

Chapter 14

DSMS and Online Algorithms

Vasileios Megalooikonomou, Dimitrios Triantafyllopoulos,
Evangelia I. Zacharaki, and Iosif Mporas

Abstract Online (real-time) analysis of medical data is crucial for automatic or semi-automatic monitoring of patients. Technologies involved in online analysis are data streaming, online data management and online data processing. In this chapter we present algorithmic methodologies and implementations for online (real-time) analysis of medical data, i.e. streaming data management and online detection of events of interest.

14.1 Online Analysis of Medical Data

Methods for the online analysis of the above modalities require the identification of features and measures that can be extracted and calculated fast. The detection of seizures at the earliest observable onset of the ictal patterns is one of the main purposes of online analysis and can be used to start more detailed diagnostic procedures during seizures and to differentiate seizures from other related disorders with seizure like symptoms. Several methods have been proposed for this purpose that can be classified depending on the modality of the recorded data (intracranial EEG, scalp EEG and ECG).

In this chapter we present online methodologies for real-time seizure detection from EEG and ECG signals. Moreover we present the implementation of online alpha rhythm detection. Finally, details about the DSMS implementation of the online tools are presented.

V. Megalooikonomou (✉) • D. Triantafyllopoulos • E.I. Zacharaki • I. Mporas
Multidimensional Data Analysis and Knowledge Discovery Laboratory,
Department of Computer Engineering and Informatics,
University of Patras, Patras, Greece
e-mail: vasilis@ceid.upatras.gr

14.2 Online Analysis Algorithms

In online analysis of medical data the challenge is the development of tools which operate robust enough as well as in real time. Typically, there is a trade-off between performance accuracy and computational complexity of the algorithm. Thus, the aim is the configuration of online methodologies in the correct operational point. In this section we present the algorithm architecture, signal processing and classification setup performed for seizure detection within the ARMOR project, as well as the implementation of it on the Stream-In-Sight (DSMS) framework.

We developed [1] a seizure detector using EEG and ECG signals based on short-time analysis with time-domain and frequency-domain features and classification using support vector machines. We evaluated a large-scale set of time-domain and frequency-domain EEG and ECG features for seizure detection, which are popular in the literature for brain and heart statistical signal processing respectively. Furthermore, we investigated the effect on the detector’s performance when using subsets of these features, with respect to a feature ranking evaluation, in order to develop online (real-time) and offline versions of it. Feature ranking investigation and evaluation of the seizure detector using subsets of features showed that the feature vector composed of approximately the 800-best ranked features provides a good trade-off between computational demands and accuracy. The block diagram of the seizure detection architecture adopted in ARMOR is illustrated in Fig. 14.1.

The data captured from the $N + 1$ sensors (where N are the EEG electrodes plus one ECG channel) have been synchronized and transmitted as streams of multidimensional signals. Thus, the input to the illustrated in Fig. 14.1 architecture consists of time-synchronous streams of EEG and ECG signal samples. As shown in Fig. 14.1, in a first step the EEG, $x_{EEG} \in \mathbb{R}^N$, and ECG, $x_{ECG} \in \mathbb{R}$, signals are pre-processed. Pre-processing consists of frame blocking of the incoming streams to epochs of constant length w with constant time-shift s . Each epoch is a $(N + 1) \times (w)$ matrix, where N is the number of EEG electrodes and $N + 1$ is the N -dimensional EEG signal appended by the ECG signal.

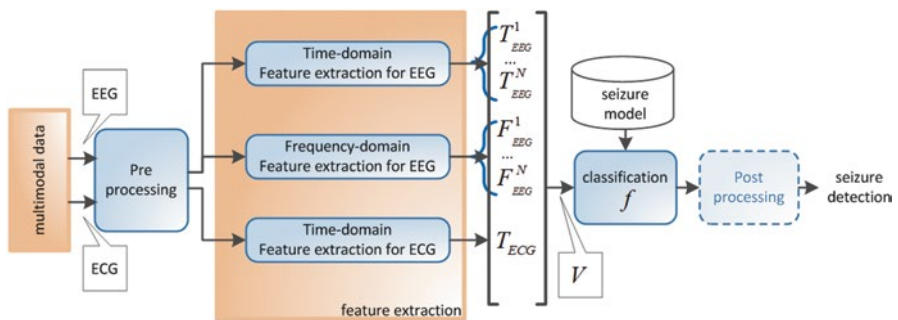


Fig. 14.1 Block diagram of the EEG-ECG based online seizure detection architecture

