HRVFrame: Java-Based Framework for Feature Extraction from Cardiac Rhythm

Alan Jovic[1] and Nikola Bogunovic[1]

[1] Department of Electronics, Microelectronics, Computer and Intelligent Systems, Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, 10000 Zagreb, Croatia
{alan.jovic, nikola.bogunovic}@fer.hr

Abstract. Heart rate variability (HRV) analysis can be successfully applied to automatic classification of cardiac rhythm abnormalities. This paper presents a novel Java-based computer framework for feature extraction from cardiac rhythms. The framework called HRVFrame implements more than 30 HRV linear time domain, frequency domain, time-frequency domain, and nonlinear features. Output of the framework in the form of .arff files enables easier medical knowledge discovery via platforms such as RapidMiner or Weka. The scope of the framework facilitates comparison of models for different cardiac disorders. Some of the features implemented in the framework can also be applied to other biomedical time-series. The thorough approach to feature extraction pursued in this work is also encouraged for other types of biomedical time-series.

Keywords: heart rate variability; computer framework; nonlinear features

1 Introduction

Analysis of cardiac abnormalities usually starts with routine diagnostic procedure called electrocardiogram (ECG). Detection of anomalies in an ECG can indicate the existence of cardiovascular and other diseases [1]. Heart rate variability (HRV) analysis examines fluctuations in the sequence of cardiac interbeat (RR) intervals. In cardiac rhythm research there is a problem with the selection of features for optimal arrhythmia description, because of the infinite dimensionality of the feature space. Therefore, researchers base their work either on physiological explicability of the features or on proven mathematical properties of the time-series. Both of these approaches do not guarantee that the entire feature space had been searched.

HRV features can be grouped in several categories depending on the type of the time-series' analysis needed to be performed for their extraction. These include: linear, time-frequency, nonlinear features, and rhythm pattern analysis [2,3,4].

Linear time and frequency domain features include statistical properties of the heart rhythm and power spectral density estimates in clinically relevant frequency bands [2]. Time-frequency features are useful for finding nonstationarities in cardiac rhythm. The time-frequency methods mostly employed for arrhythmia detection from

ECG and HRV time-series include discrete and continuous wavelet transforms [5]. Numerous nonlinear features can be grouped into three categories: chaos and phase space quantification features, entropy-based features and other nonlinear features.

There are several programs known to us that implement some of the features used in the HRV analysis: ECGLab [6], KARDIA [7], and HRV Analysis Software [8]. In this work, we develop a novel framework for HRV analysis called HRVFrame that implements most of the previously mentioned features for HRV analysis. The framework is intended to act as a bridge between cardiac rhythm data records and medical knowledge discovery. The strong point of the framework is the number of implemented features, which allows researchers to easily compare the results with other authors. Our framework is still a work in progress, albeit at an advanced stage. From scientific perspective, the framework has been evaluated in a number of successful cardiac disorder classification tasks [9,10].

2 Framework Overview

HRVFrame is an extensive Java-based framework containing many features covered in the HRV analysis literature. Its main purpose is data preparation – it is used to extract features from cardiac rhythm records and store them as feature vectors prepared for further knowledge discovery. At present, the process of feature extraction is performed offline and the framework has not been integrated in any particular knowledge-based platform. The whole process of cardiac rhythm records analysis using HRVFrame is shown in Fig. 1. HRVFrame is organized into three major logical parts: input, feature calculation, and output.

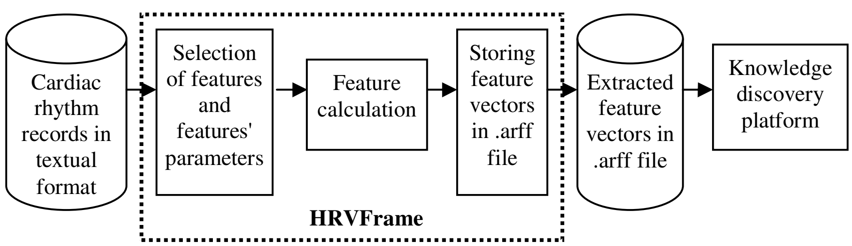

Fig. 1. Cardiac rhythm analysis using the HRVFrame computer framework

2.1 Data Input

The input part of the framework is designed to take cardiac rhythm records, to allow the selection of features and specification of features' parameters. Cardiac rhythm records in the form of textual ASCII files are acceptable as input files. The input file structure is the same as the one provided by the PhysioNet tool *rdann* [11]. The files should contain the information on the times of R peaks, types of beats, and optional rhythm annotations. All beat and rhythm annotations used in PhysioBank databases are supported by the framework. Also acceptable is the input format where only R peak times are known. Some features have parameters that determine the calculation procedure. Currently, the framework accepts a list of desired features and their parameters as arguments through a command line interface. Visual interface for feature and parameter selection is planned in the future. HRVFrame also accepts the list of cardiac rhythm files intended for analysis. Other system specifications include: start time or starting interval, end time or ending interval, RR interval series output, RR interval differences output, learning segment types from beats, etc. Features intended for extraction have to be specified by the researcher, currently there is no intelligent selection of appropriate features with respect to analyzed disorders or segment lengths, mainly because this area is still under research.

