Far East Journal of Applied Mathematics
Volume 55, Number 1, 2011, Pages 61-77
Published Online: August 4, 2011
This paper is available online at http://pphmj.com/journals/fjam.htm

TIME-STEP ANALYSIS: AN INFORMATIVE STUDY TO UNDERSTAND THE SLEEP EXPRESSIONS DATA CLUSTERS OF DROSOPHILA

LOH WEI PING

School of Mechanical Engineering
Universiti Sains Malaysia
Nibong Tebal, Seberang Perai Selatan
14300 Penang, Malaysia

Abstract

Temporal biomechanics data analysis is increasingly popular in a wide range of quantitative study concerning sleep-related expressions. Many studies involved clustering the brain sleep expression datasets from organisms. However, little knowledge is discovered about how these clustered data relates between conditional studies, e.g., the external perturbation factors during normal sleep and the active period. Therefore, this study mainly attempts a new technique of knowledge discovery using time-step quantitative analysis. We present data mining approaches using STEM clustering technique to form data clusters. Based on these clusters, cluster patterns at every time-step are observed. This study demonstrates time-step factor for identifying reasonable correlations strengths between sequential clusters generated. Our technique is implemented on the case study of fruit fly species (scientifically called the 'Drosophila') brain sleep expressions recorded under three major conditions; perturbed, unperturbed and active. We present results on conditional cluster groupings and correlation measures among clusters at 2 time point basis. The analysis observes time-step changes and the impacts under perturbed

2010 Mathematics Subject Classification: 68T30.

Keywords and phrases: clustered patterns, clustering, expression, perturbed, sleep, time-step.

This work is supported by the Short Term Grant (304/PMEKANIK/60310030), Universiti Sains Malaysia.

Received May 20, 2011

and unperturbed conditions as time duration prolongs. Our time-step analysis approach demonstrates informative learning from temporal clustered data characteristics, a step ahead of the conventional clustering methods which divide data only into groupings alone.

1. Introduction

Knowledge discovery on the mechanics of living organisms has been explored for years among researchers. Among the natural living mechanisms studied extensively in various disciplines is the sleep mechanics [2, 7, 15]. This study usually involves the understanding of mechanical principles based on how interacting forces or external perturbation affects the normal behaviors of organisms in two aspects; static and dynamic. Static mechanics considers bio-systems at rest (sleep) while the dynamic studies the systems in response to time factor of which usually exhibits strong correlation between successive points. To better understand the biomechanical system, researchers should not only consider the mechanical aspect but also the biotic components and how they interact together as well [11, 13]. Much of the early work on biomechanics knowledge discovery requires multidisciplinary collaborations from biological and engineering mechanics area aided by mathematical studies for quantitative interaction analysis. To mine data, however, clustering algorithms are often designed to allow model groupings according to similar data patterns and study conditions. Different concepts based on the nature, biological knowledge as well as computational intelligence are also integrated [10, 16, 21] such as how living organisms behave and respond to stimuli or perturbations.

In most sleep-related researches, the analyses cover the neural functions, memory consolidation in cellular processes, sleep associated disease patterns, recovery analysis following sleep loss or state of vigilance [4, 17]. Though sleep expression is a common behavior; it has complicated domain to study due to the large gene expression number and diversity of external conditional influences. From the experimental analysis of mental expressions, sleep mechanism shows dynamic phenomenon that interplays between conditional factors and disturbances continuously. This interplay has a direct influence on the mental and physical well-being of the Drosophila and thus exhibits a high correlation in inducing sleep-related disorders. As the dynamic condition suggest, time-step analysis on clustered patterns of brain expression data can be used to answer the sleep-related behaviors in an organism. A better understanding of the sleep pattern requires data clustering study on the bearing of external sleep perturbations on the active periods of Drosophila.

However, to our ultimate understanding, informative clustering to link the temporal external perturbation impacts on sleep mechanical patterns has yet to be fully tapped. In many cases, although volumes of sleep expression data are collected via endless experimental work, impact of external perturbations on active period responses remains uncertain. Though, clustering approaches are practiced to mine these data, little effort has been carried out to turn the clustered data into fruitful knowledge. Therefore, it is necessary to extract informative relations in the abundance of brain sleep expression data and other conditional factors via temporal clustered-pattern analysis to arrive at this state.

