



Copyright © IEEE.
Citation for the published paper:

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of BTH's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by sending a blank email message to pubs-permissions@ieee.org.

By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

A Hybrid Computational Method for the Identification of Cell Cycle-regulated Genes

Veselka Boeva and Petia Ivanova

Computer Systems and Technologies Department
Technical University of Sofia, branch Plovdiv
Plovdiv, Bulgaria
vboeva@tu-plovdiv.bg and petini84@yahoo.com

Niklas Lavesson

School of Computing
Blekinge Institute of Technology
SE-371 25 Ronneby, Sweden
Niklas.Lavesson@bth.se

Abstract—Gene expression microarrays are the most commonly available source of high-throughput biological data. They have been widely employed in recent years for the definition of cell cycle regulated (or periodically expressed) subsets of the genome in a number of different organisms. These have driven the development of various computational methods for identifying periodical expressed genes. However, the agreement is remarkably poor when different computational methods are applied to the same data. In view of this, we are motivated to propose herein a hybrid computational method targeting the identification of periodically expressed genes, which is based on a hybrid aggregation of estimations, generated by different computational methods. The proposed hybrid method is benchmarked against three other computational methods for the identification of periodically expressed genes: statistical tests for regulation and periodicity and a combined test for regulation and periodicity. The hybrid method is shown, together with the combined test, to statistically significantly outperform the statistical test for periodicity. However, the hybrid method is also demonstrated to be significantly better than the combined test for regulation and periodicity.

Keywords—computational method; cell cycle-regulated genes; P -value for regulation; P -value for periodicity; hybrid aggregation

I. INTRODUCTION

Microarray profiling studies of several different highly synchronized cell cultures in fission yeast (*Schizosaccharomyces pombe*) have been widely employed in recent years for the identification of periodically regulated genes during the cell cycle [1,2,3]. The latter genes are expressed only at a specific stage of the cell cycle and consequently, exhibit a periodic pattern of expression when monitored during consecutive cell cycles. Various computational methods for identifying periodically expressed genes have been developed for studying the cell cycle transcription program in a number of model organisms. However, the agreement is remarkably poor when different computational methods are applied to the same data. In view of this, we propose here a method for the identification of periodically expressed genes, which is based on a hybrid aggregation of estimations, generated by different computational methods.

In the paper of de Lichtenberg *et al.* [4], several computational methods for the identification of periodically expressed genes have been benchmarked. The authors have also proposed a new permutation-based method quantifying separately both the periodicity and the amplitude of variation, and have shown that amplitude-dependant methods perform better than the amplitude independent ones. In the benchmark study of de Lichtenberg *et al.* [4], a combined P -value was obtained by simply multiplying the P -value for regulation and the P -value for periodicity. However, this definition entails the negative side effect that the total P -value could become very low due to only one of the individual P -values. Therefore, two penalty terms, which are not intuitively well based, were further introduced in [4]. A more straightforward approach was taken in [5], where the individual P -value for regulation and periodicity are combined through their geometric mean and thus the combined P -value can be seen as a sort of trade-off between the individual P -values, since it always ranges between their minimum and maximum. The geometric mean may as well produce a significantly low total P -value due to the occurrence of a very low individual P -value. Therefore, Hermans and Tsiporkova [5], introduced two separate significance conditions that both need to be verified in order to quantify a gene as a significantly periodic. The method of periodicity estimation employed in [5] is different from the one used by de Lichtenberg *et al.* [4]. Namely, it is based on a slightly adapted version of the method used by Merges *et al.* [6,7], described originally by Shedden and Cooper [8]. Other multiple hypothesis testing methods have as well been presented in the bioinformatics literature, as *e.g.* [9,10].