After pre-processing, the extracted epochs are in parallel processed by time-domain and frequency-domain feature extraction algorithms separately for the N -dimensional EEG and the 1-dimensional ECG signals. In particular, each of the N -dimensions of the EEG signal are processed by time-domain and frequency-domain feature extraction algorithms for EEG, while the ECG signal is processed by time-domain feature extraction algorithms (based on heart rate estimation) dedicated for electrocardiogram, as shown in the block diagram of Fig. 14.1. The extracted time-domain and frequency-domain features for the EEG, $T^i \in \mathbb{R}^{|T_{EEG}^i|}$ and $F^i \in \mathbb{R}^{|F_{EEG}^i|}$, with $1 \leq i \leq N$, and the ECG signal, $T_{ECG} \in \mathbb{R}^{|T_{ECG}|}$, are afterwards concatenated to a single feature vector $V \in \mathbb{R}^{N \cdot (|T_{EEG}^i| + |F_{EEG}^i|) + |T_{ECG}|}$ representing each epoch, as shown in Fig. 14.1. The extracted sequences of feature vectors, V , are short-time parametric representations of the EEG and ECG signals representing the time and spectral characteristics of the multimodal signals. This sequence of feature vectors is afterwards used as input to a classification model in order to assign a class label (seizure class or non-seizure class) to each of the vectors, i.e. to the corresponding time-intervals of the vectors.

During the training phase of the seizure detector, a dataset of feature vectors with known class labels (labeled manually by medical experts) is used to train a binary model M (two classes: seizure vs. non-seizure) using a classification algorithm f . At the test phase the existing seizure model, M , is used in order to decide for each epoch's feature vector, V , the corresponding class using the same classification algorithm, f , as in the training phase. Thus, for each epoch i a binary label d_i , i.e. seizure or not, is decided as:

$$d_i = f(V_i, M) \quad (1)$$

and the sequence of incoming EEG-ECG data is decomposed to time-intervals of seizure or clear (non-seizure) recordings. Post-processing of the automatically detected labels can be performed for improving the performance of the architecture.

During pre-processing the time-synchronized EEG and ECG recordings were frame blocked to epochs of 1-s length, without time-overlap between successive epochs. For each epoch, time-domain and frequency domain features were extracted separately for the each of the 21 EEG channels and the ECG channel.

In particular, each of the EEG channels was parameterized using the following features: time-domain features: minimum value, maximum value, mean, variance, standard deviation, percentiles (25 %, 50 %-median and 75 %), interquartile range, mean absolute deviation, range, skewness, kurtosis, energy, Shannon's entropy, logarithmic energy entropy, number of positive and negative peaks, zero-crossing rate, and frequency-domain features: Sixth order autoregressive-filter (AR) coefficients, power spectral density, frequency with maximum and minimum amplitude, spectral entropy, delta-theta-alpha-beta-gamma band energy, discrete wavelet transform coefficients with mother wavelet function Daubechies 16 and decomposition level equal to eight, thus resulting to a feature vector of dimensionality equal to 55 for each of the 21 EEG channels, i.e. 1155 in total.

The ECG channel was parameterized using the following features: the heart rate absolute value and variability statistics of the heart rate, i.e. minimum value, maximum value, mean, variance, standard deviation, percentiles (25 %, 50 %-median and 75 %), interquartile range, mean absolute deviation, range, thus resulting to a feature vector of dimensionality equal to 12. The heart rate estimation was based on Shannon energy envelope estimation for R-peak detection algorithm, implemented as in [2]. The dimensionality of the overall feature vector V is $1155 + 12 = 1167$.

The computed feature vectors V , one for each EEG-ECG epoch were used to train binary seizure detection models, M . Specifically, we compared the support vector machines (SVMs), implemented with the sequential minimal optimization method and polynomial kernel function, a two-layered back propagation multilayer perceptron (MLP) neural network, the k -nearest neighbour (IBK) algorithm and the C4.5 decision tree. All online seizure detection models were implemented using the WEKA machine learning toolkit software [3].

