Fast Classification of Electrocardiograph Signals via Instance Selection

Krisztian Buza, Alexandros Nanopoulos, Lars Schmidt-Thieme
Information Systems and Machine Learning Lab (ISMLL)
University of Hildesheim, Germany
{*buza,nanopoulos,schmidt-thieme*}*@ismll.de*

Julia Koller
Semmelweis University
Budapest, Hungary
julcsi05@freemail.hu

***Abstract*—In clinical practice, electrocardiographs (ECG) are used in various ways. In the most simple case, directly after the ECG has been recorded, the doctor analyses it and makes the diagnosis. In other cases, e.g. when the abnormality can only be observed occasionally, at a previously unknown time, the ECG is being recorded continuously. Fast automatic recognition of abnormalities of ECG signals may substantially support doctors' work in both cases: either by immediately displaying a warning or calling the emergency service in case of danger or by pointing to the abnormal parts of a long ECG-signal in order to support analysis and diagnosis. In this paper, we focus on the (semi-)automated recognition of abnormal ECG signals. We formulate the task as a time-series classification problem, point out that state-of-the-art solutions are capable to solve this problem with a high accuracy. The recognition time is, however, crucial in our case. Therefore, as major contribution, we aim at speeding up the recognition by a new instance selection technique. We describe this technique and discuss its theoretical background. In our experiments on publicly available real ECG-data, we empirically evaluate our approach and show that it outperforms a state-of-the-art instance selection technique.**

***Keywords*-electrocardiograph (ECG), time-series classification, scaling, instance selection, hubs**

I. Introduction

Electrocardiographs (ECG) are used in various ways in clinical practice: in the most simple case, directly after the ECG has been recorded, the doctor analyses it and makes the diagnosis. In other cases, due to the general health state of the patient or when the abnormality can only be observed at a previously unknown time (in some types of arrhythmias and ischemias), the ECG signal is being recorded *continuously* for a longer time period (intensive care monitoring or monitoring with a mobile device called Holter monitor).

In such cases, automatic recognition of abnormalities in ECG signals may substantially support doctors' work. When an out-patient is wearing a mobile ECG recorder, and this device detects serious abnormalities, it can warn the patient or call the emergency service automatically. If a nurse takes care for several patients and the ECG signal becomes abnormal, the ECG recording device displays a warning so that the doctor can be called in advance. Recognizing a disease *soon* and *accurately*, either by human experts or automatically, is a difficult task: a retrospective study [1] showed that, e.g., the infants admitted to a neonatal intensive care unit had abnormal heart beating patterns 24 hours before the doctor diagnosed them with sepsis.

When an ECG signal is recorded for one day for an out-patient, the record reflects approximately one hundred-thousand heart beats. Therefore, deep analysis of the *entire* signal, due to its length, is usually impossible by human experts. Rather, the doctor focuses on the most important parts of the signal, which can be positions where an event happens (something changes) or abnormalities appear. While the ECG is being recorded with a mobile device, the patient can press a button in specific cases such as sickness, going to bed, taking pills, etc. "A special mark will be then placed into the record so that the doctors or technicians can quickly pinpoint these areas when analyzing the signal."[1] Some abnormalities, however, may be left unnoticed by the patient and therefore no marking points to the corresponding parts of the signal. The entire signal can be scanned and automatically analyzed by computers, that produce suggestions to medical experts for abnormal parts of the signal. Additionally, the system can recognize the disease and in which part of the patient's heart it happened by detecting in which lead of the ECG the disease is expressed.[2] Therefore, the approach we describe can be considered as semi-automated, because it capitalizes on automated recognition models that support human experts' diagnostic work.

As ECG signals can be considered as time series, the task can be formulated as a time-series classification problem, for which state-of-the-art solutions are based on machine learning. A recognition model, called *classifier*, is constructed based on previously collected data and evidence (such as which data corresponds to which disease, where are the symptoms of that disease expressed in the data). Although state-of-the-art classifiers are able to solve the task of recognition with high accuracy, the quality depends on the available data and evidence. In general: the more data is used to construct the classifier, the better the recognition is.

As the accuracy is crucial in medical applications, this strongly motivates the usage of very large collections of

[1] http://en.wikipedia.org/wiki/Holter_Monitor

[2] From our point of view, each lead of the ECG is a signal that reflects the electrical activity of a certain part of the heart, different leads correspond different parts. See also: http://en.wikipedia.org/wiki/Electrocardiography

data and evidence. When doing so, however, both the time required to construct the model and the recognition time can be very high. As ECG is a medical instrument used in (almost) all hospitals world-wide, and one single recording (of some hours) already contains ten-thousands of heartbeats, the amount of potentially available data is huge, which can lead to intractably high recognition times. Nearest-neighbor classifiers, that have been shown to be competitive, if not superior, to many state-of-the-art time-series classification methods [2], [3], [4], are especially affected by the aforementioned problem: whenever we want to detect abnormalities in a new ECG signal, nearest-neighbor methods search the available data for ECG signals that are most similar to the new one, and in case of very large collections, this search can take a long time and therefore the recognition time can be intractably high. In order to alleviate this problem, various speed-up techniques have been introduced, such as indexing [5], [6], lower bounding [7] or aggregation [8]. Having the common trade-off between quality and runtime in our minds, we aim at speeding-up the recognition without or with minimal loss of quality.

