Acta Technica 62 (2017), No. 7A, 1309–1320

Research on marketing strategy based on parallel fuzzy c-means clustering under new economic background

Haitao Li [1]

Abstract. With the rapid development of network technology under new economic background, especially "Internet+" and big data development in recent years, the great opportunities and challenges are brought to enterprise product marketing. Compared with traditional behavior where marketing goals are achieved through a simple and crude price war, some enterprises win the trade war by making full use of and excavating the data. Behavior information of user is excavated from big data of the Internet by utilizing data fusion technology. This paper realizes synchronization of member management and Mapreduce reduction operation by utilizing improved Mapreduce parallel computing mechanism and adopting member management protocol. The typical individual group reduction operation is implemented to replace global individual reduction operation and define two-stage buffering algorithm, to further reduce data size of Mapreduce operation at the second stage through buffering of the first stage, and reduce negative effects of marketing noise data on algorithm as far as possible. Finally, simulation results show that the algorithm can not only guarantee clustering accuracy, but also accelerate operating efficiency of algorithm effectively, which verifies effectiveness of proposed method in marketing strategy research.

Key words. New economy, Parallel computing, Fuzzy evaluation, Clustering, Marketing strategy

1. Introduction

With the development of Internet technology, network marketing has become an important way of product sales, which also is a competitive place for enterprise business war. Emergence of network has broken restriction of traditional enterprise products to sales model and caused a great impact on marketing strategy of traditional products. Traditional enterprise product marketing mode is based on a large amount of consumption of human and material resources and intermediate links,

[1] Rizhao Polytechnic, Shandong Rizhao, 276826, China

while network marketing realizes product transaction through Internet technology. Reduction of intermediate links and channels greatly reduces enterprise cost, improves transactional efficiency and optimizes resource allocation among enterprises.

On May 7, 2015, the State Council issued Opinion on Vigorously Developing E-commerce and Quickening New Economic Power Cultivation. The Opinion clearly pointed out that e-commerce should be vigorously developed, and e-commerce admittance threshold should be reduced simultaneously, to build unified, open, orderly competitive, honest and law-abiding, and safe and reliable big e-commerce market before 2020. Under e-commerce environment, enterprises can provide consumers with more individualized products, and even customized products, through network marketing. With the rise of cloud computing and big data technology, enterprise managers can even excavate potential value of data and understand needs of customers from a large number of user purchase behavior data, so as to grasp market discipline and formulate effective marketing strategies. Seen from information service perspective, marketing strategy constructed in this paper can tail after demand change of user promptly through analysis to user habit and preference behavior etc., thus actively adjusting the content and way of information service, and recommending information and services to users that they are interested in under customized mode; seen from product perspective, it can explore long tail of objects better.

At present, the relatively common product marketing recommendation strategy can be divided into 3 types: product recommendation based on content, product recommendation based on collaborative filtering, and mixed product recommendation. Product recommendation based on content is to find similar objects according to content information of objects loved by users and make recommendation. Product recommendation based on collaborative filtering is to analyze similarity among users based on historical information of user, so as to make recommendation according to the mechanism that similar users have similar preferences. Relatively common mixed product recommendation is the mixture of recommendation based on content and recommendation based on collaborative filtering. Current product recommendation service takes advantage of the interest preference of similar users, or similar objects loved by users as the basis for recommendation, but upon the personalized demands of consumers, it is basically rugged, which means that the product recommendation system treats all users as a person (or person of a class). Such recommendation result based on rugged mode will bring a quite negative user experience to potential users, for example, in e-commerce environment, occupying a user's mailbox with chronically unrelated mails in a large amount of space will cause a great disgust in the user, and the punishment for them may be simple and cruel, where their mailbox addresses are sent to the trash mailbox with light clicking to mouse, so no more promotional information can be sent to the potential customer.

Therefore, it is quite necessary to screen, filter and individually analyze recommended results according to behavior information of user. Especially in the era of big data, to make extremely rich user behavior information obtaining possible, this paper will deeply analyze and excavate user behavior, and construct collaborative filtering recommendation system based on model by utilizing parallel fuzzy c-Means clustering algorithm (PGR-PFCM), thus realizing personalized analysis and filtering to product recommendation and improving product marketing accuracy.

