Acta Technica 62 (2017), No. 7A, 1263–1274 © 2017 Institute of Thermomechanics CAS, v.v.i.

Influence analysis of national traditional sports performance in colleges and universities based on triple exponential forecast of sliding window

Luran Yang[1]

Abstract. Aiming at the defects of low accuracy and slow speed of national traditional sports performance forecast model of colleges and universities under the current situation of national traditional sports development in colleges and universities, a forecast model of national traditional sports performance of colleges and universities based on triple exponential forecast of sliding window is proposed in this paper. First of all, a large number of national traditional sports performance data of colleges and universities are collected and preprocessed, and then target is adopted to study the national traditional sport performance training samples, and the triple exponential forecast optimization algorithm of sliding window is adopted to select the targeted threshold and weight and other parameters so as to establish the optimal forecast model of national traditional sports performance of colleges and universities. Secondly, the sliding window is constructed by setting the segmentation points and triple smoothing exponential algorithm is combined for the real-time segmentation of gray model data, to obtain the real-time statistics characteristics of data, construct the function relationship between sequence error forecast and compression ratio, and make beak point judgment based on error forecast sequence to improve the robustness of inspection link. Finally, the effectiveness and superiority of the model are compared by using the national traditional sports performance data. The results show that the model can improve the forecast accuracy of national traditional sports performance of colleges and universities, and the forecast results are more reliable, which can provide valuable information for national traditional sports training in colleges and universities.

Key words. Sliding window, Triple exponential, National traditional sports in colleges and universities, Influence analysis.

[1]Sport Science and Human Body, Changzhi Medical College, Changzhi Shanxi Province, 046000, China

1. Introduction

As people's living standards are improved constantly, there are more and more sub-health phenomena in the people's body, especially the physical quality of college students is not as good as before. How to improve the physical quality of college students has aroused widespread concern in our country. The forecast of national traditional sports performance in colleges and universities can describe the physical conditions of college students and the competitive level of athletes, therefore, the modeling and forecast of the national traditional sports performance in colleges and universities can make relatively reasonable training plans for athletes and college students to improve their sports performances.

Considering national traditional sports performance data of colleges and universities as a set of data sets, model it with multiple linear regression and estimate the performance of an athlete based on the parameters. However, because linear regression requires all samples evenly distributed and presented in nonlinear growth tendancy, which is inconsistent with the actual conditions of the national traditional sports performance data of colleges and universities, the forecast result is not reliable and of less practical application value. In recent years, with the continuous development of fuzzy theory and gray theory, many non-linear national traditional sports performance forecast models of colleges and universities have emerged. The reliability of national traditional sports performance forecast in colleges and universities is superior to that of multiple linear regression model. In practice, these models also have significant limitations. For example, the fuzzy theory is difficult to grasp and requires a certain theoretical basis, while the gray theory regards the traditional national sports performance forecast process as a black box, showing poor interpretability of forecast results. At present, some scholars have put forward the target-based forecast model of national traditional sports performance in colleges and universities. The target has self-organizing and non-linear mapping ability, which can describe the changing characteristics of national traditional sports performance in colleges and universities better and constructs a better forecast model of national traditional sports performance in colleges and universities than other models. However, the relevant parameters of the target, such as the threshold and the weight, have a great influence on the performance of the target. If these parameters are determined unreasonably, the forecast results of national traditional sports performance will be very low. In order to solve the problem of determining the target parameters, some scholars have proposed using the genetic algorithm and the particle swarm algorithm, etc. to determine the threshold and weight of the target, which effectively improves the forecast accuracy of the national traditional sports performance. Genetic algorithm and particle swarm optimization belong to stochastic optimization algorithm. Usually, they can be used to find out the sub-optimal solution and fails to find out the global optimal solution. And the local optimal threshold and weight easy to be searched out make the target learning time longer and network structure more complicated, influencing the modeling effect of the national traditional sports performance in colleges and universities.