In this paper, we propose a new technique to speed-up the classification of ECG signals. The proposed technique is complementary to the above ones, as it can be applied together with them. Our approach is based on the recently observed phenomenon of hubness [9], [10], which states that some few ECG signals tend to be much more frequently nearest neighbors than the remaining ones. Our approach selects the most important ECG signals from the available data, and uses only the selected ECG signals for the classification of new ECG signals, which leads to substantial speed-up. In our experiments on publicly available real ECG-data, we evaluate our approach and show that it outperforms a state-of-the-art instance selection technique.

This paper is organized as follows: in Section II we review related work, we formally define the problem of ECG signal classification in Section III, we describe our approach in Section IV and present our experimental results in Section V before concluding in Section VI.

II. Related Work

Semi-automatic detection of irregularities in ECG signals has been explored by several researchers. An early approach was proposed by Bortolan and Willems who used neural networks for ECG classification [11]. Olszewski [12] examined feature extraction techniques for ECG classification, Melagni and Bazi used support vector machines [13], Syed and Chia presented an approach based on approximately conserved heart rate sequences [14], while Keogh et al. [15] applied a similarity-based, *unsupervised*, nearest-neighbor-like method in order to find "unusual" (and therefore likely irregular) segments of ECG signals. In contrast to [15], we formulate the problem as *supervised* classification task, which allows not only for the detection of some "unusual" segments of the signal, but also for the detection of the *type* of abnormality and many other tasks, like finding the location, where e.g. an infarct happened in the patient's body. Therefore, our approach is more generic. Furthermore, we use the Dynamic Time Warping (DTW) distance instead of the Euclidean distance used in [15]. In contrast to all the aforementioned works, we focus on *instance selection* in order to speed-up the classification of ECG signals.

As we consider the detection of abnormal segments of ECG signals as a time-series classification task, we review the related work in time-series classification domain. The intensive research efforts of the last decades resulted in a plethora of different approaches ranging from neural [16] and Bayesian networks [17] to genetic algorithms, support vector machines [18] and frequent pattern mining [19]. Nevertheless, recent research has shown that the simple nearest-neighbor (1-NN) classifier using Dynamic Time Warping (DTW) [20] as distance measure is "exceptionally hard to beat" [3]. Due to its good performance, this method has been examined in depth (a thorough summary of results can be found at [21]) with the aim to improve its accuracy [22], [23], [24] and efficiency [25].

Attempts to speed up DTW-based nearest neighbor (NN) classification fall into 4 major categories: i) speed-up the calculation of the distance of two time series, ii) reduce the length of time series, iii) indexing, and iv) instance selection.

If we implement DTW in the simple, straightforward way, the comparison of two time series of length l requires the calculation of the entries of an $l \times l$ matrix using dynamic programming, and therefore each comparison has a complexity of $O(l^2)$. A simple idea is to limit the warping window size, which eliminates the calculation of most of the entries of the DTW-matrix: only a small fraction around the diagonal remains. Ratanamahatana and Keogh [21] showed that such reduction does not negatively influence classification accuracy, instead, it leads to more accurate classification. More advanced scaling techniques include lower-bounding, like LB_Keogh [7].

Another way to speed-up time series classification is to reduce the length of time series by aggregating consecutive values into a single number [25], [8]. This makes processing faster by reducing the overall length of time series.

Indexing [5], [6] aims at *quickly* finding the time series that are most similar to the time series to be classified. Due to the "filtering" step that is performed by indexing, the execution time for classifying new time series can be considerable for large time-series data sets, since it can be affected by the significant computational requirements posed by the need to calculate DTW distance between the new time-series and several time-series in the training data set ($O(n)$ in worst case, where n is the size of the training set). For this reason, indexing can be considered complementary to instance selection, since both these techniques can be applied to improve execution time.

Instance selection (also known as *numerosity reduction* or *prototype selection*) aims at discarding most of the training time series while keeping only the most informative ones, which are then used to classify unlabeled instances. While instance selection is well explored for general nearest-neighbor classification, see e.g. [26], [27], [28], [29], [30], [31], there are just a few works for the case of time series. Xi et al. [32] present the FastAWARD approach and show that it outperforms state-of-the-art, general-purpose instance selection techniques applied for time series.

FastAWARD first calculates the optimal warping window size for DTW, then it follows an iterative procedure for discarding time series: in each iteration, the rank of all the time series is calculated and the lowest ranked time series is discarded. Thus, each iteration corresponds to a particular number of kept time time series. Xi et al. argue that the optimal warping window size depends on the number of kept time series. Therefore, FastAWARD calculates the optimal warping window size for each number of kept time series.