Conventional clustering merely summarizes about data and simply group those in order to enable biologists have a quick focus as a whole [23]. Clustering itself fails to account for the fact that the time is a continuous variable. This will cause groupings of unrelated expressions. Hence, clustered data only remains as independent groups without further cracking the intelligence to match the time factor of study problems.

To better understand clustered data patterns, we ought to deal with complex participating factors. For instance in the study case of whether the sleep mechanism is also affected by the time dependent external disturbances, factors involved perturbation impacts as well as time-step influences. From the perspectives of an unsupervised learning, the challenge is to understand the sleep characteristics subjected to sequential clustered data patterns analysis. To ensure that the developed clusters are meaningful, existing biological knowledge can be incorporated to learn from the clustered data. To the best of our knowledge, mining temporal clustered patterns into informative knowledge is a new approach in the data mining area. The important thing here is to provide enough clustered-pattern to aid sequential understanding in the biological studies. Proper clustered data analysis designed will generate knowledge representation of clusters at different time point basis.

This paper addresses the time-step analysis on data clusters implemented using the case study of Drosophila brain sleep expression. In the following sections, the term brain expression or gene expression will be used interchangeably to refer to the species' brain sleep expression values.

2. Brain Sleep Expression Data

This section describes the case study dataset: brain sleep expressions of Drosophila species, obtained from publicly available domain at Gene Expression

OMR (GEO) datasets. This data studies about the sleep deprived effects on the brain of Drosophila Melanogaster species, accumulated from changes throughout prolonged wakefulness and sleeping duration in forms of array-based data.

The raw data consists of short records of temporal sleep expression, collected under 6 experimental states: normal controlled brain sleep (NBS), brain consolidated sleep (BCS), brain sleep deprived (BD), brain sleep (BS), brain normal wake (BW) and brain during active period stimulation (APS). Both the NBS and BCS act as the control data. The expression value changes ranged from normal sleeping hours (control) to deprived sleeping hours due to perturbations followed by normal active period to active stimulated condition. The overall of 14014 raw data are expressed in form of nominal (gene accession numbers) and numerical (quantitative expression values) scales observed at 4 time points (0, 2, 4 and 6 hour basis) (Table 1). The data displayed gene expression values from a "normal" condition (NBS or BCS) that resulted in the response variables (BW or APS) upon "changed" condition.

We denote the condition NBS as 'unperturbed'. Effects of external perturbations on the species during normal sleep period are studied, denoted as 'perturbed'. The main aspect of our analysis is to analyze the stimulated prolong perturbed effects in the brain of Drosophila during their active period stimulated (APS) under similar time durations.

Table 1. Data descriptions: Temporal brain expression values of Drosophila

Species	Scale	Time-step	Time	Description	Condition
Drosophila Melanogaster	Nominal and numerical	4	0 hours	Normal brain Sleep (NBS)	Unperturbed
			2 hours	Brain sleep (BS)	
			4 hours		
			6 hours		
		4	0 hours	Brain consolidated sleep (BCS)	Perturbed
			2 hours	Brain sleep deprived (BD)	
			4 hours		
			6 hours		
		2	4 hours	Active brain (AB)	Active
			4 hours	Active period stimulation (APS)	

Our study data is displayed in the form of multi-digraph in which sleep expressions are represented by nodes at particular time points (Figure 1). Numbers shown on arrows indicate the number of replicate expressions. This representation summarizes between data sample variations and factors that affect sleep expression data at different time steps.

The multi-digraph figure shows a better picture on brain expressions at several recorded conditions. When the relationships between all factors are known, the corresponding sleep mechanisms in a dynamical view may be better understood for further cluster designs [23].

