Ensemble learning frameworks for the discovery of multi-component quantitative models in biomedical applications

Valeriy V. Gavrishchaka and Mark E. Koepke

West Virginia University, Physics Department, Morgantown, WV, USA
gavrish@verizon.net
mark.koepke@mail.wvu.edu

Olga N. Ulyanova

Moscow State Academy of Fine Chemical Technology, Department of Physical Education, Moscow, Russian Federation
Plekhanov Russian Academy of Economics, Department of Sports Industry, Moscow, Russian Federation
ulanaon@mail.ru

Abstract

Increasing availability of multi-scale physiological data opens new horizons for quantitative modeling in biomedical applications. However, practical limitations of existing approaches include both the low accuracy of the simplified analytical models and empirical expert-defined rules and the insufficient interpretability and stability of the pure data-driven models. Recently it was shown that generic boosting-based frameworks can be successfully used to address these challenges of quantitative modeling in financial applications. Boosting and similar ensemble learning techniques are capable of discovering robust multi-component meta-models from a collection of existing and well-understood base models. Accuracy and stability of such interpretable ensembles of complementary models are often significantly higher than those of the single models. Here we establish the plausibility that this ensemble learning approach can overcome such challenges also in biomedical applications.

1. Introduction

The ongoing digital revolution has provided a relatively inexpensive means to collect and store multi-scale and multi-channel physiological data. Modern hospitals and research centers are well equipped with high-resolution monitoring, diagnostic and other data collection devices. Moreover, many portable systems for real-time collection and display of physiological data have become affordable for individual use outside of specialized medical facilities. These include Holter monitor and similar devices for electrocardiogram (ECG) and heart rate recording and specialized systems for electroencephalogram (EEG), electromyogram (EMG), respiration, and temperature. Such increasing availability of data opens new horizons for quantitative modeling in biomedical applications.

Practical quantitative modeling of most adaptive complex systems with many interacting components presents serious challenges [1]. Insufficient accuracy of both the simplified analytical and other low-complexity models and the empirical expert-defined rules is a typical limitation of existing approaches. Potentially more accurate and flexible data-driven models based on various statistical and machine learning algorithms are often unstable due to training data incompleteness, intrinsic nonstationarity of the system, and low signal-to-noise ratios. Crucially important in biomedical applications is model interpretability, which is often lacking in existing models.

Challenges of quantitative modeling both in applied finance and econometrics and in biomedical applications are very similar. Time series may be very noisy and non-stationary, which is particularly common in finance. Weakly deterministic regimes can be intermittent with predominantly stochastic

dynamics. Codependency relations between different time series are also often unstable. Availability of detailed multi-scale market and fundamental data does not guarantee a straightforward solution to all existing problems. This became especially clear after recent events in global financial markets when many quantitative portfolio strategies and state-of-art risk management systems became practically useless.

Recently, generic boosting-based frameworks have been successfully used to address many challenges of quantitative modeling in financial applications [1-3]. Boosting and similar ensemble learning techniques [1-6] are capable of discovering robust multi-component meta-models or strategies from a collection of existing well-understood base models. The accuracy of such interpretable ensembles of complimentary models is often significantly higher than those of the individual components. Such model portfolios can demonstrate stable performance over a wide range of market regimes.

Here we argue that a similar generic approach could be very promising in biomedical and other scientific, engineering, and technological applications. Several specific frameworks for the discovery of multi-component models for express diagnostics and efficient quantitative representation of psycho-physiological states are outlined.

2. Challenges of quantitative modeling in biomedical applications

Even before high-resolution biomedical data become available, artificial intelligence (AI) and statistical techniques have been successfully applied in medicine. However, earlier systems mostly rely on highly-averaged low-resolution data, semi-quantitative categorical information and on empirical expert-defined rules applied to such data. Therefore, models took the form of rule-based systems capable to assist in diagnostic and treatment suggestions such as well-known MYCIN expert system (ES) [7]. Rules are extracted from the knowledge of experts. However, often experts may not know, or may not be able to formulate, what knowledge they actually use in solving their problem.