During the test phase, the EEG and ECG recordings are pre-processed and parameterized as in training. The SVM seizure detection model, M , is used to label each of the incoming EEG-ECG epochs as seizure or clear (non-seizure). In the present evaluation, no post-processing algorithm was applied on the estimated epoch-based results. The seizure detection results with respect to the use of N -best features after feature ranking are illustrated in the following figure (Fig. 14.2).

As can be seen in the evaluation results, for the three evaluated subjects the use of subset of features reduces the precision of the seizure detector. However, the exclusion of the approximately 30 % worst features still offers performance comparable to the best achieved and in combination with the reduction of the computational load of the detection architecture (both in the feature extraction stage and the classification stage) is a valuable solution for the online version.

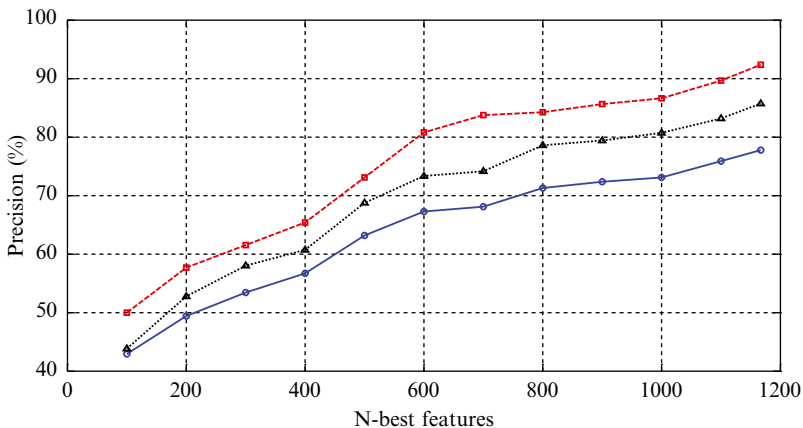


Fig. 14.2 Seizure detection precision (in percentages) for different number of features (N -best) for three subjects

Except online seizure detection, within the ARMOR framework an online alpha rhythm detection tool was developed. The online alpha rhythm detection is a simple algorithmic implementation designed to test the end-to-end system’s functionality and show in a relatively easy way its operation. It is designed in a similar way with the seizure detector, i.e. there is an incoming stream signal which is processed by a signal processing methodology through an online procedure that extracts a specific event (in this case alpha rhythms).

The algorithm estimates if the ratio of the energy in the alpha frequency band to the total signal energy exceeds a previously specified threshold, estimated either from offline analysis or empirically determined. The streamed data are also processed in windowed segments. For each segment, the detector returns a flag, true or false, regarding the existence or not of alpha rhythm. An alarm is sent if three consecutive epochs are detected with big alpha components, in order to reduce the system’s alarm load.

14.3 DSMS Implementation

One way of performing an online analysis of sensor data is a multi-layer Data Mining Model. These models are divided into four layers: Data collection, Data management, Event processing layer, Data mining service layer. The ARMOR Online platform consists of a similarly structured system whose components are developed by several partners. The data collection layer is performed from the sensors and using the xAffect tool is streamed to a StreamInsight application [4]. The streamed data are processed with a previously specified and parameterised algorithm, before the detected events are extracted (Fig. 14.3).