FastAWARD follows some decisions whose nature can be considered as ad-hoc (such as the application of an iterative procedure or the use of tie-breaking criteria [32]). Conversely, our approach is more principled: in particular, we generalize FastAWARD by being able to use several formula for scoring instances. We will explain that the suitability of such formula is based on the hubness property that holds in most time-series data sets. The presence of hubs, i.e., that some few objects tend to be much more frequently nearest neighbors than the remaining ones, has been observed for many natural and artificial networks, such as protein-interaction networs or the internet [33], it has been used in context of clustering [34], and in order to make classification algorithms more accurate [9], [10]. In this paper we exploit hubness for instance selection for time series classification algorithms. In our previous work [35] we proved that instance selection is an NP-complete problem and discussed coverage of the selected instances. Here, in contrast, we focus on hub-based instance selection for electrocardiography. Furthermore, we show that the iterative procedure of FastAWARD is not a well-formed decision, since its large computation time can be saved by ranking instances only once. Furthermore, we observed the warping window size to be less crucial, and therefore we simply use a fixed window size for our approach (that outperforms FastAWARD which uses adaptive window size).

III. Definitions and Problem formulation

In order to allow for different recognition tasks related to ECG signals, we define the problem in a generic way.

A *time series* x of length l, that represents a segment of an ECG signal in our case, is a sequence of real numeric values: $x = (v_1, ..., v_l)$. We denote the set of all considered time series as T. We are given some groups (subsets of T) of time series. These groups are called *classes*, and they are denoted as $C_1, ..., C_m$. Each time series $x_i \in T$ belongs to one of the classes, however, for some x_i, it is unknown to *which* class they belong: $\forall x_i \in T : \big((x_i \in C_1) \vee (x_i \in C_2) \vee ... \vee (x_i \in C_m)\big) \wedge \big(x_i \in C_j \Rightarrow x_i \notin C_k, k \neq j\big)$. A *labeled dataset* $\mathcal{D} = \{(x_i, c(i)\}_{i=1}^{n}$ consists of n time series together with their class labels $c(i)$. (The class label $c(i)$ shows the class of time series x_i, e.g. $c(i) = 2 \Leftrightarrow x_i \in C_2$.) The time series in a dataset $\mathcal{D}$ are also called *instances* of $\mathcal{D}$. Time series in a labeled dataset are called *labeled time series*. Other time series for which their classes are unknown, are called *unlabeled time series* respectively.

Next, we define the *time series classification problem*: we are given a labeled dataset $\mathcal{D}$, the task is to find a function $g(x) : T \rightarrow \{C_1, ..., C_m\}$ that is able to assign new, unlabeled time series to their classes. The function g is called *classifier*. More advanced classifiers, besides assigning an unlabeled time series to one of the classes, also output a likelihood (or probability) for each class.

By different concrete choices of the classes, the above definition allows for various recognition tasks related to ECG:

1) If we want to find *when* abnormalities appear in a long (e.g. 24 hours) ECG-recording, each time series in the above definition corresponds to a short segment (e.g. 1 heartbeat) of the long signal. In this case, there are two classes: *normal* signal segments belong to the first class, while *abnormal* signal segments belong to the second class. Whenever a segment is recognized as abnormal, this is a candidate that may be examined by human experts more accurately. If the classifier outputs a probability for each segment of being abnormal, the segments can be ranked according to this probability so that the most serious segments can be checked first by the human expert.
2) Another task is to find *where*, in which part of the patient's heart, the abnormality, like an infarct, happened. This is possible as signals of different leads of the ECG correspond to the electrical activity of different parts of the heart (see Footnote 2). Therefore, the task is to find in which signal the abnormality is expressed. One class corresponds to the expression of the abnormality, while the signals where the abnormality is not expressed, belong to the other class.
3) If we aim at recognizing the type of abnormality, we can define several classes, one class for each disease and an additional class for the normal signal.

Of course, we can combine all the above tasks (by e.g. using several classifiers), so that the result of the automatic recognition is a list of items describing *when* (at what time, at which position of the signal), *where* (at which part of the patient's heart) and *what* kind of abnormality was likely to happen. Such lists of abnormality-candidates can considerably support human expert's diagnostic work.

Require: Time-series dataset $\mathcal{D}$, Score Function f, Number of selected instances N
Ensure: Set of selected instances (time series) $\mathcal{D}'$

1: Calculate score function $f(x)$ for all $x \in \mathcal{D}$
2: Sort all the time series in $\mathcal{D}$ according to their scores $f(x)$
3: Select the top-ranked N time series and return the set containing them

Figure 1. Outline of Hub-based Instance Selection, our instance selection approach.

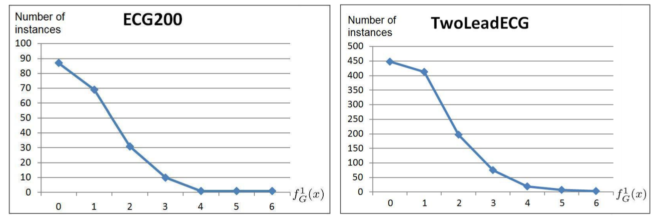


Figure 2. Distribution of $f_G^1(x)$ for the ECG200 and TwoLeadECG datasets. The horizontal axis correspond to the values of $f_G^1(x)$, while on the vertical axis we see how many instance have that value.