In order to improve the forecast accuracy of national traditional sports performance in colleges and universities, a forecast model of national traditional sports performance in colleges and universities of optimization target based on triple exponential forecast of sliding window is proposed. The results show that the model can improve the forecast accuracy of the national traditional sports performance in colleges and universities, and the forecast results are more reliable.

2. Forecast model of national traditional sports performance in colleges and universities

2.1. Factors of influence

The performance of 1,000 meters was taken as the research object. Before building a 1000-meter running forecast model, the feasibility of the actual operation was considered, and a number of factors including height (m, x_1), weight (kg, x_2), chest circumference (m, x_3), vital capacity (ten thousand, x_4), heart rate (one hundred times, x_5), leg length (m, x_6), standing long jump (m, x_7), monthly average exercise time (hrs, x_8), 1000 m running performance (s, y) are initially selected as output.

2.2. Nonlinear screening of influence factors

(1) Firstly, all normalized n factors are used as input of LSSVM, and the optimal kernel function parameters are selected by 10-fold cross-validation to obtain the training accuracy of all factors, denoted as MSE_0.

(2) Eliminate the ith $i\,(i = 1, 2, \cdots, n)$ factor and use the remaining factor as the LSS-VM input to obtain training accuracy again. Repeat all n factors to get a reject RMSE set $\{MSE_{-1}, MSE_{-2}, \cdots, MSE_{-n}\}$.

(3) If $RMSE_{\min} < RMSE_0$, indicating that the corresponding factor of $RMSE_{\min}$ can improve the forecast accuracy of the model, then we should reject the factor and change the value of $RMSE_0$ as $RMSE_{\min}$.

(4) Repeat steps (1)-(3) process (the total number of factors will be reduced by 1 in each cycle) until $RMSE_{\min} > RMSE_0$. The remaining factors at the moment are the main influence factors retained.

(5) According to the main influencing factors reserved, the data sets of national traditional sports performance in colleges and universities are processed, and the data dimension is greatly reduced.

2.3. Weighting of retention factors

After the factors are screened, the retention factor can be regarded as a factor conducive to improving the forecast accuracy of the model. However, the contribution of each factor to the model's forecast result is not necessarily the same. Therefore, the LSSVM is used to weight the retention factor one by one, with the process as follows :

of LSSVM, and the optimal kernel function parameters were selected by 10-fold cross-validation points to obtain the training accuracy denoted as RMSE as the model background accuracy.

(2) The ith retention factor is forcibly removed, and the 10-time cross-training accuracy of the remaining retention factor excluding this factor is calculated. The greater the accuracy, the worse the forecast accuracy of the model is after the retention factor is removed, indicating that the factor is more important to the forecast result.

(3) For each retention factor, after having the background deducted and being normalized, the weight value of each factor can be obtained.

3. Triple exponential forecast of sliding window

3.1. Error forecast method

The smooth exponential forecast in segmentation form can be used to obtain the forecast output S_t of the model at moment t. If the actual sports score at the moment t is y_t, then the deviation of the forecast result is $y_t - s_t$, and the forecast error meets the following defined features.

Definition 1: Assumed that T represents the time data sequence of sports performance with length of n, ΔErr_i is the forecast error of sports performance at set segmentation point i , $SKPS = \{SP_1, \cdots, SP_m\}$ represents the segmentation points set, and m is is the set length. ΔErr_i has the characteristics of random variable and satisfies the independent identical-distribution characteristics, and the normal distribution $N\left[\mu, \sigma^2\right]$ characteristics, and:

$$\mu = \frac{1}{m}\sum_{i=1}^{m}\Delta Err_i, \sigma = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(\Delta Err_i - \mu\right)^2}. \tag{1}$$

Analysis: The time series model of sports performance accords with the characteristics of random process. Although the forecast error at a certain moment is known, it can not accurately predict the forecast error of the next neighbor point, that's, the point-point forecast deviation is of no dependency and has the characteristic of independent distribution. According to the central limit theory, if the number of samples tends to infinity, a large number of data generated with random features will meet the normal distribution characteristics. Assumed cΔErr_i is a random variable, ΔErr_i is the forecast deviation at segmentation point i. If the time series of sports performance approaches the infinite length, then the forecast sequence deviations will accord to the normal distribution characteristics. Therefore, their sub-sequences ΔErr_i should also meet the normal distribution $N\left[\mu, \sigma^2\right]$ characteristics.