In view of the above, we are motivated to develop a method for the identification of periodically expressed genes, which can take into account the individual P -values, generated by different computational methods. We propose herein a hybrid aggregation algorithm, which is based on a mathematically well motivated aggregation operator developed in [11]. This aggregation algorithm will be used to integrate the individual P -values, produced by different computational methods of gene significance estimation, into a single overall value by employing a vector of different aggregation operators. Each one of these aggregation operators exhibits certain shortcomings when used individually. However, as it was

shown in [11], the aggregation operator that is defined by a vector of aggregation operators acts as a trade-off between their conflicting behaviour. In this way, different aspects of the integrated P -values will be taken into account during the aggregation process.

The proposed hybrid computational method is benchmarked against three other computational methods for the identification of periodically expressed genes: statistical tests for regulation and periodicity and the combined test for regulation and periodicity are all defined by de Lichtenberg *et al.* in [4]. The four considered computational methods are investigated and compared on gene expression time series data coming from a study examining the global cell-cycle control of gene expression in fission yeast *Schizosaccharomyces pombe* [3]. The used procedures are implemented in C# and the corresponding software is available upon request.

II. METHODS

A. P -value for Regulation

A P -value for regulation, referred to as P_{reg} , can be calculated as described by de Lichtenberg *et al.* [4]. Namely, a P -value for regulation for a particular gene is resulting from the comparison of the gene expression variance with a randomly generated variance distribution, constructed by selecting at each time point the log-ratio value of a randomly chosen gene. The P -value for regulation is calculated as the fraction of artificial profiles with a variance equal to or greater than the score of the real expression profile.

B. P -value for Periodicity

To estimate a P -value for periodicity, referred to as P_{per} , de Lichtenberg *et al.* compared the Fourier score of the observed gene expression profile for each gene to those of random permutation of the same gene [4]. Thus the P -value for periodicity is calculated as the fraction of artificial profiles with Fourier scores equal to or large than that observed for the real expression profile.

Hermans and Tsiporkova employed in [5] a method of periodicity estimation which is different from the one used by de Lichtenberg *et al.* [4]. Namely, the periodic score is based on an adapted version of the method used by Merges *et al.* [6,7]. The observed expression profile of each gene is fitted to a periodic component consisting of a sine, a cosine and an amplitude offset. The new element is the introduction of an amplitude (vertical) offset parameter to the periodicity component.

C. Combined P -value for Regulation and Periodicity

In the benchmark study of de Lichtenberg *et al.* [4], a combined P -value, referred to as P_{com} , was obtained by simply multiplying the P -value for regulation and the P -value for periodicity. However, this definition entails the negative side effect that the total P -value could become very low due to only one of the individual P -values. Therefore, two rather unintuitive penalty terms were further introduced in [4]:

$$P_{com} = P_{total} \cdot \left[1 + \left(\frac{P_{reg}}{thr} \right)^2 \right] \cdot \left[1 + \left(\frac{P_{per}}{thr} \right)^2 \right], \quad (1)$$

where $P_{total} = P_{reg} \cdot P_{per}$ and thr is a given significant threshold.

Hermans and Tsiporkova [5], took a more straightforward approach to combine the individual P -value for regulation and periodicity, namely through their geometric mean $\sqrt{P_{reg} \cdot P_{per}}$, with values always ranging between $\min(P_{reg}, P_{per})$ and $\max(P_{reg}, P_{per})$. However, the geometric mean may produce a significantly low combined P -value due to the occurrence of a very low individual P -value. Therefore, two separate significance conditions need to be verified [5]:

$$\sqrt{P_{reg} \cdot P_{per}} < thr \text{ and } \max(P_{reg}, P_{per}) < \lambda \cdot thr,$$

where λ is an individual significance trade-off $\lambda \geq 1$. A gene is qualified as significantly periodic if the P -values associated with it fulfill both conditions.

D. A Hybrid Aggregation of P -values generated by different computational methods

We propose herein a method for the identification of periodically expressed genes, which is based on a hybrid aggregation of the individual P -values, generated by different computational methods. The proposed aggregation model has been inspired by a work on non-parametric recursive aggregation, where a set of aggregation operators is applied initially over input values, and then again over the result of the aggregation, and so on until a certain stop condition is met [11]. This process defines an aggregation operator that uses a vector of aggregation operators and acts as a trade-off between their conflicting behaviour.