IV. Our approach: Hub-based Instance Selection

In order to be able to support doctor's work by automatic ECG analysis, classifiers must be able to perform recognition within an acceptable time. As already described in the Introduction, due to the huge amount of available ECG data, this is a challenge. In general, a large collection of data is required for accurate recognition and therefore reducing data in a naive way (e.g. by selecting time series randomly) could substantially harm accuracy. Instead, one should focus on selecting the most representative time series, the ones, that are most important for the recognition.

Our instance selection approach first assigns a score to each instance (instances are time series in our case). Then it selects the ones having the highest scores (see Figure 1). In this section, we examine how to develop appropriate score functions by exploiting the property of hubness.

A. The Hubness Property

In order to develop a score function that selects representative instances for nearest-neighbor time-series classification, we have to take into account the recently explored property of hubness [10]. This property states that for data with high (intrinsic) dimensionality, as most of the time-series data[3], some objects tend to become nearest neighbors much more frequently than others.

In order to express hubness in a more precise way, for a (time series) dataset $\mathcal{D}$ we define the *k-occurrence* of an instance (time series) $x \in \mathcal{D}$, denoted $f_N^k(x)$, as the number of instances of $\mathcal{D}$ having x among their k nearest neighbors. With the term *hubness* we refer to the phenomenon that the distribution of $f_N^k(x)$ becomes significantly skewed. We can measure this skewness, denoted by $\mathcal{S}_{f_N^k(x)}$, with the standardized third moment of $f_N^k(x)$:

$$\mathcal{S}_{f_N^k(x)} = \frac{E[(f_N^k(x) - \mu_{f_N^k(x)})^3]}{\sigma^3_{f_N^k(x)}} \tag{1}$$

where $\mu_{f_N^k(x)}$ and $\sigma_{f_N^k(x)}$ are the mean and standard deviation of $f_N^k(x)$. When $\mathcal{S}_{f_N^k(x)}$ is higher than zero, the corresponding distribution is skewed to the right and starts presenting a long tail.

In the presence of labeled data, we distinguish between *good hubness* and *bad hubness*: we say that the instance y is a *good (bad) k-nearest neighbor* of the instance x if (i) y is one of the k-nearest neighbors of x, and (ii) both have the same (different) class labels. This allows us to define *good (bad) k-occurrence* of a time series x, $f_G^k(x)$ (and $f_B^k(x)$ respectively), which is the number of other time series that have x as one of their good (bad) k-nearest neighbors. For time series, both distributions $f_G^k(x)$ and $f_B^k(x)$ are usually skewed, as it is exemplified in Figure 2, which depicts the distributions of $f_G^1(x)$ for the ECG200 and the TwoLeadECG datasets (we will describe the datasets in Section V). As shown, the distributions have long tails, in which the good hubs occur.

We say that a time series x is a good (bad) hub, if $f_G^k(x)$ (and $f_B^k(x)$ respectively) is exceptionally large for x. For the nearest neighbor classification of time series, the skewness of good occurrence is of major importance, because a few time series (i.e., the good hubs) are able to correctly classify most of the other time series. Therefore, it is evident that instance selection should pay special attention to good hubs.

B. Score functions based on Hubness

1) Good 1-occurrence score: In the light of the previous discussion, our approach (see Figure 1) can use scores that take the good 1-occurrence of an instance x into account. Thus, a simple score function that follows directly is the *good 1-occurrence score* $f_G(x)$:

$$f_G(x) = f_G^1(x) \tag{2}$$

When there is no ambiguity, we omit the upper index 1.

2) Relative score: While x is being a good hub, at the same time it may appear as bad neighbor of several other instances. Thus, we also consider scores that take bad occurrences into account too. This leads to scores that relate the good occurrence of an instance x to either its total occurrence or to its bad occurrence. For simplicity, we focus on the following *relative score*, however, other variations can be used too. *Relative score* $f_R(x)$ of a time series x is the fraction of good 1-occurrences and total occurrences plus one (plus one in the denominator avoids division by zero):

[3] In case of time series, consecutive values are strongly interdependent, thus instead of the length of time series, we have to consider the *intrinsic* dimensionality [9].

$$f_R(x) = \frac{f_G^1(x)}{f_N^1(x)+1} \quad (3)$$

3) Xi's score: Interestingly, $f_G^k(x)$ and $f_B^k(x)$ allows us to interpret the ranking criterion of Xi et al. [32], by expressing it as another form of score for relative hubness:

$$f_{Xi}(x) = f_G^1(x) - 2f_B^1(x) \quad (4)$$

V. Experiments

We experimentally examine the performance of our approach, Hub-based Instance Selection, with respect to classification accuracy and execution time. Instead of presenting a complete, ready-to-use application, we focus on analyzing our approach by comparing against FastAWARD, a state-of-the-art instance selection technique for time series [32].