Increasing availability of physiological and clinical data justified application of various statistical and machine learning techniques that add learning or automated knowledge discovery to ES. Examples of this kind of systems range from intelligent processing of ECG data to computer-based interpretation of medical images, e.g., for early detection of cancer [8]. At the same time, revolutionary advances in the human genome project and other achievements in molecular biology have lead to the explosive growth of the high-dimensional genomic and other micro-biological data. This opens perspectives for computational solutions of many important bioinformatics problems and *in-silico* drug discovery using advanced machine learning and statistical techniques [9,10].

However, typical problems of complex system modeling, such as "curse" of dimensionality and non-stationarity, lead to serious challenges in biomedical applications. For example, direct machine learning models in bioinformatics have very high-dimensional inputs causing training data incompleteness even with apparent abundance of the microbiological data. Indeed, training data should include sufficient part of all possible input combinations which scales as $\sim M_1 M_2 \ldots M_N$, where N is total number of inputs (basic features) and M_i is typical number of different ranges/regimes for i-th feature. Similar problems are also typical for medical image classification models based on high-dimensional features as well as models based on multi-channel and multi-scale physiological data. Non-stationarity is another challenge in modeling physiological dynamics. It is usually impractical to find and calibrate a single global multi-dimensional model that reasonably covers all different dynamical regimes.

The described problems are even more pronounced in the emerging field of personalized medicine [11,12]. Direct optimization of personal treatment based on genotype is not effective for abnormalities related simultaneously to multiple genes and influenced by history of non-genomic factors. Even more challenging is to personalize alternative therapies such as biofeedback [13] which have been shown to be effective not only in clinical settings but also in achieving optimal psycho-physiological states in professional sport [14] and other areas. It is difficult to choose an optimal protocol since noticeable effects are hard to detect in early stages without robust quantitative indicators that are sensitive to small changes of psycho-physiological states. In general, early detection of subtle positive or negative changes caused by a particular treatment is a very important and challenging task in personalization and optimization of any medical treatment strategy.

Low-complexity models based on existing knowledge and heuristics of the considered domain require much less data for calibration and demonstrate significantly more stability than pure data-driven models. For example, for several consecutive years, one of the top performers at the competition for the protein structure prediction was not the most computationally intensive data mining algorithm but much less complex model based on biological heuristics [10]. Many other low-complexity quantitative models inspired by existing knowledge of physiological processes and brain dynamics have been developed [15,16]. However, the drawback of such

models is that they can be significantly biased due to the underlying simplified assumptions and/or can be applicable only in certain dynamical regimes or physiological parameter ranges.

Limitations of such simplified base models can be compensated by model combination via different ensemble learning techniques [6]. However, typical applications of these techniques in biomedical and many other fields still rely on pure data-driven base models [1,17]. Such an approach could require even more training data and would result in much less interpretable solutions than single data-driven model. However, recently we have proposed to use boosting and similar ensemble learning techniques to combine existing low-complexity models to produce stable and interpretable meta-models [1]. Such generic frameworks have proven to be effective in several challenging financial and econometric applications [1-3]. This modeling approach is outlined in the next section.

3. Ensemble learning frameworks for the discovery of multi-component models from existing incomplete knowledge

One of the machine learning approaches to compensate for deficiency of the individual simplified models is to combine several models to form a committee [6]. One of the most flexible frameworks of this type is the mixture-of-experts (ME) model [6]. The probabilistic form of ME model can be written as

$$p(y \mid x) = \sum_{t=1}^{T} \pi_t(x)\, p_t(y \mid x)$$

Mixing coefficients $\pi_t(x)$ are known as gating functions and the component densities $p_t(y|x)$ are called experts. These components can model the distribution in different regions of input space: they are "experts" in these regions. Gating functions determine which components are dominant in which region. Extension to multilevel gating functions gives hierarchical mixture of experts (HME) [6].