We consider a hybrid aggregation algorithm, which aims at integrating vectors of values into a single vector of overall values, by employing a set of k aggregation operators A_1, A_2, \dots, A_k . This algorithm, schematically presented in Fig. 1, will be applied to obtain the vector of the overall (hybrid) P -values by aggregating the P -values assigned to each gene in a given experiment by n different computational methods of gene significance estimation. Thus we are supposed to aggregate n different P -values associated with each gene. These values can initially be combined in parallel with the above k aggregation operators. Consequently, k new P -values (one per aggregation operator) are generated. The new values can be aggregated once more, generating again k new P -values. In this fashion, each step is modeled via k parallel aggregations applied over the results of the previous step. Thus, the final result is obtained after passing a few layers of aggregation. At the first layer, we have the initial vectors of P -values (one per computational method), generated by different methods of gene significance estimation, that are to be combined. Using a vector of aggregation operators new vectors are obtained and the next step is to combine these new vectors again using these aggregation operators. This process needs to be repeated again and again until the difference between the maximum and minimum values in each position of the currently calculated vectors are small enough to stop further aggregation.

Notice that applying some information about the performance of the considered computational methods, weights may be assigned to the corresponding P -values and further used in the aggregation procedure in order to obtain more realistic overall P -values. This will reflect only the initial aggregation step of the proposed aggregation algorithm, where the weighted versions of the involved aggregation operators will be applied initially over the input P -values.

In [11], it has been shown that any recursive aggregation process, defined via a set of continuous and strict-compensatory aggregation operators, following the algorithm described herein is convergent. A more detail explanation of the hybrid aggregation algorithm can be found in [11,13,14].

The proposed hybrid computational method has several advantages. First, it is based on a mathematically well motivated aggregation operator, developed and studied in [11]. The latter one is defined by a vector of aggregation operators and acts as a trade-off between their conflicting behaviour. In this way, different aspects of P -values, produced by different computational methods, will be taken into account during the aggregation process. Second, as it was discussed above weights may easily be introduced in the hybrid P -value calculation algorithm by assigning different importance to the involved computational methods. In addition, the calculated total P -value is not dependent on a particular significant threshold, which is the case with the combined P -value, proposed by de Lichtenberg *et al.* [4]. The latter means that it will not be necessary to recalculate the total (hybrid) P -value when a different significant threshold is used.

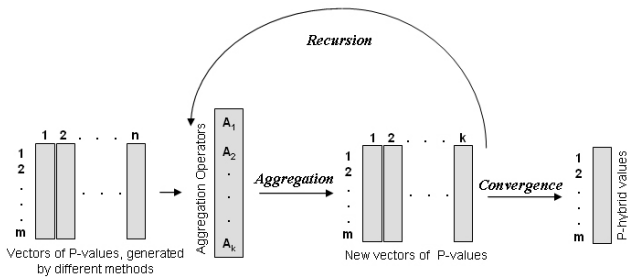


Figure 1. Schematic representation of the hybrid aggregation of P -values.

III. EXPERIMENTAL SETUP

A. Data

The considered computational methods are evaluated and compared on gene expression time series data obtained from a study examining the global cell-cycle control of gene expression in fission yeast *Schizosaccharomyces pombe* [3]. The study includes 8 independent time-course experiments synchronized respectively by 1) elutriation (three independent biological repeats), 2) *cdc25* block-release (two independent biological repeats, of which one in two dye-swapped technical replicates, and one experiment in a *sep1* mutant background), and 3) a combination of both methods (elutriation and *cdc25* block-release as well as elutriation and *cdc10* block-release). Thus, the following 9 different expression test sets are

available: 1) *elu1*, 2) *elu2*, 3) *elu3*, 4) *cdc25-1*, 5) *cdc25-2.1*, 6) *cdc25-2.2*, 7) *cdc25-sep1*, 8) *elu-cdc10-br*, 9) *elu-cdc25-br*.