2. Personalized recommendation and marketing strategy on user behavior

User behavior information is all behaviors of user on the Internet, it has various forms of performance. For example, behavior information before the purchase can deeply reflect purchasing psychology and purchase intention of potential customers. Through analysis and understanding of such behavior information, intimate service and personalized recommendation to user are made. For example, in e-commerce, such personalized recommendation service generally has quite good effect, which can improve purchase intention of user, shorten purchase path and time, and capture optimal purchase impulse of user at relatively proper moment, and also reduce unwarranted harassment of traditional marketing mode to user, and improve user experience.

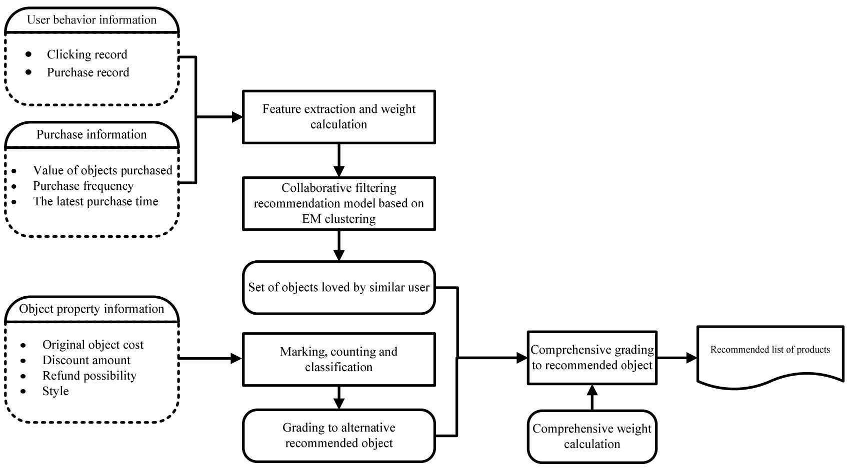

Fig. 1. Recommended process for comprehensive calculation of user behavior and product properties

This paper excavates personalized demand of user by utilizing user behavior data, to realize product recommendation and personalized marketing service. Quantify past clicking behavior and purchase behavior etc. of user, value of objects purchased by user, purchase frequency and the latest purchase time etc., and generate user behavior space vector, so as to construct collaborative filtering recommendation model by utilizing EM clustering algorithm, and realize clustering analysis of similar user; mark, count and classify original object cost, discount amount, refund possibility, and style etc., and grade preference condition of alternative recommended object at current stage. Final recommended results can be obtained by endowing different weight factors to recommended results related to factors in two aspects and making weight calculation. The recommended process for comprehensive calculation of user behavior and product properties is as shown in Fig.1.

3. PGR-PFCM marketing recommendation model

3.1. Model introduction

Mapreduce model is a relatively mature distributed computing model of big data at present, and also is main execution framework realized by parallel clustering algorithm of big data. The name of Mapreduce is sourced from two main operations of the framework: map and reduce operation. Map operation is similar to a kind of mapping operation, being a kind of operation aimed at all members of dataset, and return to result list after map operation. Reduce operation executes parallel algorithm aimed at feedback result of map operation, as shown in Fig.2.

In Mapreduce model, problem processes are decomposed into mutually independent sub-processes that can be operated parallelly, so that computing performance of computer cluster can be brought into full play. Map and reduce operation framework is mainly designed and operated according to key-value in data:

$$Map : (k_1, v_1) \to [(k_2, v_2)] \ . \tag{1}$$

$$Reduce : (k_2, [v_2]) \to [(k_3, v_3)] \ . \tag{2}$$

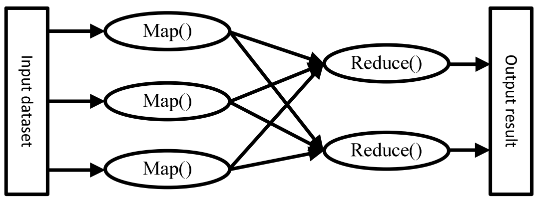

Fig. 2. Mapreduce model

Under system operation, management host will dispatch quantity of computers participating in parallel computing in real time according to breakpoint condition of input data, and computer failure problem can also be handled in real time through management host. Therefore, programmer is allowed to easily use large distributed data resources without operating experience and hardware operation experience.