To increase accuracy and flexibility, complex machine learning models (e.g., NNs) can be used as components in both ME and HME frameworks [6]. However, increasing complexity of the models often leads to practical problems in the training of such a multi-component system, to poor out-of-sample performance and instability. Thus, it is highly desirable to have a committee of the well-understood and low-complexity expert models that consistently demonstrate acceptable combined performance [1].

Adaptive boosting and similar ensemble learning algorithms combine many desirable features [1,4-6]. The majority of ensemble learning algorithms, including "random sample" techniques such as bagging, can reduce only the variance part of the model prediction error, i.e. they make a combined model more stable. Boosting, on the other hand, can reduce both bias and variance parts of the model error, i.e. it makes the model more accurate and more stable. It means that one can start with simple model ("rule of thumb") with low accuracy slightly above the random guess and discover a committee of complimentary models with much higher accuracy. Therefore, boosting can be applied to the pool of the well-understood low-complexity models to produce interpretable, but qualitatively different, combined model with significantly higher accuracy and stability as discussed in [1]. Moreover, boosting tries to maximize margin to ensure good out-of-sample performance, i.e. it is a large-margin classifier [4-6].

A typical boosting algorithm such as AdaBoost [4-6] for the two-class classification problem (+1 or -1) consists of the following steps:

$$w_n^1 = 1/N \tag{1}$$

$$\varepsilon_t = \sum_{n=1}^{N} \left(w_n^t\, I(-y_n\, h_t(x_n)) \right) \tag{2}$$

$$\gamma_t = \sum_{n=1}^{N} \left(w_n^t\, y_n\, h_t(x_n) \right) \tag{3}$$

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1+\gamma_t}{1-\gamma_t}\right) - \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right) \tag{4}$$

$$w_n^{t+1} = w_n^t \exp\left(-\alpha_t\, y_n\, h_t(x_n)\right)/Z_t \tag{5}$$

$$H(x) = \sum_{t=1}^{T} \alpha_t\, h_t(x) \Big/ \sum_{t=1}^{T} \alpha_t \tag{6}$$

Here N is the number of training data points, x_n is a model input value of the n-th data point and y_n is class label, T is the number of iterations, $I(z) = 0$ ($z<0$), $I(z)=1$ ($z>0$), w_n^t is the weight of the n-th data point at t-th iteration, Z_t is normalization constant, $h_t(x_n)$ is the best model at t-th iteration, ρ is a regularization constant, and $H(x)$ is the final combined model (meta-model).

Boosting starts with equal and normalized weights for all training data (step 1). Base classifiers $h_t(x)$ are trained using weighted error function ε_t (step 2). The best $h_t(x)$ is chosen at the current iteration. The data weights for the next iteration are computed in steps (3)-(5). At each iteration, data points misclassified by the current best model (i.e., $y_n\, h_t(x_n) < 0$) are penalized by the weight increase for the next iteration. AdaBoost constructs progressively more difficult learning problems that are focused on hard-to-classify patterns defined by the weighted error function (step 2). Steps (2)-(5) are repeated at each iteration until stop criteria

occur. The final meta-model (Eq.6) classifies the unknown sample as class +1 when $H(x) > 0$ and as -1 otherwise.

In our previous works we have demonstrated boosting effectiveness for a combination of low-complexity econometric models [1] and proposed a boosting-based optimization framework for the discovery of portfolios of trading strategies from the low-complexity base strategies [2,3]. It was shown that such frameworks are capable to discover interpretable multi-component portfolio strategies with stable performance across a wide range of market regimes and robust generalization ability [2,3].