In the time series data of sports performance, the compression rate is the key parameter to perform sequence segmentation. Based on the above definition, Lemma 1 can be obtained, which can characterize the relationship between segment compression rate and forecast error.

Lemma 1: Assumed that the compression rate of the time series data of sports performance is p, then the following inequality relation can be obtained:

$$2\Phi\left(2x\right) - 1 \leq 1 - p\,. \tag{2}$$

In the formula (16), x is the deviation oscillation of mean, Φ is distribution accumulation function, meeting the normal distribution characteristics.

Analysis: Assumed that the compression segmentation rate of the time series data of sports performance is p, then the probability of the existence of the segmentation points is less than $1 - p$. The closer to μ, the larger the probability of the existence of the segmentation point; and the farther to μ, the smaller the probability of the existence of the segmentation point. The smooth exponential segmentation forecast process assumes that the distribution range is in line with the standard deviation of the mean, therefore, the ΔErr_i at segmentation point position is distributed within the interval of $[\mu - 2x\sigma, \mu + 2x\sigma]$, and the probability of the existence of segmentation point is less than $1 - p$. Assumed that R represents the random variable, then it can be obtained that:

$$P\left\{\mu - 2x\sigma < R <\right\} \leq 1 - p\,. \tag{3}$$

The formula (3) can be transformed to obtain:

$$P\left\{\frac{(\mu - 2x\sigma) - \mu}{\sigma} < \frac{R - \mu}{\sigma} < \frac{(\mu + 2x\sigma) - \mu}{\sigma}\right\} \leq 1 - p\,. \tag{4}$$

According to the central limit theorem, after the transformation of random variables, the $(R - \mu)/\sigma$ obtained should accord with the normal distribution $N\left(0, 1\right)$ characteristics. Therefore, formula (4) can be used to deduce that formula (5) holds.

With the forecast error, the segmented subdata sequence of the sports performance time series data can be obtained, which satisfies a single trend and has a certain degree of stability in the independent interval. If a data trend deviates greatly from the trend of the current sequence, the point is the possible segmentation point of data sequence.

3.2. Construction of segmentation point

Based on the second smoothing exponential, the triple smoothing exponential model can be constructed as follows [12].

$$\begin{cases} S_t^{(1)} = \alpha X_t + (1 - \alpha) S_{t-1}^{(1)} \\ S_t^{(2)} = \alpha S_t^{(1)} + (1 - \alpha) S_{t-1}^{(2)} \\ S_t^{(3)} = \alpha S_t^{(2)} + (1 - \alpha) S_{t-1}^{(3)} \end{cases} \tag{5}$$

In formula (5), $t = 2, 3, \cdots, S_t^{(1)}$ is the first smoothing exponential value at moment t; $S_t^{(2)}$ is the second smoothing exponential value at moment t; $S_t^{(3)}$ is the triple smoothing exponential value at moment t; X_t is the actual collection data of time series data of sports performance $\{X_t\}$ at moment t; α is the smoothing factor, meeting the interval of $\alpha \in (0,1)$.