3.2. Typical individual operation scheme

Mapreduce parallel computational model under traditional Hadoop platform is a kind of parallel calculation strategy, and the model adopts group individual reduction strategy[9], which means to adopt the operation of replacing group members with typical individual so as to reduce data processing quantity (as shown in Fig.3). In the typical individual operation mode, design operation process identifier list (pID) to map ID of process into group IDS (mID), and then make global reduction strategy to group. Time complexity of global individual reduce operation algorithm is $\log_2(p)$, but in grouping scheme based on typical individual operation, reduce operation algorithm time complexity is $\log_2(p')$, where $p' = \max_k(|P_k|)$, P_k is the number of typical individual of process subset, and therefore, in algorithm implementation process, we hope $p' \sim p$, which can be realized by controlling grouping quantity and typical individual selection condition.

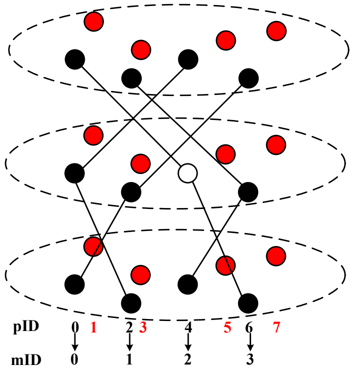

Fig. 3. Data reduction strategy

Literature [10] proposes that group-based typical individual reduction strategy can realize information communication by defining MPI communication function in group. But because operation process identifier list will change continuously with iteration, which means that object combination for group reduction changes continuously, and process identifier list needs to be changed for each iteration, such that excessive communication overhead will be caused through MPI communication function mode. Because with the increase of iteration, gravity center can get closer to the center of clustering, clustering mode that only has 1 gravity center and 1 process subset shall be considered firstly. To solve the problem that MPI communication function time occupation ratio is excessive, synchronized group member management protocol is designed for traditional MPI communication model for replacement, and operation based on group reduction can be synchronously operated in process subset in parallel and mutually independent mode, applicable for parallel algorithm embedding.

3.3. Synchronization management protocol

A typical individual dynamic of protocol group-reduce shall broadcast the pID list by encapsulating the control information of dynamic group management in the operation information and not need to add extra communication overhead of algorithms as each group reduces operations to reach synchronous modification. The operation steps of protocol group-reduce can be shown as follows:

Step1: The group typical individual management protocol shall be defined, and definite clearly that leader shall manage the typical individual. Use the reference list L_i^{cur} of typical individual for the course p_i, and the next operation shall be in accordance with L_i^{cur}.

Step2: Initializing and the reference list L_i^{cur} shall be broadcasted by leader, and initializing L_i^{cur} includes all default courses.

Step3: The reduction operation for group shall be carried out in accordance with reference listL_i^{cur}, the course p_i will make a partial pID list $L_{i,s}$ $(L_{i,0} = L_i^{cur})$ for the step s of communication, which includes in the control information of group reduction operation.

Step4: Information shall be exchanged in the course i, j when step s, which includes the combination with long-distance and local pID list, and combination formula is $L_{i,s+1} = L_{i,s} \oplus L_{j,s}$.

Step5: If the contribution value g_i of the course p_i is lower than threshold value $T_{\min}$, the course shall be taken out of pID list. The contribution value [10]:

$$g_i = \frac{L_i}{\sum\limits_{j=1}^{n} L_j}\,. \tag{3}$$

The external course p_m can transmit the contribution value to the local group contacted with leader by asynchronization, which includes the contribution degree of the next reduction operation for p_m, and add ID of the external course p_m into the mID of local group. If the leader accepts the course p_m, the pID list shall be sent to the course p_m.

Step 7: Output the result mID.