A. Datasets

We performed experiments on two ECG datasets, ECG200 and TwoLeadECG from the dataset collection used in [3]. As this is one of the most frequently used publicly available collections of labeled time series datasets, we assist comparability and reproducibility with this choice.[4]

1) ECG200: The ECG200 dataset contains 200 ECG signals, each of them consisting of 96 measured values (each time series reflects 1 heartbeat). Out of the 200 time series, 133 are labeled as normal while the remaining 67 are labeled as abnormal [12]. Time series are segments of a long ECG signal, therefore the experiments on this dataset simulate the scenario when the automatic recognition system is supposed to support the doctor while she or he is searching for abnormal parts of a long ECG signal (first task listed in Section III).

2) TwoLeadECG: This dataset contains 1162 ECG signals of length 82 (each time series reflects 1 heartbeat). In the TwoLeadECG dataset, two different leads of the ECG are considered, each signal originates from one out of these two leads. An abnormality, infarct, is expressed with different intensity in these both leads.[5] As the two classes correspond to two different leads of the ECG, experiments on this dataset simulate the second scenario listed in Section III, when we aim at finding in which part of the patient's heart the abnormality occurred.

B. Experimental Protocol

We performed 10-fold-cross validation. We divided the data into 10 splits, out of which 1 was reserved as *test* data while the remaining 9 splits constituted the so called *training* data. We selected the most representative instances (time series) from the training data and then we constructed the recognition system (i.e. trained the classifier) using these selected instances. While the instances are being selected and the classifier is being constructed (trained), the data in the *test* split is unknown both for the instance selection algorithm and the classifier. At the end, the classifier is used to determine the class labels of the test data, which is then compared to the true class labels in order to allow for quantitative evaluation of the quality of the classifier. The whole process of instance selection, classifier construction and evaluation is repeated 10 times, in each of the 10 rounds a different data split serves as test data.

In our experiments we used two instance selection algorithms: (i) our approach (see Figure 1) with the score functions in section IV-B and (ii) the competitor, FastAWARD. We refer to our approach as "Hub-based Selection".

Both for our approach and FastAWARD, as classifier, we used 1-NN, i.e. the nearest neighbor algorithm with $k = 1$ (i.e. we considered always the first nearest neighbor).

As distance function, both for the classifier and for the calculation of the scores, we used Dynamic Time Warping (DTW). One of the parameters of DTW is the size of warping window. In contrast to FastAWARD, which determines the optimal warping window size r_{opt}, for our approach, Hub-based Selection, we set the warping-window size to a constant of 5%. (This selection is justified by the results presented in [21], which show that relatively small window sizes lead to higher accuracy.) In order to speed-up the distance calculations, we used the LB_Keogh lower bounding technique [7] both for Hub-based Selection and FastAWARD.

C. Results on Effectiveness

We first compare Hub-based Selection and FastAWARD in terms of classification accuracy (i.e. ratio of correctly recognized signals) that results when using the instances selected by these two methods. Table I presents the average accuracy and corresponding standard deviation for each data set, for the case when the number of selected instances is equal to 10% of the size of the training set. For both datasets, all variants of our approach (Hub-based Selection with all score functions) clearly outperform FastAWARD. Figure 4 shows the classification of some signals.

We also compared Hub-based Selection and FastAWARD in terms of the resulting classification accuracy for varying number of selected instances. Figure 3 illustrates that Hub-based Selection compares favorably to FastAWARD. In order to keep the presentation simple, we only present results for the case when we used $f_G(x)$ score function. We note, however, that we observed very similar tendencies for the other score functions.

Besides the comparison between Hub-based Selection and FastAward, what is also interesting to examine, is their relative performance compared to the case of using the entire training data (i.e., no instance selection is applied). For the ECG200 dataset, selecting 10% of the training data using our Hub-based Selection algorithm with $f_G(x)$, the accuracy is