Then the forecast value of sports performance at moment $t+m$ is:

$$\hat{X}_{t+m} = a_t + b_t m + c_t m^2 . \tag{6}$$

In formula (20), m is the predicted visual length of the sports performance data, generally taken as positive integer $1, 2, 3, \cdots$. For example, $m = 1$ represents the one-step predictive value of the sports performance data; besides, $\hat{X}_1 = X_1$ and $\hat{X}_2 = X_2$ are set. Then the forecast parameters can be expressed as:

$$\begin{cases} a_t = 3S_t^{(1)} - 3S_t^{(2)} + S_t^{(3)} \\ b_t = \dfrac{\alpha}{2(1-\alpha)^2}[(6-5a)S_t^{(1)} - (10-8a)S_t^{(2)} + (4-3a)S_t^{(3)}] \\ c_l = \dfrac{\alpha^2}{2(1-\alpha)^2}(S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)}) \end{cases} \tag{7}$$

There are two problems in the triple smoothing exponential forecast process needing to be dealt with: (1) the setting of initial values $S_1^{(1)}$, $3S_1^{(2)}$, $S_1^{(3)}$ of the smoothing parameter; (2) the setting of smoothing parameter α. For a more reasonable distinction between segmentation point and outliers, the practice is to set flag bit of check link to determine the possible segmentation point. And then check the next adjacent data point. If the point also does not accord with their historical trend of the sports performance data, then the point shall be the desired segmentation point and then set mark to show that the segmentation point is the desired outlier while clearing the set flag point.

In the above steps, the number of segmentation points can be increased: $n\,(n \geq 2)$, if n continuous sports performance data of candidate adjacent segmentation points accord with the historical trend, then set the corresponding flag bit, and set the point as the segmentation point. Here $n = 1$ is chosen to check the data sequence of segmentation point. During the data transmission, as the probability of data outliers produced by disturbance is small, the probability that both continuous points are outliers is smaller.

3.3. Algorithm flowchart

In the segmentation process of sports performance time series, the main purpose is to reduce the dimension of the time series data and the key problem is the rapid detection of segmentation points. The use of the smoothing exponential forecast of sliding window is the use of segmentation algorithm for the trend forecast of the historical time series data of sports performance to obtain the mean square value between the historic value and forecast value of data. The compression rate and forecast error of sequence data are used to determine segmentation points, and it is found out that two sub-sequences formed by two segmentation points have independent distribution trend with a small volatility. The use of segmentation

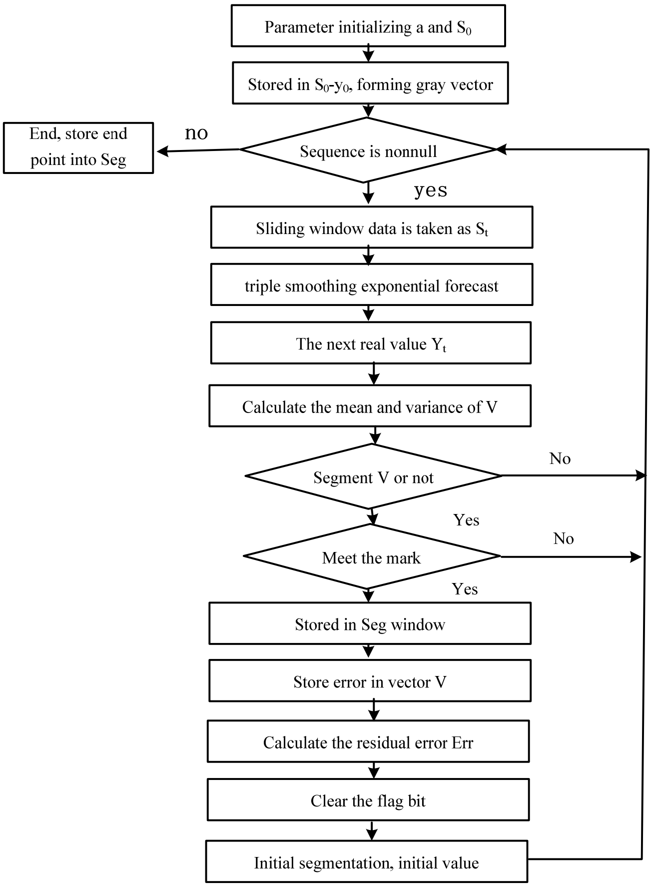

Fig. 1. Flow chart of exponential smoothing forecast segmentation algorithm

points forecast to further remove the outlier disturbance point in the check link is the target and main manner of improving the searching accuracy of segmentation points. Above searching methods only browse data once, therefore, no additional storage space is required for data storage, and the algorithm complexity can be reduced, and the real-time data requirement of the application environment can be better satisfied. The specific process is shown in Fig. 1.