4. PGR-PFCM Clustering Algorithm

4.1. Fuzzy c-Means Algorithm

Suppose that there are n data point $x_1, x_2, \cdots, x_n$, of which, the data $x_i \in R^d$, the traditional clustering method is that each data point shall be given a gather (or category), and Suppose that c kinds of categories can be generated, c of clustering center $\{v_1, v_2, \cdots, v_c\}|\, v_j \in R^d$ shall be firstly calculated, v_j can be called as prototype point of clustering. Fuzzy degree of membership function c of each data point shall be calculated for fuzzy c-means algorithm, which can be represented as $u_{ji} \in [0,1]$, $j = 1, \cdots, c$, $i = 1, \cdots, n$. Especially for the random $i \in [1, \cdots, n]$, meet the condition $u_{ji} \in [0,1]$, $\sum_{j=1}^{c} u_{ji} = 1$, the concept of membership function can be equivalent to the traditional accuracy clustering algorithm. The objective function of fuzzy c-Means algorithm is as follow:

$$J = \sum_{i=1}^{n} \sum_{j=1}^{c} (u_{ji})^m \|x_i - v_j\|^2\,. \tag{4}$$

Meets:

$$\sum_{j=1}^{c} u_{ji} = 1\,. \tag{5}$$

In formula (4), $\|\cdot\|$ represents inner product norm. Fuzzy c-Means algorithm uses the following formulas for the value of v_j and u_{ji} to do iteration, so as to achieve the purpose to minimize the objective function J:

$$v_j = \frac{\sum_{i=1}^{n} (u_{ji})^m x_i}{\sum_{i=1}^{n} (u_{ji})^m} . \tag{6}$$

$$\begin{cases} u_{ji} = \left(\sum\limits_{k=1}^{c} \left(\dfrac{d_{ji}}{d_{ki}} \right)^{\frac{2}{m-1}} \right)^{-1} \\ d_{ji} = \|x_i - v_j\| \end{cases} \tag{7}$$

Suppose that the end condition of iteration is $Max \left\{ \left\| u_{ji}^{(t+1)} - u_{ji}^{(t)} \right\| \right\} < \varepsilon$, of which, $m \in [1, \infty)$ acts as weight coefficient, clustering will be gradually clear and definite with $m \to 1$. Clustering will be gradually fuzzy with $m \to \infty$, $m = 2$ is usually taken for calculation. Pseudocode of fuzzy c-Means algorithm (PFCM) can be shown as follows:

Procedure 1: Pseudocode of PFCM algorithm

```
% Pseudocode of fuzzy c-Means algorithm (PFCM)
Function P = PFCM();
randomise my_uOld[j][i] for each x[i]
do {
 maxErr=0;
 for j=1 to c
   myUsum[j]=0;
   reset vectors my_v[j] to 0;
   reset my_u[j][i] to 0;
 endfor;
 for i=myid*(n/P)+1 to (myid+1)*(n/P)
   forj=1 to c
     update myUsum[j];
     update vectors my_v[j];
   endfor;
 endfor;
 for j=1 to c
   Allreduce(myUsum[j], Usum[j], SUM);
   Allreduce(my_v[j], v[j], SUM);
   v[j] = v[j]/Usum[j];
 endfor;
  for i=myid*(n/P)+1 to (myid+1)*(n/P)
   for j=1 to c
      update my_u[j][i];
     maxErr=max{|my_u[j][i]-my_uOld[j][i]|};
     my_uOld[j][i] = my_u[j][i];
   endfor;
  endfor;
  Allreduce(maxErr, Err, MAX);
  } while (Err>=epsilon)
```

4.2. Description of PGR-PFCM clustering algorithm

Using a node type similar to Combiner, Mapper operation and reduce operation shall be performed in series in the algorithm iteration process, the steps of PGR-PFCM algorithm can be shown as follows:

Step1: Input parameter. The clustering amount K_t in intermediate buffering, clustering amount K, and length of data segment L.

Step 2: Mapper operation. Perform map operation (inValue, inKey, outValue, outKey)

(a)Load clustering data in inValue and perform map operation.

(b)Perform the operation of $PFCM\,(K_t)$ map clustering in outValue data.

(c)Reformat the value of outValue according to clustering result.

(d)(outValue, outKey) Output (outValue, outKey).

Step3: Reducer operation. Perform protocol group-reduce operation of synchro management protocol for (inValue, inKey, outValue, outKey) operation.

(a)Load clustering intermediate result of Mapper operation in inValue and perform protocol group-reduce operation.