[4] http://www.cs.ucr.edu/~eamonn/time_series_data/

[5] http://users.eecs.northwestern.edu/~hdi117/listfile/VLDB08_datasets.ppt

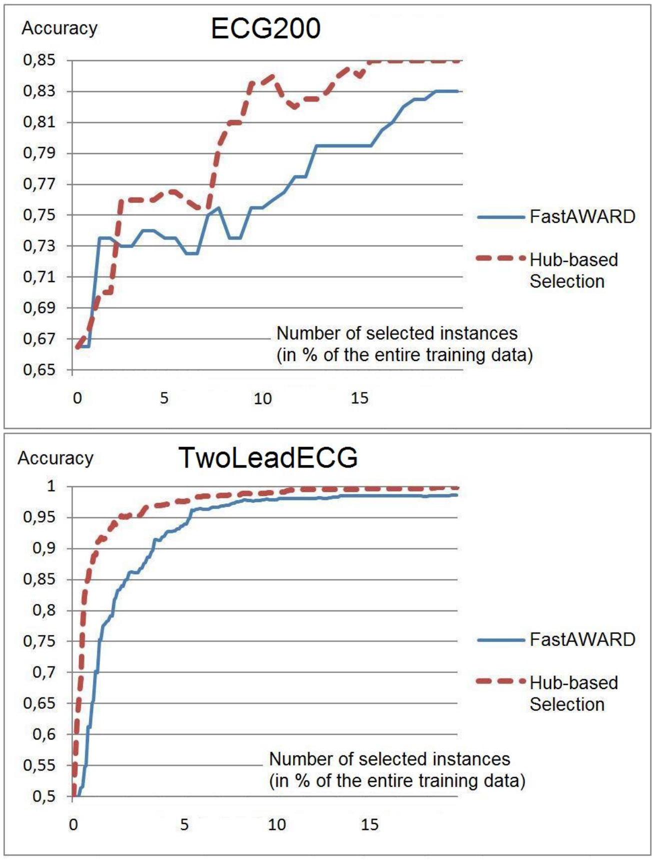


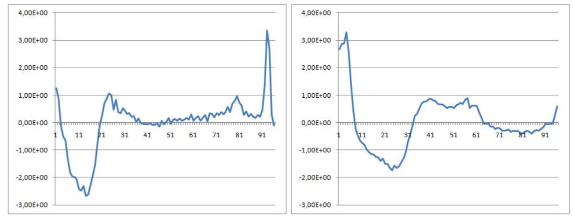
Figure 3. Accuracy as function of the number of selected instances (in % of the entire training data) for FastAWARD and Hub-based Selection with $f_G(x)$ on the ECG200 (top) and TwoLeadECG (bottom) datasets.

Table I
ACCURACY $\pm$ STANDARD DEVIATION FOR SELECTING 10 % OF THE ENTIRE TRAINING DATA WITH OUR APPROACH, CALLED HUB-BASED SELECTION, AND FASTAWARD (BOLD FONT: WINNER).

	ECG200	TwoLeadECG
FastAWARD	0.755±0.113	0.978±0.013
Hub-based Selection with $f_G(x)$	**0.835±0.090**	**0.989±0.012**
Hub-based Selection with $f_R(x)$	0.820±0.071	**0.989±0.012**
Hub-based Selection with $f_{Xi}(x)$	0.820±0.071	**0.989±0.012**

approximately 0.05 worse compared to the case of using the entire training data. For FastAWARD, however, this number is about 0.13. For TwoLeadECG, our approach wins against FastAWARD again with 0.01 vs. 0.02.

Next, we investigate the reasons for the presented difference between Hub-based Selection and FastAward. In Section IV-A, we identified the skewness of good k-occurrence, $f_G^k(x)$, as a crucial property for instance selection to work properly, since skewness renders good hubs to become representative instances. In our examination, we found that using the iterative procedure applied by FastAWARD, this skewness has a decreasing trend from iteration to iteration. Figure 5 exemplifies this by illustrating the skewness of



True label	2 (abnormal)		2 (abnormal)	
Classification results		Correct classification?		Correct classification?
FastAWARD	1 (normal)	NO	1 (normal)	NO
Hub-based Selection	2 (abnormal)	YES	2 (abnormal)	YES

Figure 4. Some signals from the ECG200 dataset, their true class labels and the class labels output by 1-NN after selecting instances with FastAWARD and our approach, Hub-based selection. Note that in most cases, both algorithms classified the signals correctly, the above examples aim at illustrating the differences. Also note that all variants of our approach (Hub-based with different score functions) agreed on the classification of these signals.

Table II
EXECUTION TIMES (IN SECONDS, AVERAGED OVER 10 FOLDS) OF INSTANCE SELECTION WITH OUR APPROACH, CALLED HUB-BASED SELECTION, AND FASTAWARD

Dataset	FastAWARD	Hub-based Selection with $f_G(x)$
ECG200	634	2
TwoLeadECG	12 946	45

$f_G^1(x)$ for the ECG200 and TwoLeadECG datasets as a function of iterations performed in FastAWARD. In order to quantitatively measure skewness we use the standardized third moment, see Equation 1. The reduction in the skewness of $f_G^1(x)$ means that FastAWARD is not able to identify representative instances in the end, since there are no pronounced good hubs remaining. Note that FastAWARD iteratively drops *bad* instances, the instances *remaining at the end* are considered as the selected ones that are used for classification, therefore, the reduction of skewness is crucial.

The above observation justifies that the reduced effectiveness of FastAWARD stems from its iterative procedure and not from its score function, $f_{Xi}(x)$ (Eq. 4). In the last row of Table I, we show our approach, Hub-based Selection with FastAWARD's score function, $f_{Xi}(x)$. This variant of Hub-based Selection, similarly to the other ones, clearly outperforms FastAWARD, which indicates the robustness of our approach with respect to the score function.

D. Results on Efficiency

In our analysis we focus on the computational time of two steps, in particular the (i) recognition time, i.e. time required to detect the classes of new instances, and the (ii) instance selection time, i.e. the time required to select the instances.

