicator. Experiments were conducted [18,17] to show that a change in activity of mussels may be more subtle in response to various concentrations of substances such as *cyanotoxins, herbicides, salt* or *LPS (Lipopolysaccharide)*. These publications dealt with an analysis of the life rhythms described in [16] and took under consideration changes that may appear in the elementary behaviour to suggest the appearance of harmful substances. A typical *elementary event* (or *behaviour*) consists of the following stages: closing, opening and resting (see Fig. 2, all these stages usually include some perturbations).

It turns out that the behaviour of the bioindicator could be slightly changed at the level of elementary behaviours. Moreover, it may be the only change compared to the control mussels. In practice, this means that the elementary event is disturbed in some of the three phases. Some examples of distorted elementary movements are placed in Fig. 3.

Our contribution in this area is the development of feature extraction algorithm, so that we can automatically extract elementary events from the behaviour of the mussels as described above. Then we assign to such behaviour a set of features used later in risk analysis.

Since the behaviour of the zebra mussels is recorded as a long series of shell states, logged at every second, we needed efficient analytic tools such as wavelets [10,21]. Note that Fourier Transform (FT) technique can be applied to analyse the frequency spectrum, but it does not provide any insight into which frequency component is present and when. Therefore, to fulfil the requirements of our system, we applied wavelets and kernel methods. These tools are better suited to study the long-term records of sudden changes in the animal behaviour; they can gain information about the time and duration of the peak in the reaction of the living being (see Section 5.4.1).

3. Data selection

In this section we describe a set of data originating from a biological experiment, which will be further analysed.

The data are a record of freshwater mussel activity, measured as the distance between the two halves of the shell. Mussels are in flow-through aquarium and are attached to its wall (due to the stationary lifestyle, it does not significantly affect their behaviour). A sensor measures changes in the magnetic field of the magnet placed on the one bivalve shell. The measurements are recorded with one second rate and are transmitted to the database. The system monitors eight mussels at the same time and each sensor generates a measurement from the interval [0, 45] with a possible measurement error of ± 3.

For the analysis we used the data from experiments carried out in laboratory conditions where the effect of salt and herbicides on the behaviour of mussel was examined. In the experiment we used mussels from the same colony and all experiments were conducted with at least 50% of the mussels in control conditions. In this paper, we analyse a total of 853 min of herbicide, 986 min of salt and 2193 min of control measurements.

Due to customary maximum time response of BEWS system [11], we analyse data set of length 1000 s (16 min and 40 s), which from now on will be called *fragments* (or *moving windows*, since these fragments are moving right over time with the velocity 1 time point per 1 s). In the data preparation steps we remove the white noise and average the signal using a high-pass filter (wavelet filter), which passes high frequency components and completely rejects all others.

4. An overview of the wavelet theory

In this section, we briefly present the basic theory of wavelets and comment on their properties that are useful in the digital signal processing.

The basic theory of wavelets was established by Haar in 1909, and then developed in the mid 1960s (cf. [22]). Wavelets are very successful as an analytical tool in the representation of signals, denoising, data compression and time-scale analysis of a time series. In many situations the results obtained by this method are better than those obtained by a Fourier analysis or other methods of filtration, because wavelet transform can be used to analyse even non-stationary signals. Moreover, the coefficients in wavelet expansion include both information about the frequency and time of the peak (*multiscale resolution analysis*). This differentiates them from the Fourier transform and makes them particularly useful