(b)Operate $PFCM\,(K)$ clustering operation in clustering intermediate result of protocol group-reduce operation.

(c)Output the final clustering result.

In the above steps, the value of parameter K_t directly decides the operating speed and clustering quality of PGR-PFCM algorithm. The first stage of PGR-PFCM algorithm can play a role in compressing data quality by defining a clustering amount K_t in intermediate buffering. The size of parameter K_t decides the size of data input by PGR-PFCM algorithm in the second stage, so as to influence the operation speed and accuracy of algorithm, the larger K_t is, the slower speed and the higher accuracy will be. The smaller K_t is, the faster speed and the higher accuracy will be.

5. Experimental analysis

5.1. Experimental data

The experiment uses the data gotten from online bookstore of Amazon as experimental data set, which includes 543 users, records of 2682 books and scores (1-5) for 100000 times. Each user will at least mark and grade to 5 book records or interesting books, and each book will be graded by one user at least. Training set and test set provided in data set account for 80% and 20%, respectively, use training set for training, and forecast the score of unknown books, and compared and experimented with the actual scores of books graded by users in test set.

5.2. Evaluation criterion

For recommendation effect of system, the performance of recommendation strategy is evaluated by calculating mean absolute error (MAE). Suppose that a user u

and goods i in test set, set r_{ui} is the actual score of good i for user u, but $\tilde{r}_{ui}$ is prediction score given by recommended algorithm, MAE can be defined as:

$$MAE = \frac{\sum_{u,i \in T} |r_{ui}, \tilde{r}_{ui}|}{|T|}.$$

5.3. Experimental results and analysis

The recommendation of goods consists with collaborative filtering based on user behavior and user rating in this paper, of which proportion shall be calculated by balance factor α. The value range of α can be set as $[0, 1]$, increase 0.1 at each time, compared the influence on MAE, which can be shown as Figure 4.

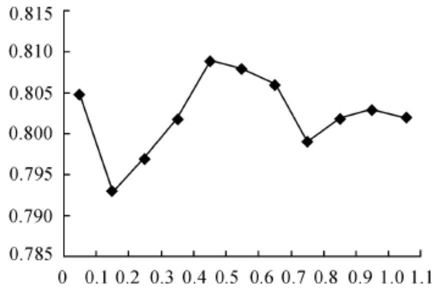

Fig. 4. Experiment for balance factor influence on MAE

When $\alpha = 0.2$, MAE is minimum and with optimum recommendation effect from Figure 4, which shows that users have a greater impact on the content of the book information, and is consistent with the actual situation. In real life, users have a preference for book content, and the types of books purchased or be interested in are similar, even though users may not be able to score highly for some books in the category, there is still a tendency to choose books.

The performance of the algorithm is compared from the accuracy and efficiency of the evaluation algorithm. Eight types of true marketing data sets shall be used for generating larger mixed data sets as test data set, which are composed of 10000 vector quantities, respectively, the dimension of test data set in each node can meet the geometric multiples of 2 (2,4,8,16), and test marketing data sets obey the normal distribution of [0, 1]. Literature [11] selects clustering cost to evaluate the accuracy of K-means II algorithm, three algorithms of K-means II, FPCM[12] and PGR-PFCM are simulated by this method in the experiment, all experiments run for 30 times to get the average value and the simulation results can be shown in Figure 5-6.

Figure 5 shows the costs (iteration steps) of the three algorithms in the marketing test data set, from which can be seen that the costs of the three algorithms are not significantly different at low dimensions. With the increase of marketing data dimension, computational costs of the three algorithms are rapidly increasing, and the cost gap between each other is widening, which reflects the performance superiority of the algorithm proposed in the high-dimensional marketing data set in this paper. Figure 6 shows the time operation comparison of the algorithm, which shows that due to the small difference of the cost algorithm in the low-dimensional marketing data, while time operation of the algorithm increases due to the increase of the protocol list operation in the PGR-PFCM algorithm. However, with the increase of marketing data dimension, the growth rate of operation time for PGR-PFCM algorithm is less than the other two comparison algorithms. Figure 3-4 reflects the operation efficiency advantage of the PGR-PFCM algorithm. Table 2 gives the clustering success rate of the three algorithms under each dimension data set (the algorithm runs 30 times to get the average value), of which can be seen that the average clustering success rate of PGR-PFCM in the three algorithms is the highest, followed by K-means II algorithm, and the success rate of FPCM algorithm is the lowest.