Assuming that we select the same number of instances, the recognition time is equal both for our approach and

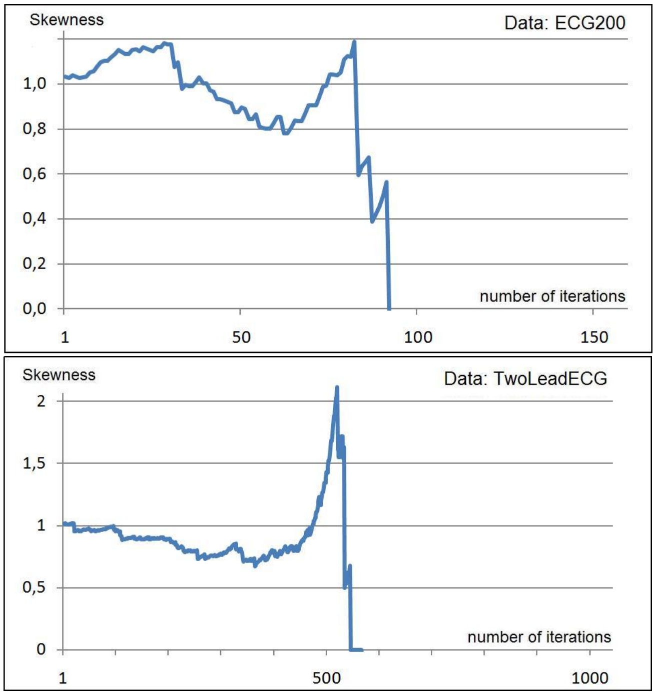


Figure 5. Skewness of the distribution of $f_G^1(x)$ as function of the number of iterations performed in FastAWARD for ECG200 and TwoLeadECG datasets. On the trend, the skewness decreases from iteration to iteration.

for FastAWARD, because in both cases we perform nearest neighbor classification using the same number of selected instances. As shown in the experiments, if we select the same number of instances, our approach, Hub-based Selection, achieves higher accuracy. If we aim at achieving the same accuracy as with FastAWARD, in case of our approach, it is sufficient to select less number of instances, which makes the classification faster.

Regarding the time required to select the instances, the computational complexity of Hub-based Selection depends on the calculation of the scores of the instances and on the selection of the top-ranked instances. Thus, for the examined score functions, the computational complexity of our approach is $O(n^2)$, n being the number of training instances, since it is determined by the calculation of the distance between each pair of training instances. For FastAWARD, its first step (leave-one-out nearest neighbor classification of the training instances) already requires $O(n^2)$ execution time. However, FastAWARD performs additional computationally expensive steps, such as determining the best warping-window size and the iterative procedure for excluding instances. For this reason, Hub-based Selection is expected to require reduced execution time compared to FastAWARD. This is verified by the results presented in Table II, which shows the execution time needed for Hub-based Instance Selection and instance selection with FastAWARD. As expected, our approach outperforms FastAWARD drastically.

VI. Conclusion

In this paper, we introduced a new instance selection approach in order to speed-up the classification of electrocardiograph signals. Allowing for a semi-automated diagnosis, in the clinical practice, our approach can support human experts' work by (i) quickly detecting abnormal segments of a long ECG signal, i.e. *when* the abnormality occurred, (ii) delivering suggestions regarding the disease, and (iii) finding in which lead of the ECG the abnormality is expressed, which can help in finding *where*, in which part of the heart of the patient, the abnormality occurred. We evaluated our approach on publicly available, real-world ECG data that allowed to simulate two of the aforementioned use-cases. In both cases, we found that our approach outperforms the state-of-the-art instance selection technique: our approach allows faster recognition at the same level of accuracy, and, more importantly, more accurate recognition at the same execution time.

Acknowledgments

We thank Eamon Keogh for making the datasets available and the Reviewers for their comments.

References

[1] M. Griffin and J. Moorman, "Toward the early diagnosis of neonatal sepsis and sepsis-like illness using novel heart rate analysis," *Pediatrics*, vol. 107, no. 1, p. 97, 2001.

[2] T. Rath and R. Manmatha, "Word image matching using dynamic time warping," in *Conference on Computer Vision and Pattern Recognition*, vol. 2. IEEE, 2003.

[3] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh, "Querying and mining of time series data: experimental comparison of representations and distance measures," in *Proceedings of the VLDB Endowment*, vol. 1, no. 2, 2008, pp. 1542–1552.

[4] E. Keogh, C. Shelton, and F. Moerchen, "Workshop and challenge on time series classification," 2007. [Online]. Available: http://www.cs.ucr.edu/~eamonn/SIGKDD2007TimeSeries.html

[5] K. Chakrabarti, E. Keogh, M. Sharad, and M. Pazzani, "Locally adaptive dimensionality reduction for indexing large time series databases," *Transactions on Database Systems*, vol. 27, pp. 188–228, 2002.

[6] D. Gunopulos and G. Das, "Time series similarity measures and time series indexing," in *Proc. SIGMOD International Conference on Management of Data*. ACM, 2001, p. 624.

[7] E. Keogh and C. Ratanamahatana, "Exact indexing of Dynamic Time Warping," *Knowledge and Information Systems*, vol. 7, no. 3, pp. 358–386, 2005.

[8] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*. ACM, 2003, pp. 2–11.

[9] M. Radovanovic, A. Nanopoulos, and M. Ivanovic, "Time-Series Classification in Many Intrinsic Dimensions," in *SIAM International Conference on Data Mining*, 2010, pp. 677–688.