In the smoothing exponential forecast process shown in Fig. 1, S_t is smoothing exponential forecast value, y_t is the real time series value of sports performance, s_0 is the initial exponential set value of segmentation algorithm, $s_0 = (y_0 + y_1 + y_2)/3$. α is the set weight of smoothing exponential, $\alpha = 0.2$, V is used to store the vector of forecast error, Seg is the set of segmentation points stored, and Err is the residual error value of segmentation point between the original sequence and the fitting sequence.

Seen from Fig. 1, in the smoothing exponential forecast algorithm, t is the current time point of the real-time data, V is used to store the forecast error values;

and based on the segment update strategy, the mean square error of V is calculated to obtain the forecast interval of the error. Based on the set compression rate and Lemma 1, the mean deviation degree x is calculated. After the parameter initializing, the smoothing single-exponential forecast is used to calculate the smoothing value S_t, which is regarded as the forecast value of the next moment of moment t. Then, the actual sports performance data y_t at moment t is collected and $y_t - s_t$ forecast error value is obtained, and store the value in the vector V. Thereafter, the standard deviation δ and mean μ of the vector are calculated and both parameters are updated. The fourth step, formula (16) is used to verify whether the segmentation point is appropriate, if not, then continue performing the next cycle point; if satisfied, then set the flag bit and the store the segmentation point. If the next data points also meets the requirements of the segmentation point, then store the point in Seg, and re-initial the initial S_0 at segmentation section and continue performing the cycle operation.

4. Experimental analysis

4.1. Experimental environment

In order to test the effectiveness of the forecast model of national traditional sports performance in colleges and universities based on triple exponential forecast of sliding window, 500 100m running results (unit: s) of Wuchang Institute of Technology were selected as the experimental subjects, and the model was realized by VC ++ 6.0 programming. The results of 10m running are shown in Fig. 2.

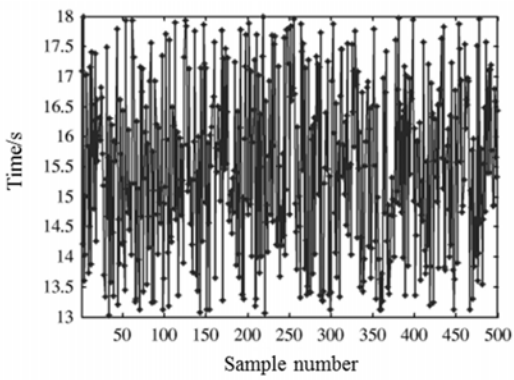

Fig. 2. 100m running performance

4.2. Normalizing operation of sports performance data

Conduct linear regression stepwise screening on the influence parameters of sports performance shown in Table 1 to obtain the regression data in Table 1.

Table 1. Regression data of influence parameters of sports performance

	Partial correlation	t value	p value
$r(y, X1) =$	0.6928	1.6659	0.1937
$r(y, X3) =$	-0.7086	1.7813	0.7653
$r(y, X4) =$	0.2094	0.3796	0.7138
$r(y, X5) =$	0.9618	5.3795	0.0137
$r(y, X6) =$	0.3719	0.6873	0.5261
$r(y, X7) =$	-0.8608	2.9468	0.0593
$r(y, X8) =$	-0.9046	3.7045	0.0352