The elutriation datasets and the *cdc25* block-release datasets consist of 18 to 20 time points covering 2 full cell cycles, the combined elutriation/block-release datasets contain 21 to 22 time points, however covering only one cycle. The elutriation synchronization method produces a very homogenous population of small cells resulting in expression data that contain less noise, while the *cdc25* method is based on temperature-sensitive cell-cycle mutants generating a slightly better synchrony, but introducing some artifacts, which may affect the quality of the expression [3].

In the pre-processing phase the rows containing more than 25% missing entries have been filtered out from each expression matrix and any other missing expression entries have been imputed by the DTWimpute algorithm [12]. In this way nine complete expression matrices have been obtained. Further, the set of overlapping genes has been found across all nine datasets. Subsequently, the time expression profiles of these genes have been extracted from the original data matrices and thus nine new matrices have been constructed.

The implemented version of the hybrid aggregation algorithm combines two individual P -values, generated by the permutation-based methods of regulation and periodicity estimation proposed in [4], by using three different aggregation operators: arithmetic, geometric and harmonic means. Each one of these aggregation operators exhibits certain shortcomings when used individually. For instance, the arithmetic mean values are strongly influenced by the presence of extremely low or extremely high values. This may in some cases lead to an averaged overall P -value, which does not adequately reflect the individual P -values. In case of the geometric mean, the occurrence of a very low individual value is sufficient to produce a low overall P -value, no matter what the individual P -values are. The harmonic mean behaves even more extreme in situations when an individual entry with a very low value is present.

The proposed hybrid computational method has been benchmarked against three other computational methods for the identification of periodically expressed genes: statistical tests for regulation and periodicity and the combined test for regulation and periodicity, all defined by de Lichtenberg *et al.* in [4]. In order to validate and compare the performance of the considered computational methods, they have been applied on the expression matrices included in our test corpus. As benchmark sets we have used 407 genes identified by Rustici *et al.* [3] as cell-cycle regulated and 218 significantly regulated genes identified in [13].

B. Evaluation

The performance of the considered methods in the identification of genes from the benchmark sets is measured as follows: $\text{cov} = n^2 / s \cdot s_b$, where n is the number of overlapping genes across two sets (identified and benchmark), s and s_b are the number of genes in the newly obtained and benchmark sets, respectively. As it was shown in [14], higher values of the fraction (coverage) cov implies better

performance of the underlying computational algorithm on the corresponding test matrix.

The identification of periodically expressed genes may also be viewed as a classification problem for which the objective of each studied method is to distinguish positives (periodically expressed genes) from negatives. A true classification is achieved when a method classifies a gene correctly, otherwise the classification is false. Since a gene may be either a positive or a negative, there are four possible outcomes of gene classification: 1) true positive, 2) false positive, 3) true negative, and 4) false negative. These four outcomes may also be associated with different costs, depending on the application.

Using the classification perspective, the performance of the four studied methods can also be analyzed using ROC curves and evaluated using the area under the ROC curve measure (AUC), which is based on the true positives rate (TPR) and the false positives rate (FPR). Two suitable characteristics of AUC are that it does not depend on equal class distribution or misclassification costs [15]. ROC analysis in general and the calculation of AUC in particular is described in detail and motivated by Fawcett [16]. Using AUC as performance metric, we may compare the methods as follows: the null hypothesis of interest is that the difference in performance between any of the studied methods is zero. The test of this hypothesis involves the comparison of more than two methods and on multiple data sets. Thus, a suitable method to apply when testing the null hypothesis is the non-parametric Friedman's test [17] and the corresponding Nemenyi post hoc test [18]. This combination of tests is comparable to the ANOVA test and the Tukey post hoc test for situations where ANOVA's assumptions may be violated [19]. Friedman's test is based on calculating the average rank of each method on the studied data sets rather than using the actual AUC scores.