2.2 Feature Calculation and Output

HRVFrame is used to calculate values for the specified HRV features. The list of currently supported features is given in Table 1. A thorough search of the available literature was performed that resulted in implementation of most of the features applied in HRV research. Features are calculated for one record segment at a time.

The framework allows users to first create an .arff file that only contains the declaration of features. After the feature calculation process is complete, the framework creates output feature vectors and stores them in the .arff file. Data are appended to the end of the file as long as the new segments and records are continued to be analyzed. The .arff file can then be used by several data mining and knowledge discovery platforms like Weka [16] and RapidMiner [17] for disorder classification.

Table 1. Features implemented in the HRVFrame framework.

Feature(s)	Category	Parameters
Mean; SDNN; RMSSD; SDANN; SDSD; pNN*X*; Fano factor; Allan factor; HRV triangular index; TINN; Total PSD; ULF; VLF; LF; HF; LF/HF [2,5]	Linear, time and frequency	SDANN: number of segments; pNN*X*: miliseconds *X*; Fano and Allan factor: counting time; Fast Fourier Transform; window (Hanning, Hamming, none); Burg AR model; model order
Discrete Haar wavelet standard deviation [5]	Time-frequency	Scale
Largest Lyapunov exponent; correlation dimension; spatial filling index; central tendency measure; SD1/SD2; CSI, CVI; sequential trend analysis [4,5,6,12]	Nonlinear - phase space	Embedding dimension, trajectory length, lag
Detrended fluctuation analysis (DFA) α_1, α_2; Hurst exponent; Higuchi's fractal dimension [5,13]	Nonlinear - fractal estimates	Higuchi's kmax
Entropies: approximate (ApEn), sample (SampEn), multiscale sample (MSampEn), Rényi, spectral [4,14]	Nonlinear - entropy	ApEn, (M)SampEn: embedding dimension, radius, maximum (Y/N); Rényi: order; spectral: frequency band
Multiscale asymmetry index [15]	Other	Scale

3 Related Work and Framework Applications

ECGLab [6] and KARDIA [7] are Matlab-based tools for ECG and HRV analysis. Both tools extract some of the simple linear time and frequency domain features. In addition, current version of ECGLab implements time-frequency features, and KARDIA implements DFA scaling analysis. Both tools lack direct implementation of more recent nonlinear HRV features and depend on Matlab for execution. HRV Analysis Software [8] is freely-available Windows software with support for basic HRV analysis (linear and SD1/SD2 features) and is not intended for scientific data exploration, but rather as a decision support system for medical professionals.

The main application of HRVFrame is in cardiac data preparation for automated classification tasks. HRVFrame is oriented toward scientific exploration of different feature combinations for optimal models of cardiac abnormalities. It can be used for result comparison from different authors and for investigation of the research results reported by the authors. It should be however noted that this type of analysis requires that the protocol was clearly stated by the authors and that the used data is publicly available, which is often not a case. The framework can be used both for the analysis of cardiac disorders and for cardiac rhythm description in each record segment with arrhythmia detection in mind [10]. The framework always designates one rhythm type per segment based on rhythm priority, which can be specified by the user.

HRVFrame contains many features used by researchers in HRV analysis. It is interesting that the majority of the features are not specific for cardiac rhythm analysis and can be used in the analysis of other biomedical series (ECG, EEG, EMG, gait, skin resistance...). Application of our framework in non-cardiac domain is possible, e.g. if one investigates couplings between several different biomedical time-series.

Biomedical data analysis presents significant scientific challenges. Finding the appropriate feature combination for cardiac disorder detection and description is difficult. It includes both obtaining a significant number of feature vectors for valid conclusions as well as finding the optimal feature selection method and classifier models for each diagnosis. Any efficient and accurate methodology that would perform such a task would have to rely on a carefully selected subset of appropriate domain features. It would also need to integrate a number of features from diverse biomedical time series. Constructing thorough frameworks such as HRVFrame for each biomedical domain might be a favorable way for accomplishing such a task.