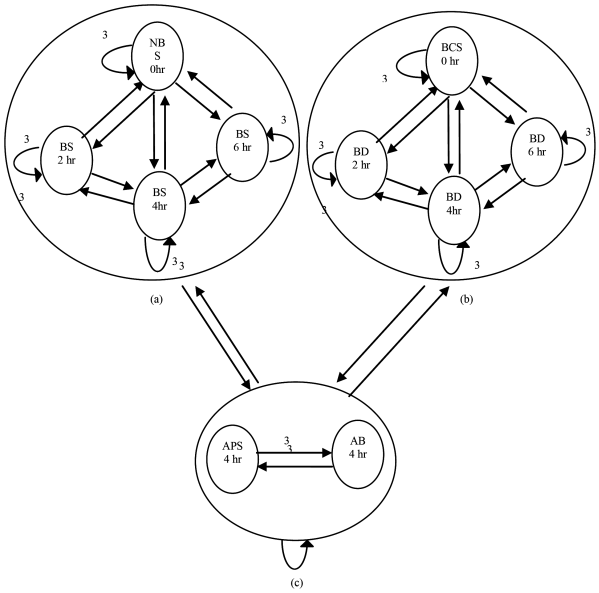

Figure 1. Multi-digraph of Drosophila Melanogaster brain expressions at hour 0, 2, 4 and 6 of (a) unperturbed condition, (b) perturbed condition and (c) active period.

3. Methodology

3.1. Mining time-step expression data

The characteristics of temporal brain sleep mechanisms of Drosophila can be learned using data mining approaches involving data clustering, and clustered-pattern analysis to obtain information. Each of these levels addresses the qualities of reliable data which also serves as a prior step for higher stages in the data mining processes.

Prior to clustering time-series microarray data, data standardization is important. Brain expression data usually consists of non-normal random noise, small number of measurements per gene or non-uniform distributions due to technological problems and biological variations. According to Wang et al. [18], even with the utmost accurate data, if the normality of the data is not assessed, then it turns out with little beneficial value.

Data clustering involves the grouping work of similar data patterns, behaviors and characteristics exhibited under several recorded time points or conditions. As temporal data analysis is rapidly growing in the realm of unsupervised learning research, knowledge representation techniques to deal with uncertain knowledge are required. These techniques potentially yield fruitful medical and biological insights for improving biomechanical system [10, 16, 21]. Due to the unavailability of trained time-step data, clustering effort gains the importance [1]. Matter arises from the clustering analysis that the clusters generated might not agree well with the actual biological meaning of data [3]. On the other hand, there are high tendencies of gene expressions being misgrouped according to similar characteristics and result in mixture of expression values collected at various time durations, leading to inaccurate correlations and interpretations.

Here, we use clustering approach to extract fundamental understanding of sleep mechanics subjected to perturbations at different time steps. To turn accumulated systematic effects into a standardized scale, data transformation process is formerly conducted. Significances of sequential time-step analysis to characterize the effects of perturbations performed on brain sleep mechanics of Drosophila are also considered. Like most sleep-related researches, the fruit fly species, Drosophila is chosen since this species potentially exhibits characteristics that mimic mammalian sleep responses [12]. To retrieve knowledge from clusters, statistical analysis is usually required to find the significant relations. Historical understanding on the study data, on the other hand, is a good solution to retrieve information about underlying relationships like time order between variables. However, in sleep mechanics, the main problem is that the brain data are usually recorded in an unsupervised fashion (without knowledge) about the characteristics or responses of data. For the purpose of generating cluster profiles, we adopted the STEM clustering technique facilitated by Short Time-Series Expression Miner (STEM) tool [5]. The declared goal is pursued further by searching for sequential comparisons of clustered-pattern analysis based on idea of time-step data correlations.

This study includes the impacts of perturbations on normal sleep mechanisms of Drosophila to discover temporal patterns of brain expresses under conditions of normal brain sleep (NBS) and brain deprived (BD) due to external perturbations. Our approach considers the significances of sequential time point basis analysis to characterize information concealed in short time-series microarray clustering analysis. The concept of time-step analysis on sleep mechanics study moves us beyond intuition, to identify continuous changes in the sleep mechanisms.

In the subsequent sections, STEM clustering technique and time-step clustered-pattern analysis are detailed.