However, success of the described boosting-based frameworks still relies on the availability of sufficient historical time series data. Potentially interesting strategies that directly exploit rare events or emerging regimes cannot be obtained in this way due to the lack of training data. A promising approach to learn novel/rare classes or patterns could be single-example learning frameworks developed mostly for computer vision applications. One of them is single-example learning using representation by similarity [18,19]. In this framework, a novel class is characterized by its similarity to several previously learned classes. If a system is already trained to classify several classes using sufficient amount of training data it can be extended to classify an additional novel class using just one or several examples of this class.

Assume that n well-learned classes C_i exist, F_i is a classifier for the i-th class, $C(p)$ is the class to which a pattern p belongs, and $F_i(p)$ is the output of F_i on pattern p (i.e., probability-like number that p belongs to C_i). $F_i(p)$ can be interpreted as the similarity of pattern p to the class C_i. To learn a novel class C from a single example E, the similarity of E to the familiar classes is recorded in a feature vector:

$$F_1^n(E) = [F_1(E)...F_n(E)]^T . \quad (7)$$

Other instances of class C are expected to have similarity/feature vectors (Eq.7) resembling that of E. Therefore, nearest-neighbor classification can be used. New patterns p are classified as belonging to class C when the distance between their similarity vectors and that of E is below a chosen threshold.

Boosting-based and single-example learning frameworks can be naturally used together. For the classification problems with a sufficient amount of training data, boosting can be used to discover robust multi-component models (Eq.6). When several stable meta-models are available, they can be used as classifiers F_i to represent novel classes via the feature vector (Eq.7). Moreover, since each F_i is a weighted collection of base models, even one meta-model can provide a detailed feature vector with components representing the state of each base-model scaled by its weight. In a financial context, novel class could represent new emerging market regime. The presented approach is generic and can be used in many different fields. Potential biomedical applications of these frameworks are outlined in the next section.

4. Multi-component quantitative models in biomedical applications

The proposed approach provides practical tools for the optimal combination of existing low-complexity models and could be applied to different problems discussed in section 2. However, because of our current research interests, we would like to focus only on models for express diagnostics and quantitative representation of psycho-physiological states based on easy-to-collect physiological data. On-site and remote express diagnostics with real-time capabilities are important in intensive care medicine, professional sport rehabilitation and performance enhancement, military training, and many other applications. Collection of diagnostic models can also be used as sensitive indicators of personal psycho-physiological state dynamics. Such indicators could play a key role in rapid discovery of personalized treatment strategies and in achieving optimal psycho-physiological states in sport and other applications.

One of the most easily collectable physiological data offering rich and informative dynamics is ECG time series. Complex physiological signals including ECG are typically non-stationary, but not statistically random [15,20]. This necessitates quantifying the complexity of physiological dynamics with the help of different indices. Methods from nonlinear dynamics (NLD) have shown new insights into heart rate variability (HRV) changes under various physiological and pathological conditions, providing additional prognostic information and complementing traditional time- and frequency-domain analyses [15,20].

Many types of such measures inspired by NLD have been proposed [20]. These include fractal measures such as power-law scaling exponent, indices from detrended fluctuation analysis, and multiple scaling exponents from multifractal analysis. A different set of measures consists of entropy measures: approximate/sample entropy, multiscale entropy, and compression entropy. Different symbolic dynamic measures (entropies and probabilities) represent one more family of measures. Other practical types of HRV signal complexity measures have also been proposed [15,20].

Any such measure is potentially useful for classifying between normal and pathological cases [15,20]. The obvious limitation of a single measure is insufficient accuracy (bias). More importantly, most of such measures require long-enough periods of data to provide stable results, thus limiting their practical value for the express diagnostics and real-time low-

noise indicator of psycho-physiological state dynamics. It is clear that an intelligent model/measure combination using discussed frameworks could significantly improve accuracy and stability as well as drastically reduce requirements for the data-acquisition duration.