300 100-m-run results were used to form a training sample set that are trained with the target. First of all, the optimal weight and threshold of the optimal target were found based on the triple exponential forecast optimization algorithm of sliding window, and then the structure of the target was determined according to the optimal connection weight and threshold, and the performance forecast model of 100m run was established. Finally, the remaining 200 100m running performance were forecast, and the deviation between the forecast value and measured value of 100m running is as shown in Fig. 3. Through the analysis of the forecast results of 100m running performance in Fig. 3, it can be found out that the forecast value of 100m running was very close to the measured value, and the coincidence accuracy between both was very high, indicating that through the selection between the connection weight and threshold value of target based on the triple exponential forecast optimization algorithm of sliding window, a relatively better 10m running performance forecast model can be established, and the error between the forecast value and measured value of 100m running performance was minor and can be negligible totally. The change interval of error was narrow and the result verified the effectiveness of national traditional sports performance forecast model in colleges and universities based on triple exponential forecast algorithm optimization target of sliding window, and the forecast result was reliable with small error.

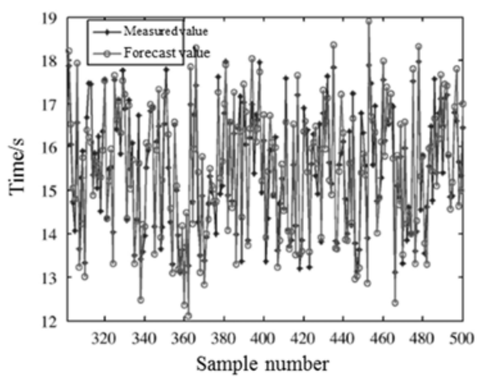

Fig. 3. Forecast effect of 100m running based on triple exponential forecast of sliding window

The proposed method in this paper is used to forecast the sports performance trend of the city from 2004 to 2014 and the SVM algorithm, the gray scale SVM algorithm and the triple exponential forecast algorithm are selected as the comparison algorithm. The forecast results obtained are shown in Table 2.

Table 2. Comparison of forecast performance

Forecast model	Root mean square error	Operating time (s)
SVM algorithm	3503.67	35.62
Gray scale SVM algorithm	2659.21	39.58
triple exponential forecast algorithm	1986.25	25.49
Propose algorithm in this paper	568.19	14.27

Seen from the data in Table 3, in terms of forecast accuracy, the proposed algorithm in this paper outperforms SVM algorithm, gray scale SVM algorithm, and triple exponential forecast algorithm. The proposed algorithm in this paper has the smallest root mean square error and it obviously outperforms the comparison algorithms selected in the operating time, showing the advantages of the proposed algorithm in this paper in the calculation accuracy and operational efficiency.

Through the analysis of forecast accuracy in Table 2, the following conclusions can be drawn: (1) multiple linear regression model of 100m running performance forecast has the lowest forecast accuracy, which shows that multiple linear regression model can not reflect the change characteristics of 100m running performance, and the model established has large forecast error and lower practical application value. (2) The forecast accuracy of 100 m running performance of genetic algorithm optimization target and particle swarm optimization target is obviously better than that of multiple linear regression model because the target is a non-linear strong modeling ability algorithm, and it can reflect the change of 100m running results and obtain better forecast effect. However, the forecast result of individual point is not satisfactory.

(3) The forecast accuracy of 100m running performance based on triple exponential forecast of sliding window is higher than that of genetic algorithm optimization and particle swarm optimization target. This is because the triple exponential forecast optimization algorithm of sliding window solves the problem that genetic algorithm and particle swarm optimization (PSO) hardly find out the connection weight and threshold of the global optimal target, reflects the change trend of the 100m running result more accurately and obtains a better forecast result of the 100m running performance.

5. Conclusion

In order to improve the forecast accuracy of national traditional sports performance in colleges and universities, a forecast model of national traditional sports performance in colleges and universities based on triple exponential forecast of sliding window is proposed aiming at the difficulty in determining the weight and threshold of target. First of all, the sports performances of national traditional sports performance in colleges and universities are prepossessed to generate training sample and testing sample of the target. Then triple exponential forecast optimization algorithm of sliding window is used to determine the connection weight and threshold of the target. Through training sample learning, the national traditional sports performance forecast model in colleges and universities is established. Finally, the forecast effect of the model is tested through the specific simulation experiment. The results show that the triple exponential forecast of sliding window improves the forecast accuracy of the national traditional sports performance in colleges and universities, and solves the limitations of other forecast models of the national traditional sports performance in colleges and universities, and its forecast results are more reliable, which can provide scientific decision-making basis for the national traditional sports training in colleges and universities.