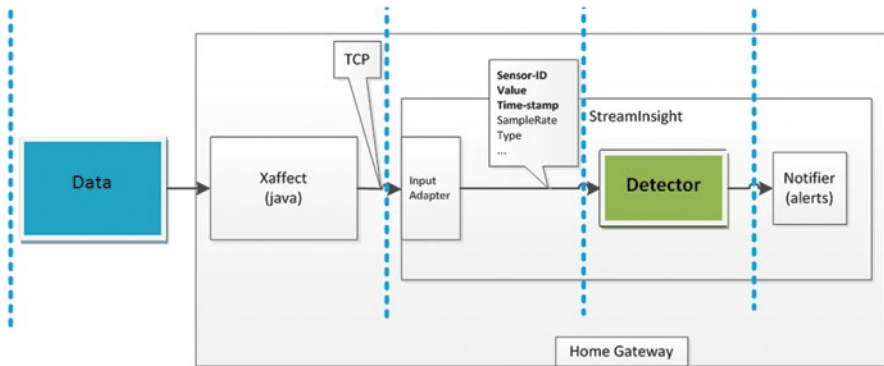


Fig. 14.3 Interface between different data management modules

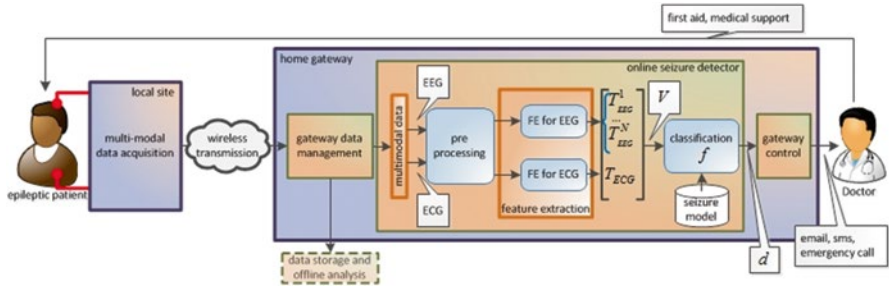


Fig. 14.4 Block diagram of the online seizure detection module

The seizure detection module includes the EEG and ECG recordings, while EMG and EOG, which are mainly used for movement/artifact detection, were not integrated. EMG and EOG were not included in the online seizure detection architecture in order to avoid further increasing the feature vector dimensionality and since the literature review showed limited use of these modalities with no significant advantages in online detection. The block diagram of the seizure detection architecture is illustrated above (Fig. 14.4).

The captured multimodal data (EEG and ECG) are wirelessly transmitted by a wearable solution to a local gateway for online processing. During online seizure detection, the EEG and ECG signals are initially pre-processed.

Real time applications are characterized by a limited amount of time available to process the data stream. The real time processing interval is the time necessary to process each data stream frame which corresponds to 1 s of data. If this interval is bigger than 1 s then the system will eventually reach memory limits and the processing of additional events will be delayed or postponed. Thus for each data block of 1 s the processing time should be less than a second. In our study we tested the performance of the detection algorithm for several numbers of EEG electrodes. The data stream was framed with non overlapping windows of 1 s. In Table 14.1 the average real time processing interval for each set of electrodes is shown. Each experiment was performed using the depicted number of electrodes in addition to the ECG signal.

The core of each of our detection algorithms is implemented in Mathworks' Matlab environment. Matlab offers many features that make the design of an algorithm more efficient and effective as well as shortening the time needed for the development of the algorithm. An online application on the other hand requires a Data Stream Management System, which offers the ability to process online data in a time and memory efficient way. The system used in our applications is StreamInsight from Microsoft.

To introduce our algorithms to the Microsoft's StreamInsight environment, the .Net compiler and API of Mathworks are used. In more detail, Mathworks provide a compiler for .Net packages. The result of this procedure is a set of libraries

Table 14.1 Real time processing for several numbers of EEG electrodes

Number of electrodes	Real time processing
19	0.977
17	0.874
15	0.776
13	0.681
11	0.588

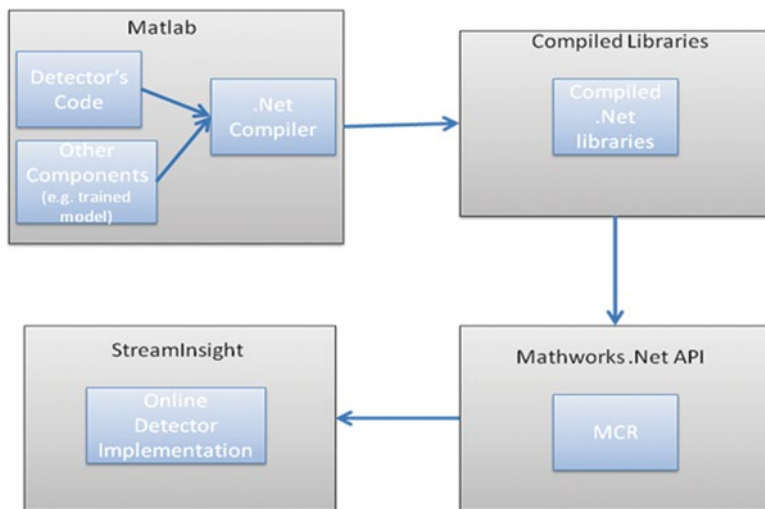


Fig. 14.5 Development scheme for using Matlab implementations of detectors in Microsoft StreamInsight

containing the algorithms and all the necessary components (functions, data such as training models etc.) necessary for each algorithm.