Since it has been shown that Friedman's statistic, based on the χ^2_F distribution, is too conservative [20], the recommended alternative statistic, F_F , based on the F -distribution, is used instead:

$$F_F = \frac{(m-1)\chi_F^2}{m(l-1) - \chi_F^2}$$

Given that m represent the number of data sets and l represent the number of compared methods, hypothesis testing is conducted at $p < 0.05$ and with $l-1$ and $(l-1)(m-1)$ degrees of freedom. If the null hypothesis is rejected, the Nemenyi post hoc test can be used to determine whether the performance of two particular methods is significantly different. In order for the difference to be significant, the corresponding average ranks of the methods in question must differ by at least the critical difference:

$$CD = q_\alpha \sqrt{\frac{l(l-1)}{6m}},$$

where q_α is the critical value for the two-tailed Nemenyi test at $p < \alpha$. The corresponding critical value for the Nemenyi post hoc test in our case is 2.569. Thus, the critical difference is approximately 0.951.

C. Software

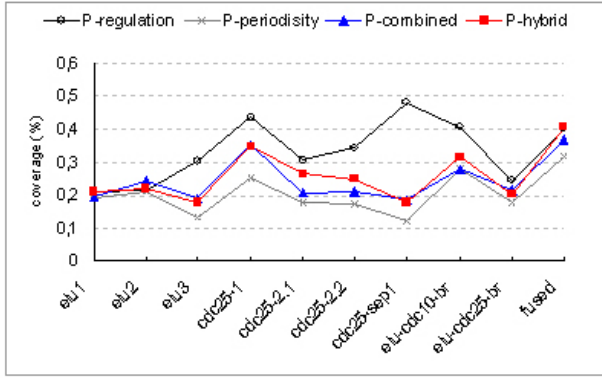
Four computational methods for the identification of periodically expressed genes have been implemented in the developed software: 1) P -value for regulation; 2) P -value for periodicity; 3) combined P -value for regulation and periodicity; 4) hybrid P -value for regulation and periodicity. The implemented statistical tests for regulation and periodicity are based on the permutation-based algorithm described in [4] (see Sections II.A and II.B). The combined P -value for regulation and periodicity is calculated according to the definition proposed by de Lichtenberg *et al.* in [4] (equation (1)). Four text files (one per computational method) containing the calculated P -values are generated as a result of the software execution. In addition, the software proposes a possibility for comparing the performance of the implemented methods in the identification of genes from a given benchmark set by calculating the corresponding coverage values.

IV. RESULTS AND DISCUSSIONS

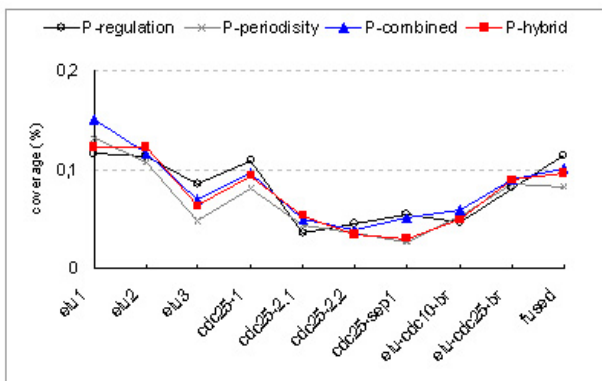
Fig. 2 benchmarks the calculated coverage values of the individual test matrices and the integrated (fused) matrix separately on each benchmark set. The integrated matrix is obtained by applying the hierarchical merge procedure, developed in [13], which fuses together multiple-experiment expression profiles of the individual matrices. It can be seen that the performance of P -regulation test is almost always superior to the one of P -periodicity test. The greatest difference in performance between two methods is seen for *cdc25-sep1* dataset, which may be due to the quality of the data. In addition, the P -regulation test in the majority of cases outperforms the other three methods especially on the benchmark set containing 218 cell-cycle regulated genes identified in [13] (see Fig. 2(a)). A possible explanation for this phenomenon is the used benchmark set, which was built in [13] on the base of overall P -regulation values of multiple-experiment expression profiles from Rustici *et al.* data [3]. The P -hybrid coverage values are comparable to (in some case even higher than) those given by the combined P -value method. Moreover, the performance of the P -hybrid method is almost always better than that of the P -periodicity method. The latter one is observed to have the worst performance.