The current version of the framework is available free-of-charge, and only for non-commercial, scientific purposes. The framework is not open source at the moment. Please contact the author by e-mail: alan.jovic@fer.hr for further instructions.

4 Conclusion

A novel framework for HRV analysis is presented in this work. The framework allows researchers to thoroughly analyze cardiac rhythm records with different linear and nonlinear features and their combinations. It can be used in the analysis of cardiac disorders, arrhythmias, and in other biomedical domains. The main advantage of the framework compared to the existing solutions lies in the larger number of implemented HRV features and features' parameters. Future work will be focused on improving the interface of the framework and on implementation of other existing nonlinear methods for cardiac rhythm analysis.

References

1. Garcia, T.B., Holtz, N.E.: 12-Lead ECG: The Art of Interpretation. Jones and Bartlett Publishers, Sudbury, MA, USA (2001)
2. Malik, M., Bigger, J.T., Camm, A.J., Kleiger, R.E., Malliani, A., Moss, A.J., Schwartz, P.J.: Heart rate variability guidelines: Standards of measurement, physiological interpretation, and clinical use. Eur. Heart J. 17(3), 354--381 (1996)
3. Yang, A.C.-C., Hseu, S.-S., Yien, H.-W., Goldberger, A.L., Peng, C.-K.: Linguistic Analysis of the Human Heartbeat Using Frequency and Rank Order Statistics. Phys. Rev. Lett. 90(10), 108103 (2003)
4. Acharya, R.U., Joseph, K.P., Kannathal, N., Lim, C.M., Suri, J.S.: Heart rate variability: a review. Med. Bio. Eng. Comput. 44, 1031--1051 (2006)
5. Teich, M.C., Lowen, S.B., Jost, B.M., Vibe-Rheymer, K., Heneghan, C.: Heart-Rate Variability: Measures and Models. In: Akay, M. (ed.) Nonlinear Biomedical Signal Processing, vol. II, Dynamic Analysis and Modeling, ch. 6, pp. 159--213. IEEE Press, New York (2001)
6. de Carvalho, J.L.A., da Rocha, A.F., Nascimento, F.A.O., Neto, J.S., Junqueira, L.F.: Development of Matlab Software for Analysis of Heart Rate Variability. In: Proc 6th Int Conf Sig Proc ICSP 2002, pp. 1488--1491, IEEE Press, Beijing, China (2002)
7. Perakakis, P., Joffily, M., Taylor, M., Guerra, P., Vila, J.: KARDIA: A Matlab software for the analysis of cardiac interbeat intervals. Comp. Meth. Prog. Biomed. 98, 83--89 (2010)
8. Niskanen, J.-P., Tarvainen, M.P., Ranta-aho, P.O., Karjalainen, P.A.: Software for advanced HRV analysis. Comp. Meth. Prog. Biomed. 76, 73--81 (2004)
9. Jovic, A., Bogunovic, N.: Random Forest-Based Classification of Heart Rate Variability Signals by Using Combinations of Linear and Nonlinear Features. In: Bamidis, P.D., Pallikarakis, N. (eds.) Proc. XII Mediterranean Conf. Medical and Biological Engineering and Computing MEDICON 2010, pp. 29--32, Springer, Berlin (2010)
10. Jovic, A., Bogunovic, N.: Electrocardiogram analysis using a combination of statistical, geometric, and nonlinear heart rate variability features. Artif. Intell. Med., in press, doi:10.1016/j.artmed.2010.09.005
11. PhysioNet tool *rdann*, http://www.physionet.org/tutorials/physiobank-text.shtml#ann
12. Faust, O., Acharya, R.U., Krishnan, S.M., Min L.C.: Analysis of cardiac signals using spatial filling index and time-frequency domain. BioMedical Engineering OnLine 3, 30 (2004)
13. Peng, C.-K., Havlin, S., Stanley, H.E., Goldberger, A.L.: Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos 5(1), 82--87 (1995)
14. Costa, M., Goldberger, A.L., Peng, C.-K.: Multiscale entropy analysis of biological signals. Phys. Rev. E 71, 021906 (2005)
15. Costa, M., Goldberger, A.L., Peng, C.-K.: Broken asymmetry of the human heartbeat: Loss of time irreversibility in aging and disease. Phys. Rev. Lett. 95, 198102 (2005)
16. Witten I.H., Frank, E.: Data mining: Practical machine learning tools and techniques. Morgan Kaufmann, San Francisco (2011)
17. RapidMiner, http://rapid-i.com/content/view/181/190/