[10] ——, "Nearest neighbors in high-dimensional data: The emergence and influence of hubs," in *Proceedigns of the International Conference on Machine Learning (ICML)*, 2009, pp. 865–872.

[11] G. Bortolan and J. Willems, "Diagnostic ECG classification based on neural networks," *Journal of Electrocardiology*, vol. 26, p. 75, 1993.

[12] R. Olszewski, "Generalized feature extraction for structural pattern recognition in time-series data," Ph.D. dissertation, School of Computer Science, Carnegie Mellon University, 2001.

[13] F. Melgani and Y. Bazi, "Classification of electrocardiogram signals with support vector machines and particle swarm optimization," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 5, pp. 667–677, 2008.

[14] Z. Syed and C.-C. Chia, "Computationally generated cardiac biomarkers: Heart rate patterns to predict death following coronary attacks," in *SIAM International Conference on Data Mining*, 2011.

[15] E. Keogh, J. Lin, A. Fu, and H. Van Herle, "Finding the unusual medical time series: Algorithms and applications," *Transactions on Information Technology in Biomedicine, Special Post-conference Issue "Mining Biomedical Data/CBMS2005"*, 2005.

[16] A. Kehagias and V. Petridis, "Predictive modular neural networks for time series classification," *Neural Networks*, vol. 10, no. 1, 1997.

[17] P. Sykacek and S. Roberts, "Bayesian time series classification," *Advances in Neural Information Processing Systems*, vol. 2, pp. 937–944, 2002.

[18] D. Eads, D. Hill, S. Davis, S. Perkins, J. Ma, R. Porter, and J. Theiler, "Genetic algorithms and support vector machines for time series classification," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 4787, 2002, pp. 74–85.

[19] P. Geurts, "Pattern extraction for time series classification," in *Principles of Data Mining and Knowledge Discovery (PKDD)*. Springer, 2001, pp. 115–127.

[20] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *Transactions on Acoustics, Speech and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.

[21] C. Ratanamahatana and E. Keogh, "Everything you know about dynamic time warping is wrong," in *SIGKDD Int'l. Workshop on Mining Temporal and Sequential Data*, 2004.

[22] K. Buza, A. Nanopoulos, and L. Schmidt-Thieme, "Time-Series Classification Based on Individualised Error Prediction," in *13th International Conference on Computational Science and Engineering*. IEEE, 2010, pp. 48–54.

[23] ——, "Fusion of Similarity Measures for Time Series Classification," in *6th International Conference on Hybrid Artificial Intelligence Systems (HAIS-2011)*. Springer, 2011.

[24] C. Ratanamahatana and E. Keogh, "Making time-series classification more accurate using learned constraints," in *SIAM International Conference on Data Mining*, 2004, pp. 11–22.

[25] E. Keogh and M. Pazzani, "Scaling up dynamic time warping for datamining applications," in *SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2000, pp. 285–289.

[26] D. Aha, D. Kibler, and M. Albert, "Instance-based learning algorithms," *Machine Learning*, vol. 6, no. 1, pp. 37–66, 1991.

[27] H. Brighton and C. Mellish, "Advances in instance selection for instance-based learning algorithms," *Data Mining and Knowledge Discovery*, vol. 6, no. 2, pp. 153–172, 2002.

[28] N. Jankowski and M. Grochowski, "Comparison of instance selection algorithms I. Algorithms survey," in *Proc. 7th International Conference on Artificial Intelligence and Soft Computing - ICAISC*, ser. LNCS, vol. 3070. Springer, 2004, pp. 598–603.

[29] M. Grochowski and N. Jankowski, "Comparison of instance selection algorithms. II. Results and comments," in *Proc. 7th International Conference on Artificial Intelligence and Soft Computing - ICAISC*, vol. 3070. Springer Verlag, 2004, pp. 580–585.

[30] H. Liu and H. Motoda, "On issues of instance selection," *Data Mining and Knowledge Discovery*, vol. 6, no. 2, pp. 115–130, 2002.

[31] R. Paredes and E. Vidal, "Learning prototypes and distances: A prototype reduction technique based on nearest neighbor error minimization," *Pattern Recognition*, vol. 39, no. 2, pp. 180–188, 2006.

[32] X. Xi, E. Keogh, C. Shelton, L. Wei, and C. Ratanamahatana, "Fast time series classification using numerosity reduction," in *Proceedings of the 23rd International Conference on Machine Learning*. ACM, 2006, pp. 1033–1040.

[33] A. Barabasi, *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*. Plume, 2003.

[34] N. Tomasev and M. I. M. Radovanovic, D. Mladenic, "The role of hubness in clustering high-dimensional data," in *15th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, ser. LNCS. Springer, 2011, vol. 6634, pp. 183–195.

[35] K. Buza, A. Nanopoulos, and L. Schmidt-Thieme, "INSIGHT: Efficient and Effective Instance Selection for Time-Series Classification," in *15th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, ser. LNCS. Springer, 2011, vol. 6635, pp. 149–160.