3.2. STEM clustering approach

Cluster analysis aims to group data according to similar properties or mechanism at work in the domain from which data points are drawn; a mechanism that causes some instances to be strongly resembles one another than the remaining Keogh [24]. Das et al. [25], however, define clustering from mathematical perspective in which assumptions are made where by $X \in R^{m \times n}$ a set of data items representing a set of *m* points x_i in R^n. The attempt of clustering is to partition *X* into *K* groups such that every data which belong to same group are more alike than other data in different groups. Each of the *K* group is called a *cluster*.

In sleep expressions, data can be studied in terms of expression profiled values' comparisons by type, replications or conditions. Nevertheless, the most fundamental idea of clustering similar data is can also be carried out using the similarity measure called the *distance metric*. As our case study consists of volumes of numerical brain sleep expressions, these data can be clustered into *n*-dimensional vectors. For instance two clusters of data points in form of $x = (x_1, x_2, x_3, ..., x_n)$ and $y = (y_1, y_2, y_3, ..., y_n)$ as *n*-dimensional vectors.

As the sleep expressions also involved time-steps, the impact of the time parameter towards our cluster analysis is important. Time-series generally refers to sequences of data point measurements at different times that can be observed in terms of changes across time steps. In clustering expression values, the set of time parameter is either clustered or ignored to obtain typical patterns in data. At the same time, there are indications that sequential time point's data may actually characterize relations and data dependencies. Thus, there is a division of which expressions clustered into profile groups could be further analyzed in finer aspects.

At the preliminary stage, the raw data is standardized by log normalizing uncontrolled expression values, v_i at time-step, $i = 1, 2, 3, ...$ with respect to the controlled, $v_0 : \left(0, \log_2\left(\frac{v_1}{v_0}\right), \log_2\left(\frac{v_2}{v_0}\right), ..., \log_2\left(\frac{v_n}{v_0}\right)\right)$. Log normalization considers systematic differences across data and to provide numerals meaningful from both biological and mathematical aspects. Besides, the noise, outliers or extreme values identified at this stage can be verified and eliminated. Thus log normalization and transformation make the data distribution more symmetrical and almost normal [9], [14], [19]. Expression values differences based on maximum-minimum value records of every single genotype expressions is key to the significances of expression replicates. This variation enables outliers to be identified and individual expression to be distinguished from one experimental condition to another. The $\log_2$ transformation is chosen here because STEM tool is programmed with the base 2 logarithm. However, it is not major whether the data are transformed into $\log_2$, $\log_e$ or $\log_{10}$ as long as the log transformations are carried out consistently [22]. The standardized data are later subjected to clustering analysis whereby STEM clustering method is employed. The STEM algorithm is adopted since it is specifically programmed for short records of time-series expressions. The main parameters to consider using this technique include the correlation, maximum number of profiles, number of permutations and the significance level.

In our study case, we made comparisons among several replicates of expression data at every time point. Based on STEM clustering technique, data profiles which are assigned to a particular cluster ought to show correlation values above a minimum correlation benchmark. To set this boundary, the major emphasis is either on the minimum correlation or the minimum correlation percentile. For sleep expressions be grouped into profiles according to same responses, however, two model profiles ought to have the correlation higher than minimum correlation factor. The maximum correlation plays the indicator for selecting the sleep expressions to avoid the phenomenon of identical profiles opted. Number of permutations, whereas, ensures the appropriate permutations be based on the time point aspects prior to cluster profiles assignment.

Data clustering using correlation factor, however, tolerates with the common similarity measure in the sense that, set of expressions that characterize similarly are likely to have high correlations to fall into identical cluster profiles. Pearson

correlation coefficient is thus chosen to rank linear relationships between pairs of dimensions as shown in equation (1):

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}, \quad (1)$$

where

n = number of points

r = correlation coefficient (between –1 and 1).

The basic objective of expression clustering at this point is to extract the fundamental patterns of the brain sleep expressions. The next challenge is to turn the clustered data profiles into information. In this article, we deliver the novel idea of time-step analysis that links sequential clustered changes impacts on expression values. This is to extend the informative knowledge discovery from clustered data patterns into finer time-basis aspects. The patterns retrieved ought to conserve the dynamical aspects from clustered data patterns obtained. The time-step analysis concept for informative pattern analysis is further discussed in the following sections.