Table 2. Comparison of clustering success rate (operation for 30 times

Data Set Dimension	Algorithm	Clustering Success Rate
2	FPCM	73.3%
	K-means II	82.7%
	PGR-PFCM	94.6%
4	FPCM	72.3%
	K-means II	77.5%
	PGR-PFCM	90.1%
8	FPCM	66.8%
	K-means II	71.2%
	PGR-PFCM	86.9%
16	FPCM	60.8%
	K-means II	68.6%
	PGR-PFCM	82.4%

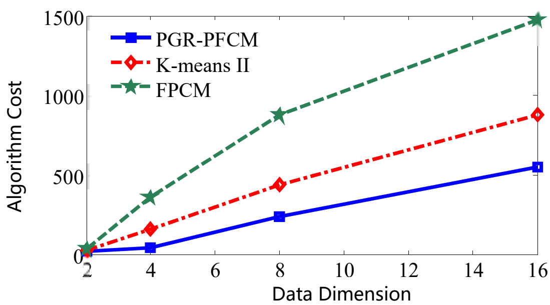

Fig. 5. Algorithm computation cost comparison

6. Conclusion

This paper explores the new way of network product for recommendation and marketing according to the situation of enterprise management under the current

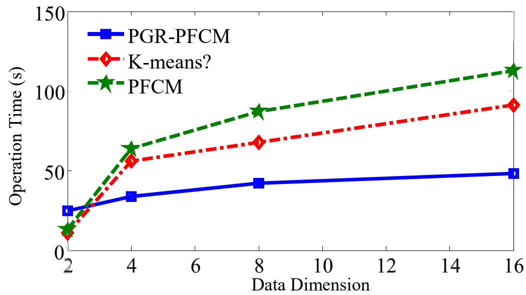

Fig. 6. Algorithm computation time operation comparison

e-commerce environment. Especially, the development of "Internet plus" and big data technology in recent years has brought great opportunities and challenges to enterprise product marketing. Compared with the traditional marketing purposes by simple and crude price war, the use of data fusion technology to dig for users' behavior information from Internet big data is more conducive to analyzing the individual needs of consumers, recommending the products they may like to consumers and carrying out personalized active marketing services, Meanwhile, it is more conducive to establishing the corresponding personalized product marketing strategy, so as to improve the number of product sales and success rate of product recommendation.

References

[1] D. S. Abdelhamid, Y. Y. Zhang, D. R. Lewis, P. V. Moghe, W. J. Welsh, and K. E. Uhrich: *Tartaric Acid-based Amphiphilic Macromolecules with Ether Linkages Exhibit Enhanced Repression of Oxidized Low Density Lipoprotein Uptake*, Biomaterials *53* (2015), 32–39.

[2] W. Pan, S. Z. Chen, Z. Y. Feng: *Automatic Clustering of Social Tag using Community Detection.* Applied Mathematics & Information Sciences *7* (2013), No. 2, 675–681.

[3] Y. Y. Zhang, E. Mintzer, and K. E. Uhrich: *Synthesis and Characterization of PEGylated Bolaamphiphiles with Enhanced Retention in Liposomes*, Journal of Colloid and Interface Science *482* (2016), 19–26.

[4] N. Arunkumar, K. M. Sirajudeen: *Approximate Entropy based ayurvedic pulse diagnosis for diabetics - A case study.* TISC 2011 - Proceedings of the 3rd International Conference on Trendz in Information Sciences and Computing, art. no. 6169099 (2011) , 133–135.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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).

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

[12] 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.

[13] M. P. Malarkodi, N. Arunkumar, N. V. Venkataraman: *Gabor wavelet based approach for face recognition.* International Journal of Applied Engineering Research *8* (2013), No. 15, 1831–1840.

[14] L. R. Stephygraph, N. Arunkumar: *Brain-actuated wireless mobile robot control through an adaptive human-machine interface.* Advances in Intelligent Systems and Computing *397* (2016), 537–549.

Received May 7, 2017