The above initial information about the performance of the studied computational methods can be applied to introduce weights in the definition of hybrid P -value. In this way, its calculation will be fitted to the observed performance of P -regulation and P -periodicity tests. Namely, the obtained coverage values of each method on the nine test datasets can be used to define weights, which will be assigned to the corresponding P -values and used in the calculation of P -hybrid values. Evidently, the P -regulation and P -periodicity tests will have different contribution to the produced hybrid P -value. In

the considered context, the P -regulation and P -periodicity weights have been estimated to 0.7 and 0.3, respectively.



(a) 218 cell-cycle regulated genes identified in [13].



(b) 407 cell-cycle regulated genes identified in [3].

Figure 2. Comparison of the coverage results of the test matrices.

TABLE I. Area under the ROC curve (AUC) scores for all methods on nine data sets and the corresponding average rank of each method.

Data set	AUC scores			
	P-regulation	P-periodicity	P-combined	P-hybrid
elu1	0,85702	0,66861	0,83245	0,84404
elu2	0,86763	0,67224	0,84947	0,85914
elu3	0,92801	0,65785	0,90086	0,91789
cdc25-1	0,92892	0,82191	0,90886	0,91920
cdc25-2.1	0,94902	0,76596	0,91692	0,93531
cdc25-2.2	0,95187	0,79089	0,91316	0,93164
cdc25-sep1	0,94405	0,69769	0,91935	0,93149
elu-cdc10	0,94267	0,85937	0,92987	0,93548
elu-cdc25	0,92336	0,85170	0,90838	0,91416
Avg. rank	1	4	3	2

The weighted version of hybrid P -value method has been benchmarked against the other three computational methods by

conducting Friedman’s test on the AUC scores of each method on the nine studied data sets (see Table I). In classification, the AUC for a certain classifier is calculated by ordering instances according to the classifier’s assigned probability that the instances are positives. However, in contrast, the P -values generated by the studied methods represent the probability that the instances are negatives. Thus, we may obtain the required type of probability for each instance by calculating $1-P$ -value. The ordered list of instances is used to generate a ROC curve from which the AUC metric can be derived. Statistically, the AUC represents the probability that a randomly selected instance, which has been classified as a positive, is instead a negative. In general, an AUC score close to 1 represents a significant ability to correct classify genes, while a score equal to 0.5 implies that the classification performance of the studied method is equal to that of a random guesser. The AUC values in the presented study have been generated by plotting ROC curves with ROCR in the R-project environment.

The aforementioned results are confirmed by Friedman’s test, as well. The test statistic, F_F , is distributed according to the F -distribution with 3 and 24 degrees of freedom. At $p < 0.05$, the null hypothesis can be rejected since $F_F > 3.72$. The subsequent Nemenyi post-hoc test reveals that P -regulation is significantly better than the other methods. Moreover, P -combined and P -hybrid are significantly better than P -periodicity since the differences in average rank are larger than the critical difference ($CD = 0.951$). Finally, P -hybrid is shown to be significantly better than P -combined in terms of AUC performance.

V. CONCLUSION

In this paper, we have proposed a method for the identification of periodically expressed genes, which is based on a hybrid aggregation of estimations, generated by different computational methods. The proposed hybrid computational method has been benchmarked against three other computational methods for the identification of periodically expressed genes: statistical tests for regulation and periodicity, and a combined test for regulation and periodicity, all three developed in [4]. The four computational methods have been investigated and compared on gene expression time series data coming from a study examining the global cell-cycle control of gene expression in fission yeast (*Schizosaccharomyces pombe*). The four methods have also been subjected to ROC analysis. Area under the ROC curve (AUC) scores were obtained for all methods on nine datasets. Friedman’s test and the Nemenyi post hoc test show that P -regulation is significantly better than the remaining methods. P -hybrid and P -combined both significantly outperformed P -periodicity. However, P -hybrid was additionally shown to be significantly better, in terms of AUC, than P -combined.