Those compiled libraries can be accessed by a StreamInsight application using the API that Mathworks provides. This API is the Matlab Compiler Runtime, in our case the 8.1 version 32 bit, which provides all the necessary components in order to use the algorithm in a stream application.

The figure below shows the pipeline followed when an online algorithm, designed in Matlab, is introduced in a StreamInsight application (Fig. 14.5).

It should be noted that each algorithm was designed to operate with segments of the data, as is necessary in an online application. In order to pre-process the streamed data, before having them processed with the detection algorithm, a set of StreamInsight tools were used. Depending on the procedure followed by the detection algorithm the streamed data should be aggregated or transformed in matrices whose rows and columns correspond to channels and samples of each segment,

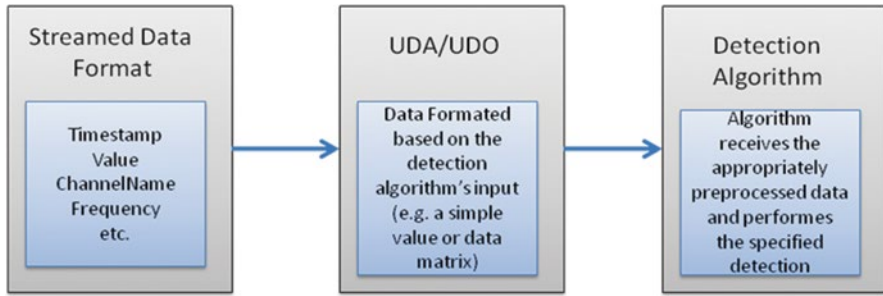


Fig. 14.6 StreamInsight's extensibility framework for detection algorithm's integration

respectively. In more detail for our alpha rhythm detection algorithm, an aggregation procedure of the segmented data should be followed. For every segment of the streamed data, the percentage of the Alpha band's energy relatively to the whole energy of the signal is returned. This is performed by using a User Defined Aggregate (UDA), a service provided by the StreamInsight framework. The detection procedure is performed after this step. In the case of the seizure detector for each time window of the streamed data, a matrix should be formed. Each row of this matrix contains the data values for one channel during the specific time window, whereas the columns of this matrix contain the values for all channels for a specific time point. This transformation is possible due to StreamInsight's service of User Defined Operator (UDO).

The figure above shows the procedure followed in order to adapt the streamed data to the input format of a detection algorithm (Fig. 14.6).

14.4 Conclusions

We presented algorithmic methodologies for online analysis of medical data. These methodologies serve as tools for online analysis of streaming data operated by a DSMS system. Both the online (real-time) processing of the medical data and the management of them are crucial when monitoring patients, and especially for the case of chronic diseases such as epilepsy. In contrast to offline analysis where performance is the key, when designing and implementing online tools there is a trade-off between tool's performance and computational complexity. Thus, fine-tuning of the tools in order to meet real-time processing demands with acceptable performance, i.e. accuracy in online detection of events of interest, is essential.

References

1. Mporas I, Tsirka V, Zacharaki EI, Koutroumanidis M, Megalooikonomou V (2014) Online seizure detection from EEG and ECG signals for monitoring of epileptic patients. In: 8th Hellenic conference on artificial intelligence (SETN 2014), Ioannina
2. Sabarimalai MM, Soman KP (2012) A novel method for detecting R-peaks in electrocardiogram (ECG) signal. *Biomed Signal Process Control* 7(2):118–128
3. Witten HI, Frank E (2005) *Data mining: practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco
4. Ali M et al (2009) Microsoft CEP server and online behavioral targeting. In: VLDB 2009 (demonstration)