3.3. Clustered patterns time-step analysis

When dealing with temporal datasets for instance $x_t = (x_1, x_2, x_3, ..., x_n)$ at the time interval $1 \le t \le n$, the internal structure of data such as autocorrelation, trend or seasonal variation is always monitored. Similar datasets $x_i, 1 \le i \le n$ recorded at time interval $1 \le t \le m$, when clustered into

$$x_{it} = ((x_{11}, x_{21}, x_{31}, ..., x_{n1}), (x_{12}, x_{22}, ..., x_{n2}), ..., (x_{1m}, x_{2m}, ..., x_{mm})),$$

meaning that each bracket represents similar pattern groupings at particular time steps. In each grouping, pattern variations can be judged. The clustered patterns analyzed usually unable to show sequential relations from a time-step to another. In short, the sequence of the clustered data is rather important. From the initial point of data clustering approach, the importance is to group data according to their similar characteristics (at single time-step). Clustered data patterns, however, posit a trend which is then suppose to be correlated to other grouped trends at different time-steps.

Clustered patterns in temporal data are usually not considered in biomechanics studies. In particular, the difference among clustered profiles at every time-step has yet to be properly tapped. For instance, if there exists more than a parameter in a case study, then the independence or relations associated to time and external parameters need deeper exploration.

The clustered-pattern can be understood in terms of time duration as a whole. Yin and Chiang [20] initiate to study the trend of expressions in terms of data variation over time measured. This technique studies the grouping profiles and patterns exhibited from expressions. The profile patterns are compared from the initial state till the last time-step. However, the information retrieved from overall time duration judgement of clustered patterns might turn out meaningless if the patterns fail to indicate changes of expressions in sequence (from one time point to another). Therefore, our response here is to link the cluster patterns of brain expressions through sequential time-step analysis.

Pattern analysis can also lead to interpretations such as in our case study; whether perturbations have impact on the active period of an organism. Apparently, informative pattern discovery through time-step analysis plays the key role to extract such message. A simpler manner to consider the time-step clustered pattern is via our proposed idea: time-step analysis on clustered data correlations. These also include interactions and dependencies between brain expressions (1) in cluster, (2) between clusters and (3) across time steps. The key idea here is to make good use of the correlation measure to link binary cluster profiles at sequential time-step basis.

4. Results and Discussions

The STEM clustering technique generates clustered data profiles including the number of sleep expressions assigned to each and every cluster (Table 2). Table 2 shows the cluster profiles formed under 3 main conditions: time-steps of which brain expression data was collected, profile numbers of clustered data, number of brain expressions clustered into particular profiles as well as either the groupings are significant or insignificant. The results from Table 2, show that the profiles numbers 46-49 are significant since these profiles appear to be significant in all experimental conditions: unperturbed, perturbed, active. Meaning that in each time-step, these profiles show identical graphical patterns for matching. For instance at hour 0 of unperturbed, both profile numbers 48 and 46 show similar characteristics while

profile numbers 47 and 49 illustrate another pattern. Meanwhile profile number 45 which unable to match these two characteristics; thus considered as insignificant. For all time points, there is a single cluster with has insignificant numbers of expressions assigned.

Based on Table 2, we consider clustered data profiles and the corresponding experimental conditions. The comparisons of independent clustered profiles are carried out according to 4 experimental states; unperturbed vs. unperturbed, perturbed vs. perturbed, unperturbed vs. perturbed and wake vs. stimulated. These states were also compared at every binary time-step basis; hours 0-2, hours 2-4, hours 4-6.

The permutations on brain expressions allowed *p*-value enumerations. The significance parameter could show the number of data assigned to profiles compared to expected number of brain expressions assigned for *p*-value computation. This evaluation enables us to differentiate between the actual (number of brain sleep expressions) correctly clustered and expected size (number of expressions expected to be assigned based on permutation tests). STEM clustering method puts the emphasis on minimum correlation and its percentile value. This means that in a particular cluster, two model profiles should indicate higher correlation than the minimum correlation parameter to get an actual grouping. Nevertheless, the overall judgement of clustered profiles formed still fail to indicate changes of expressions from a time point to another. We thus initiate to work out the relations between patterns of brain sleep expressions on time-step basis.