ACKNOWLEDGMENT

This research is partially supported by U904-17/2007 Bulgarian National Science Fund grant.

REFERENCES

- [1] A. Oliva *et al.*, “The cell cycle-regulated genes of *Schizosaccharomyces pombe*,” PLOS, vol. 3(7), 2005, pp. 1239-60.
- [2] X. Peng *et al.*, “Identification of cell cycle-regulated genes in fission yeast,” Mol. Biol. Cell, vol. 16, 2005, pp. 1026-1042.
- [3] G. Rustici *et al.*, “Periodic gene expression program of the fission yeast cell cycle,” Nat. Genetics, vol. 36, 2004, pp. 809-17.
- [4] U. de Lichtenberg *et al.*, “Comparison of computational methods for the identification of cell cycle-regulated genes,” Bioinformatics, vol. 21(7), 2004, pp. 1164-1171.
- [5] F. Hermans and E. Tsiporkova, “Merging microarray cell synchronization experiments through curve alignment,” Bioinformatics, vol. 23, 2007, pp. e64-e70.
- [6] M. Menges *et al.*, “Cell cycle-regulated gene expression in *Arabidopsis*,” Journal of Biol. Chem., vol. 277, 2002, pp. 41987-42002.
- [7] M. Menges *et al.*, “Genome wide gene expression in an *Arabidopsis* cell suspension,” Plant Mol. Biol., vol. 53, 2003, pp. 423-442.
- [8] K. Shedden and S. Cooper, “Analysis of cell-cycle-specific gene expression in human cells as determined by microarrays and double-thymidine block synchronization,” PNAS, vol. 99, 2002, pp. 4379-4384.
- [9] T. Golub *et al.* “Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring,” Science, vol. 286, 1999, pp. 531–537.
- [10] V.G. Tusher, R. Tibshirani, and G. Chu, “Significance Analysis of Microarrays Applied to Ionizing Radiation Response,” In Proc. Natl. Acad. Sci., vol. 98, 2001, pp. 5116-5121.
- [11] E. Tsiporkova and V. Boeva, “Nonparametric Recursive Aggregation Process,” Kybernetika, J. of the Czech Society for Cybernetics and Inf. Sciences, vol. 40(1), 2004, pp. 51-70.
- [12] E. Tsiporkova and V. Boeva, “Two-pass imputation algorithm for missing value estimation in gene expression time series,” Journal of Bioinformatics and Computational Biology, vol. 5(5), 2007, pp.1005-1022.
- [13] E. Tsiporkova and V. Boeva, “Fusing Time Series Expression Data through Hybrid Aggregation and Hierarchical Merge,” Bioinformatics, vol. 24(16), 2008, pp. i63-i69.
- [14] V. Boeva and E. Kostadinova. “A Hybrid DTW based Method for Integration Analysis of Time Series Data,” Proc. in ICAIS'09, Austria, 2009, pp.49-54.
- [15] F. Provost, T. Fawcett, R. Kohavi, “The case against accuracy estimation for comparing induction algorithms,” In: 15th international conference on machine learning. Morgan Kaufmann Publishers, San Francisco, 1998, pp 445–453.
- [16] T. Fawcett, “ROC graphs—notes and practical considerations for data mining researchers,” Tech. Rep. HPL-2003-4, Intelligent enterprise technologies laboratories, Palo Alto, 2003.
- [17] M. Friedman, “A comparison of alternative tests of significance for the problem of m rankings,” Annals of Mathematical Statistics, vol. 11, 1940, pp. 86-92.
- [18] P. B. Nemenyi, “Distribution-free multiple comparisons,” PhD thesis, Princeton University, 1963.
- [19] J. Demzar, “Statistical Comparisons of Classifiers over Multiple Data Sets,” Journal of Machine Learning Research, vol. 7, 2006, pp. 1-30.
- [20] R. L. Iman and J. M. Davenport, “Approximations of the critical region of the Friedman statistic,” Communications in Statistics, 1980, pp. 571-595.