In addition to direct classification for diagnostics, the probability-like output of the multi-component model (Eq.6) could also be used as a sensitive indicator for early detection of treatment effects, thus allowing optimizing of a treatment strategy for a particular patient. Existing collections of multi-component models (Eq.6) could also be very effective in the context of single example learning. This would allow even faster discovery of an optimal treatment strategy for a particular patient associated with a model state (Eq.7) that resembles that of the patient that was successfully treated in the past.

The concept of individual zones of optimal functioning introduced in sport psychology revealed multi-featured (multi-dimensional) nature of personal psycho-physiological states associated with the best performance [21]. However, it is difficult to quantify such optimal zones. Therefore, in the most cases, these zones are just roughly estimated from subjective questionnaires [21]. On the other hand, optimal zones based on objective physiological measures such as heart rate (HR) have been identified in several different sports [14]. Output from the collections of multi-component physiological models (Eq.6), expressed as feature vector (Eq.7), could be an objective multi-dimensional indicator of psycho-physiological states in sport and similar applications. Current feature vector (Eq.7) could be compared to the feature vectors associated with previous optimal performances and used as sensitive indicator of progress for personalized tuning of different methods for achieving such optimal states (psychological training, biofeedback, etc.).

5. Application example

In this section potential applications of the proposed framework are discussed in the context of using beat-to-beat heart rate (RR) time series for diagnostics and psycho-physiological state characterization. The RR-interval signal can be extracted with high accuracy from even noisy ECG time series (e.g. those collected by portable devices outside medical facilities) while extraction of other ECG features could often be seriously affected by noise. Thus, RR time series can be used in many real-life applications where consistent usage of full ECG and other more complex data such as EEG is not yet technologically feasible.

Many original and current applications of statistical physics and nonlinear dynamics methods in physiological modeling successfully use RR time series [15,20]. Majority of such models are formulated to discriminate (classify) between normal and various abnormal states of cardiovascular system based on different complexity measures of RR time series [15,20]. However, HR and HRV sensitivity to emotions and other complex psycho-physiological states [22-24] significantly expands potential application areas of models based on RR data.

Majority of publications on linear and NLD-inspired analyses of RR time series focus on illustrating the discriminative power of a chosen single complexity measure, β. This majority includes discrimination between normal and pathological cases [15,20] as well as between different psycho-physiological states [22]. However, although many indicators have been shown to provide statistically significant differentiation between classes of interest, the considerable overlapping of regions prohibits practical usage of such indicators in many important applications. More importantly, majority of such indicators require long periods of data (~ several hours) for adequate calculations. This could drastically limit practical usability of these indicators in many applications, including express diagnostics, real-time monitoring for operator alertness [22], early detection of subtle directional changes during personalization of medical treatment or psychological training, and intensive-care medicine.

For illustration of these problems and their potential resolution using our framework, we will consider just one well-known indicator, based on detrended fluctuation analysis (DFA), that was successfully applied to RR data analysis in the past [15,20]. DFA was proven to be useful in revealing the extent of long-range correlations in time series. First, the investigated time series with N samples is integrated. Next, the integrated time series is divided into n boxes of equal length. In each box, a least-square line is fit to the data with y coordinate denoted by $y_n(k)$. Finally, the integrated time series, $y(k)$, is detrended as follows:

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left[y(k) - y_n(k)\right]^2} \quad . \qquad (8)$$

A linear relationship on the plot of log $F(n)$ vs. log n indicates power law (fractal) scaling characterized by a scaling exponent β (slope of the fitted straight line).

Peng et al. found that $F(n)$ computed from RR times series is characterized by two scaling exponents β_1 and β_2 (cross-over phenomena) computed over a small ($4<n<16$) and a larger ($16<n<64$) intervals, respectively [15]. Here we reproduced these results using up to 24 hours of RR data from 8 subjects with normal sinus rhythm and 8 subjects with congestive

heart failure (http://www.physionet.org/physiobank). The two scaling exponents are computed over approximately 2-hour segments (8192 beats) and presented in figure 1. It is clear that the two scaling exponents (complexity measures) provide distinctive clustering of the normal and pathological RR segments. For example, β_1 for healthy subjects is usually larger than that of subjects with abnormality. However, it is also clear that direct practical usage of (β_1,β_2)-based indicator can lead to ambiguous interpretation because of significant overlapping of normal and pathological classes in (β_1,β_2) space.