Our hypothesis from the profiles results is that the time parameter and the experimental conditions are related to each other. Thus, the clustered patterns observed from Table 2 are evaluated from both quantitative and qualitative standpoints to link clustered-pattern profiles via time-step for comparisons. For example, Figure 2 shows a comparison set of unperturbed condition at hours 0-2 in Drosophila Melanogaster brain expressions. The figure illustrates boxes that represent different temporal expression profiles. Numbers indicated at the top left corner show the profile numbers while the bottom left corner shows number of expressions assigned to that particular profile.

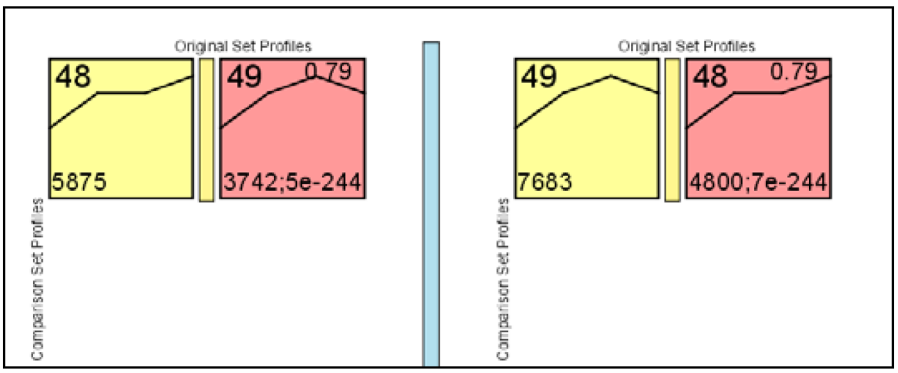

Figure 2. Comparison profiles of Drosophila Melanogaster brain expressions at hours 0 vs. 2 of unperturbed condition. Cluster profiles to the left of yellow bar are clustered at hour 0 whilst to the right represents cluster profile at hour 2. The top right figure to the right of yellow bar, indicates the correlation value between compared profiles. Meanwhile the bottom left corner value indicates number of genes assigned followed by the *p*-value of genes in the compared intersection.

However expressions in clusters representing high similarities may not necessarily exhibit similar patterns at different time steps. Often, researchers explore clusters in terms of strong correlation of data to reflect consistencies or similarities in a cluster itself. These patterns will neither relate the temporal effect nor the study conditions as they only accumulate data according to similar measures. To better link the generated clusters, we find information through the cluster profiles developed. Evaluations in terms of different study states, for instance correlations based on brain expression types, perturbation effects and the time-step conditions are important. This requires not only problem understanding, but also techniques of discovering time-step analysis on clustered patterns which is a new step ahead of the conventional data clustering analysis.

The overall time-step clustered-pattern profile for the expression data is summarized in Table 3. This table shows the results comparisons in terms of correlations. A potential difference observed here is the level of correlations: strong or perfect. In the conditions of unperturbed vs. unperturbed and perturbed vs. perturbed, correlation values evaluated at 2 time point basis show an increment from approximately 0.8 (strong) to 1.0 (perfect). In the case of unperturbed vs. unperturbed, a drop of correlation from 1.0 (perfect) to 0.79 (strong) is observed.

In terms of relations throughout all time steps and conditions, at subsequential time-step basis, we deal with binary relations. We name sets *A*, *B*, *C* and *D* to represent clusters of brain expressions for the 4 time steps. From Table 3, a dotted line across the diagonal of the table is drawn to represent the symmetrical axis. Here, we can see that relations, *R* at all conditions and time-steps can be explained mathematically according to the following notations:

i. Reflexive if $aRa, \forall a \in A$. For instance: unperturbed (hr 2) ***R*** unperturbed (hr 2), perturbed (hr 4) ***R*** perturbed (hr 4).

ii. Symmetric if aRb, then $bRa, \forall a \in A, b \in B$. For instance: unperturbed (hr 2) ***R*** unperturbed (hr 4), then unperturbed (hr 4) ***R*** unperturbed (hr 2).

iii. Transitive if aRb, bRc, then $aRc, \forall a \in A, b \in B$ and $c \in C$. For instance: unperturbed (hr 0) ***R*** unperturbed (hr 2), unperturbed (hr 2) ***R*** unperturbed (hr 4), then unperturbed (hr 0) ***R*** unperturbed (hr 4).