References

[1] Y. Z. Chen, F. J. Tang, Y. Bao, Y. Tang, G. Chen: *A Fe-C coated long period fiber grating sensor for corrosion induced mass loss measurement.* Optics letters *41* (2016), 2306–2309.

[2] N. Arunkumar, S. Jayalalitha, S. Dinesh, A. Venugopal, D. Sekar: *Sample entropy based ayurvedic pulse diagnosis for diabetics.* IEEE-International Conference on Advances in Engineering, Science and Management, ICAESM-2012, art. no. 6215973 (2012) , 61–62.

[3] Y. Song, N. Li, J. Gu, S. Fu, Z. Peng, C. Zhao, Y. Zhang, X. Li, Z. Wang, X. Li: *β-Hydroxybutyrate induces bovine hepatocyte apoptosis via an ROS-p38 signaling pathway.* Journal of Dairy Science *99* (2016), No. 11, 9184–9198.

[4] N. Arunkumar, K. R. Kumar, V. Venkataraman: *Automatic detection of epileptic seizures using permutation entropy, Tsallis entropy and Kolmogorov complexity.* Journal of Medical Imaging and Health Informatics *6* (2016), No. 2, 526–531.

[5] R. Hamza, K. Muhammad, N. Arunkumar, G. R. González: *Hash based Encryption for Keyframes of Diagnostic Hysteroscopy*, IEEE Access, https: //doi.org/10.1109/ACCESS.2017.2762405 (2017).

[6] J. W. Chan, Y. Y. Zhang, and K. E. Uhrich: *Amphiphilic Macromolecule Self-Assembled Monolayers Suppress Smooth Muscle Cell Proliferation*, Bioconjugate Chemistry *26* (2015), No. 7, 1359–1369.

[7] N. Arunkumar, K. Ramkumar, S. Hema, A. Nithya, P. Prakash, V. Kirthika: *Fuzzy Lyapunov exponent based onset detection of the epileptic seizures.* 2013 IEEE Conference on Information and Communication Technologies, ICT 2013, art. no. 6558185 (2013), 701–706.

[8] J. J. Faig, A. Moretti, L. B. Joseph, Y. Y. Zhang, M. J. Nova, K. Smith, and K. E. Uhrich: *Biodegradable Kojic Acid-Based Polymers: Controlled Delivery of Bioactives for Melanogenesis Inhibition*, Biomacromolecules *18* (2017), No. 2, 363–373.

[9] N. Arunkumar, V. Venkataraman, Thivyashree, Lavanya: *A moving window approximate entropy based neural network for detecting the onset of epileptic seizures.* International Journal of Applied Engineering Research *8* (2013), No. 15, 1841–1847.

[10] Y. J. Zhao, L. Wang, H. J. Wang, and C. J. Liu: *Minimum Rate Sampling and Spectrum Blind Reconstruction in Random Equivalent Sampling.* Circuits Systems and Signal Processing *34* (2015), No. 8, 2667–2680.

[11] S. L. Fernandes, V. P. Gurupur, N. R. Sunder, N. Arunkumar, S. Kadry: *A*

novel nonintrusive decision support approach for heart rate measurement. Pattern Recognition Letters. https: //doi.org/10.1016/j.patrec.2017.07.002 (2017).

[12] N. Arunkumar, K. Ramkumar, V. Venkatraman, E. Abdulhay, S. L. Fernandes, S. Kadry, S. Segal: *Classification of focal and non focal EEG using entropies.* Pattern Recognition Letters *94* (2017), 112–117.

Received May 7, 2017