More problems become evident when RR data-acquisition duration is reduced below several hours as naturally required by many potential applications. The ambiguity problem is clearly illustrated in figure 2 where averaged β_1 for all healthy subjects (solid line) and β_1 range for one of the subjects with congestive heart failure (dotted lines) is plotted as a function of RR segment length. While for large RR segments one can clearly differentiate this subject with abnormality from the average healthy subject for all RR segments, for small-length segments the noise level (or instability) cannot be tolerated by any practical application (jumping from good separation to very significant overlapping).

To increase overall accuracy and to improve stability for short RR segments, one can use classification models based on linear and nonlinear combinations of several β values instead of direct usage of these values as indicators. In general, such a classification model takes the form of the nonlinear mapping:

$$y = F(\beta_1,...,\beta_K), \qquad (9)$$

where F is defined implicitly in most cases. For example, linear and nonlinear measures (including DFA scaling exponents) extracted from RR data and used as inputs to neural networks, support vector machines, and other machine learning algorithms have been proven to be successful in producing accurate classifiers for cardiac arrhythmias [25,26]. Although not studied directly, one can assume that these classifiers could also be noticeably more stable for short RR segments compared to β values themselves. The probability-like real-value output of such classifier (y) could also be used as a single-number composite measure for early detection of the positive or negative changes during treatment optimization as well as change of psycho-physiological state in general.

However, there are several serious limitations of such complex "black-box" classifiers. Even when inputs are based on well-understood measures, the final nonlinear classifier still lacks interpretability and clear understanding of stability ranges. Compensating or eliminating these shortcomings is very important for biomedical applications. Such an approach also lacks the scalability to include all existing partial knowledge in the form of many different complexity measures and other simple models. Indeed, a direct increase of number of inputs corresponding to different complexity measures β_i, will lead to a complexity increase of NN or similar algorithm and finally will make it impossible to train such a classifier on any data set of practical size (which is always incomplete) without overfitting. Usage of dimensionality reduction preprocessing, such as PCA or its nonlinear extensions [27], is often of limited value since class membership information (for classification models) is not taken into account. Also, the possible diversity of measures $\beta(\boldsymbol{p})$ due to parameter vector $\boldsymbol{p}$ variations are not fully used (only best-on-average parameters are employed).

The limitations of the individual indicators β_i and the drawbacks of "black-box" classifiers (Eq.9) can be self-consistently resolved by boosting-like ensemble learning frameworks. One can employ low-complexity base classifiers that depend on just a few different measures β. For example, it is clear from figure 1 that even a two-level classification tree which uses simple thresholding on β_1 and β_2 at the 1-st and 2-nd level respectively (or vice versa) could be promising as a base model. In the general case of many potentially useful measures, $\beta_1 \ldots \beta_K$, classifiers based on any two measures take the form:

$$y = h(\beta_i[p_i], \beta_j[p_j], \gamma_1, \gamma_2, \gamma_3), \qquad (10)$$

where γ_i are threshold levels used at 1-st and 2-nd levels of this classification tree. Next, applying boosting steps (1)-(6) to the base classifier (Eq.10) with all possible (i,j) combinations and optimizing over (p_i, p_j, γ_1, γ_2, γ_3) at each boosting iteration, we obtain multi-component meta-classifier (Eq.6). Since, at each iteration, low-complexity models are optimized one by one, we can use any number of measures β (i.e., all existing incomplete knowledge) relevant to the considered problem without direct threat of overfitting contrary to (Eq.9). Note, however, that boosting often finds compact meta-classifier by including only a small part of the complementary measures in the final model (Eq.6). Also, unlike direct approach (Eq.9), optimization over (p_i, p_j) at each boosting iteration will be able to find not only a single parameter set optimal on-average but also much less obvious complementary measures which could significantly contribute to the final meta-classifier. Finally, the obtained ensemble of models is interpretable and each of the constituents is simple enough for further analysis that may lead to discovery of other types of low-complexity base models.