The increase of correlation values in the first two compared conditions explains that as the sleeping time duration of Drosophila prolongs, the species gains sufficient time to express itself till perfect expression values are observed. In the case of unperturbed vs. perturbed sleeping durations, the decrease correlation suggests that disturbances during sleeping hours have mainly affected the brain expressions of the species during their active period. Therefore, the species performs differently compared to its original habits.

Table 2. Cluster profiles of brain expressions different conditions and time-steps

Conditions	Hour	Profile no.	Significant	No. of expressions
Unperturbed	0	48,46	Yes	7169
		47,49	Yes	6839
		45	No	2
	2	48,49	Yes	14002
		40,45,47	No	12
	4	48,49	Yes	14005
		46,47	No	5
	6	48,49	Yes	13982
		45,46	Yes	21
		47	No	0
Perturbed	0	48,49	Yes	13989
		45	Yes	8
		46,47	No	13
	2	48,49	Yes	13775
		46	Yes	223
		40,45,47	No	12
	4	48,49	Yes	13977
		45,46	Yes	28
		47	No	5
	6	48,49	Yes	13977
		47	Yes	28
		45,46	No	5
Active	Wake	48,49	Yes	12869
		45,47	Yes	29
		46	No	6
	Stimulated	46,48	Yes	7309
		47,49	Yes	6701

Table 3. Number of intersections and average correlation values of clustered patterns in Drosophila brain expressions

		Number of cluster profile intersections (correlation value)									
		Unperturbed time (hr)				*Perturbed time (hr)*				*Wake(hr)*	APS*(hr)*
Conditions (hr)		0	2	4	6	0	2	4	6	4	4
Unperturbed	0		2 (0.79)	2 (0.79)	3 (0.76)	2 (1.00)				1 (0.79)	2 (1.00)
	2	2 (0.79)		2 (1.00)	2 (1.00)		3 (0.97)				
	4	2 (0.79)	2 (1.00)		2 (1.00)			2 (0.79)		2 (1.00)	
	6	3 (0.76)	2 (1.00)	2 (1.00)					2 (0.79)		
Perturbed	0	2 (1.00)				3 (0.75)	2 (1.00)	2 (1.00)	1 (0.79)	3 (0.95)	
	2		3 (0.97)			3 (0.75)		3 (0.73)	3 (0.73)		
	4			2 (0.79)		2 (1.00)	3 (0.73)		2 (1.00)		2 (1.00)
	6				2 (0.79)	2 (1.00)	3 (0.73)	2 (1.00)			
Active	wake 4	1 (0.79)		2 (1.00)		1 (0.79)					2 (0.79)
	APS 4	2 (1.00)				3 (0.95)		2 (1.00)		2 (0.79)	

5. Conclusion

The time-step analysis technique incorporates temporal data information and correlation analyses for identifying and describing the brain sleep changes. This method can consider clustered patterns of sleep mechanics system under conditional (unperturbed, unperturbed and active) environments. Time-step analysis performs well to extract information according to sequential time steps (one time point after another) as in many cases clustered results at a preliminary time point potentially affect the following outcomes. This can be described relations between sequential clusters for observing the time-step changes and external impacts in a system. As in many applications of biomechanics, fundamental mathematical relations are important. Here, our work has demonstrated the idea of time-step analysis which can be further adopted to design new machine learning devices for more efficient temporal data clustering.

References

[1] L. Bréhélin, O. Gascuel and O. Martin, Using repeated measurements to validate hierarchical gene clusters, Bioinf. 24 (2008), 682-688.

[2] C. Cirelli, Searching for sleep mutants of Drosophila Melanogaster, Bioessays 25 (2003), 940-949.

[3] A. Clare and R. D. King, How well do we understand the clusters found in microarray data?, Silico Biol. 2 (2002), 511-522.

[4] J. M. Donlea, N. Ramanan and P. J. Shaw, Use-dependent plasticity in clock neurons regulates sleep need in Drosophila, Sci. 324 (2009), 105-108.