Applying boosting steps (1)-(6) with base classifier (Eq.10) that includes just DFA scaling exponents β_1 and β_2, it is quite easy to achieve out-of-

sample classification hit rate of more than 80% on the considered data set (i.e., ratio of correctly classified RR segments to the total number of segments). Unlike the significantly volatile β_1 value of the patient with congestive heart failure in figure 2, the classification hit rate of the obtained meta-classifier remains stable at the level of ~95% for all segment lengths considered in figure 2. Also, compared to a single classifier that uses fixed suggested best-on-average n-values ($4<n<16$ and $16<n<64$), the boosted meta-classifier classifier with optimized n-values at each iteration is capable of increasing the out-of-sample classification hit rate by 10-20%. Obviously, by including other existing measures besides DFA scaling exponents into base models (Eq.10), we obtain further impressive improvements in meta-classifier performance. Additional detailed examples will be presented elsewhere.

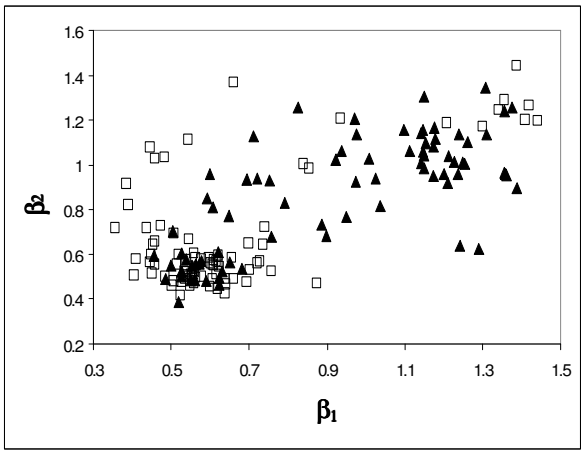

Figure 1: DFA scaling exponents for the healthy subjects (triangles) and subjects with congestive heart failure (squares) for RR sample size = 8192 beats.

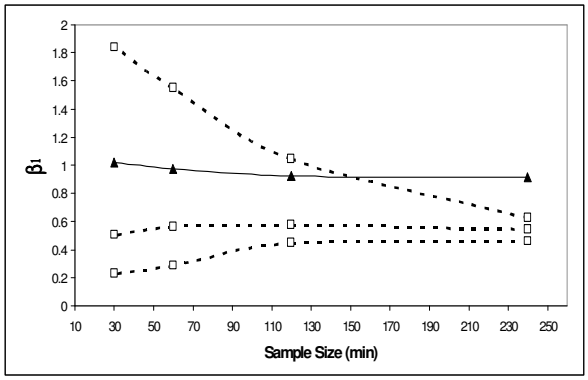

Figure 2: DFA scaling exponent vs RR sample size: (1) triangles (solid line) – average for all considered healthy subjects, (2) squares (dotted lines) – average, max, and min for one of the subjects with congestive heart failure.

6. Conclusions

Ensemble learning frameworks for the discovery of multi-component biomedical models from the existing incomplete knowledge have been outlined. Possible applications of such models include express diagnostics from limited data as well as representation of personal psycho-physiological states for automated monitoring and for discovery of personalized treatment strategies. The same approach could be also effective in quantification of optimal psycho-physiological states in professional sports and in similar performance-metric endeavors.

7. References

[1] V.V. Gavrishchaka, Boosting-based frameworks in financial modeling: Application to symbolic volatility forecasting, *Advances in Econometrics*, 20B, 123, 2006.

[2] V.V. Gavrishchaka, Boosting-based framework for portfolio strategy discovery and optimization, *New Mathematics and Natural Computation*, 2, 315, 2006.

[3] V.V. Gavrishchaka et al, Discovery of multi-component portfolio strategies with continuous tuning to the changing market micro-regimes using input-dependent boosting, in the *Proceedings of Computational Finance and its Applications*, 2008.

[4] R.E. Schapire, The design and analysis of efficient learning algorithms, *PhD thesis*, MIT Press, 1992.

[5] G. Ratsch, Robust boosting via convex optimization: Theory and applications, *PhD thesis*, Potsdam University, 2001.

[6] C.M. Bishop, *Pattern recognition and machine learning*, Springer, 2006.

[7] C.A. Perry, Knowledge bases in medicine: a review, *Bull. Med. Libr. Assoc.*, 78(3), 1990.

[8] G.D. Magoulas, A. Prentza, Machine learning in medical applications, in *Machine Learning and its applications*, Springer, 2001.

[9] P. Baldi, S. Brunak, *Bioinformatics: The machine learning approach*, MIT Press, 2001.

[10] B. Bergeron, *Bioinformatics Computing*, Prentice Hall PTR, 2003.

[11] M.D. Kane et al, Drug safety assurance through clinical genotyping: Near-term considerations for a system-wide implementation of personalized medicine, *Personalized Medicine* 5(4), 387-397, 2008.

[12] J. Davis et al, Machine learning for personalized medicine: Will this drug give me a heart attack?, in the *Proceedings of International Conference on Machine Learning (ICML)*, 2008.

[13] J.R. Evans, A. Abarbanel, *Introduction to quantitative EEG and neurofeedback*, Academic Press, 1999.

[14] B. Blumenstein et al, *Brain and body in sport and exercise*, John Wiley & Sons, 2002.

[15] J. Belair et al., *Dynamical disease: Mathematical analysis of human illness*, American Institute of Physics, 1995.

[16] E. Gordon, *Integrative neuroscience*, Hardwood Academic Publishers, 2000.

[17] H. Moon, Ensemble methods for classification of patients for personalized medicine with high-dimensional data, *Artificial Intelligence in Medicine*, 41, 197-207, 2007.

[18] S. Edelman, *Representation and recognition in vision*, MIT Press, 1999.

[19] E. Bart and S. Ullman, Single-example learning of novel classes using representation by similarity, *in Proc. BMVC*, 2005.

[20] A. Voss et al., Methods derived from nonlinear dynamics for analyzing heart rate variability, *Phil Trans. R. Soc. A*, 2008.

[21] Y.L. Hanin, Emotions and athletic performance: Individual Zones of Optimal Functioning model, *European Yearbook of Sport Psychology*, 1, 29-72, 1997.

[22] P. Smrcka et al., Fractal and multifractal properties of heartbeat interval series in extremal states of the human organism, *Measurement Science Review*, 3, 13, 2003.

[23] T.T. Yang, et al., Increased amygdale activation is related to heart rate during emotion processing in adolescent subjects, *Neuroscience Letters*, 48, 109-114, 2007.

[24] Van Hecke et al., Electroencephalogram and heart rate regulation to familiar and unfamiliar people in children with autism spectrum disorder, *Child Development*, 80, 1118-1133, 2009.

[25] B. Anuradha and V.C. Veera Reddy, ANN for classification of cardiac arrhythmias, *ARPN Journal of Engineering and Applied Sciences*, 3, 1, 2008.

[26] G. Kheder, et al., HRV analysis using wavelet package transform and least square support vector machine, *Int. J. of Circuits, Systems and Signal Processing*, 2, 18, 2008.

[27] A.J. Izenman, *Modern multivariate statistical techniques: Regression, classification and manifold learning*, Springer, 2008.