[5] J. Ernst, G. J. Nau and Z. Bar-Joseph, Clustering short time-series expression data, Bioinf. 21 (2005), i159-i168.

[6] GEO (Gene Expression OMR Datasets), http://www.ncbi.nlm.ni.gov.geo. Accessed 2 May 2007.

[7] S. T. Harbison, T. F. C. Mackay and R. R. H. Anholt, Understanding the neurogenetics of sleep: progress from Drosophila, Cell Press 25 (2009), 262-269.

[8] A. Joshi, Y. Van de Peer and T. Michoel, Analysis of a Gibbs sampler method for model-based clustering of gene expression data, Bioinf. 24 (2008), 176-183.

[9] A. Long, H. Mangalam, B. Chan, L. Tolleri, G. W. Hatfield and P. Baldi, Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework, J. Biol. Chem. 276 (2001), 19937-19944.

[10] I. Rojas, M. Golebiewski, R. Kania, O. Krebs, S. Mir, A. Weidemann and U. Wittig, SABIO-RK: A database for biochemical reactions and their kinetics, BMC Systems Biology 1 (2007), S6.

[11] S. D. Ross, M. L. Tanaka and C. Senatore, Detecting dynamical boundaries from kinematic data in biomechanics, Chaos 20 (2010), doi:10.1063/1.3267043.

[12] G. M. Rubin, Around the genomes: the Drosophila genome project, Gen. Res. 6 (1996), 71-79.

[13] P. D. Schreuders and A. Johnson, A systems approach for bioengineering, Int. J. Eng. Education 15 (1999), 243-248.

[14] T. P. Speed, Hints and Prejudices-always Log Spot Intensities and Ratios. Univ. of California, Berkeley, 2000.

[15] R. Stickgold, J. A. Hobson, R. Fosse and M. Fosse, Sleep, learning, dreams: off-line memory reprocessing, Sci. 294 (2001), 1052-1057.

[16] M. Strauch, J. Supper, C. Spieth, D. Wanke, J. Kilian, K. Harter and A. Zell, A two-step clustering for 3-D gene expression data reveals the main features of the Arabidopsis stress response, J. Integrative Bioinformatics 4 (2007), 54-66.

[17] S. Sunderam, N. Chernyy, N. Peixoto, S. L. Weinstein, S. J. Schiff and B. J. Gluckman, Improved sleep-wake and behavior discrimination using MEMS accelerometers, J. Neurosci. Methods 162(2) (2007), 373-383.

[18] R. Y. Wang, H. B. Kon and S. E. Madnick, Data quality requirements analysis and modelling, Proc. of the 9th Int. Conf. of Data Eng. (ICDE), 1993, pp. 670-677.

[19] Y. H. Yang, M. J. Buckley, S. Dudoit and T. P. Speed, Comparison of methods for image analysis on cDNA, Technical Report, Univ. of California, Berkeley, 2000.

[20] Z. X. Yin and J. H. Chiang, Novel algorithm for coexpression detection in time-varying microarray data sets, IEEE Trans. Comput. Biol. Bioinf. 5 (2008), 120-135.

[21] H. Zhu, Y. Tang, L. Ivanciu, M. Centola, C. Lupu, F. B. Taylor and F. Lupu, Temporal dynamics of gene expression in the lung in a baboon model of E. coli sepsis, BMC Gen. 8 (2007), 58.

[22] S. Knudsen, A Biologist's Guide to Analysis of DNA Microarray Data, John Wiley and Sons, USA, 2002.

[23] W. P. Loh and Y. Abu Hasan, Clustering sleep deprivation effects on the brain of Drosophila Melanogaster, The 2nd Regional Conf. on Ecological and Environmental Modeling (ECOMOD07), 2007.

[24] E. Keogh, Exact indexing of dynamic time warping, Proceedings of the 28th International Conference on very large data bases, 2002, pp. 406-417.

[25] G. Das, K. Lin, H. Mannila, G. Renganathan and P. Smyth, Rule discovery from time series, Proceedings of the 4th International Conference on knowledge discovery and data mining, 1998, pp. 16-22.