2J2-OS-16a-5

Sleep Pattern Visualization via Clustering on Sound Data

Hongle Wu*1 Takafumi Kato*2 Tomomi Yamada*2 Masayuki Numao*1
Ken-ichi Fukui*1

*1 The Institute of Scientific and Industrial Research, Osaka University
*2 Graduate School of Dentistry, Osaka University

The quality of a good sleep is important for a healthy life. Recently, several sleep analysis products have emerged on the market; however, many of them require additional hardware or there is a lack of scientific evidence regarding their clinical efficacy. We proposed a novel method via clustering of sound events for discovering the sleep pattern. This method extended conventional self-organizing map algorithm by kernelized and sequence-based technologies, obtained a fine-grained map that depicts the distribution and changes of sleep-related events. We introduced widely applied features in sound processing and popular kernel functions to our method, evaluated their performance, and made a comparison. Our method requires few additional hardware, and by visualizing the transition of cluster dynamics, the correlation between sleep-related sound events and sleep stages was revealed.

1. Introduction

Sleep is an important physiological state of the human body. Almost one third of the time in a person's life is spent sleeping. The quality of sleep is very important to a person's health. Therefore, sleep monitoring technology has become an indispensable content in modern personal sleep management [Chen 13].

Currently, there are many products on the market that aim to make sleep assessment portable at a reduced cost. Besides traditional polysomnography (PSG), actigraphy has also been used as an alternative tool. One of the problems of these products is that they are invasive to users, which means that users have to wear an additional device or place a device on their bed during sleep. According to a recent survey, many people are resistant to wearing a device during sleep [Choe 10]. Even if users accept to wear the device, it is not easy to properly place the sensors in the correct position. Also, according to [Mantua 16], medical experts do not suggest to use the results from these consumer equipment for medical research, which means they are not reliable enough.

Moreover, additional devices add extra financial burden to the user. The efforts in the market to reduce the cost are mostly through mobile apps. Mobile apps use a smartphone's built-in sensors, and hence, users do not need to purchase additional hardware. However, according to [Behar 13], very few of the apps are based on published scientific evidence.

To solve the problems mentioned above simultaneously, and considering that many types of sleep disorder are respectively related to a distinctive type of sound, such as snoring, tooth grinding, limb movement and sleep talking, we proposed a method for sleep analysis based on clustering of sound data.

Contact: Hongle Wu, The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka, 567-0047, Japan, +81-6-6879-8426, wu@ai.sanken.osaka-u.ac.jp

We extracted sound clips of events from the recorded sound data, applied Fast Fourier Transform (FFT) to get the frequency spectrum as input vectors, and then applied self-organizing map (SOM) [Kohonen 95] algorithm to the data to obtain cluster maps. In our previous work [Wu 16], we calculated the Euclidean distance between frequency spectrum as the only similarity measure between sound events in standard SOM, in this paper, in order to make a comparison, we applied Mel Frequency Cepstral Coefficent (MFCC) [Davis 80] which is a feature widely used in automatic speech recognition as another metric. Besides the standard SOM, kernel SOM [Fukui 11] was also used. Since the Euclidean distance applied in the standard SOM treats each discrete point as an independent variable, in [Wu 16], Kullback-Leibler (KL) kernel was introduced through kernel SOM as a similarity measure in order to capture the distribution structure of a frequency spectrum. In this paper, to make a comparison, we tried radial basis function (RBF) kernel and polynomial kernel besides KL kernel. According to experiment results, the KL kernel SOM (KL-KSOM) obtained the best effect.

2. Methodology

In this section, we introduce the key methodologies applied in this study.

2.1 MFCC

In sound processing, the mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency.

MFCCs are coefficients that collectively make up an MFC [Davis 80]. They are derived from a type of cepstral representation of the audio clip (a nonlinear "spectrum-of-a-spectrum"). The difference between the cepstrum and the mel-frequency cepstrum is that in the MFC, the frequency bands are equally spaced on the mel scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal cepstrum. This frequency warping can allow for better representation of sound, for example, in audio compression. In our experiment, 12 MFCC coefficients were extracted for each sound clip as a MFCC vector, and the Euclidean distance between the MFCC vectors were applied as the similarity between sound events.

2.2 Kernel SOM

We used the frequency spectrum as input vector. The standard SOM uses Euclidean distance as a similarity measure of data points, so the distribution structure of a frequency spectrum cannot be captured since each discrete point is treated as an independent variable. The authors in [Fukui 11], proposed the use of Kullback-Leibler (KL) divergence to introduce a distribution structure into a similarity measure of frequency spectrum of acoustic emission events and obtained a good effect. In this study, KL kernel, RBF kernel and polynomial kernel were introduced to SOM through kernel SOM[Andras 02] [Boulet 08] to cluster the sleep-related sound events.

The RBF kernel function is defined as

$$K_{RBF}(\mathbf{x}_i, \mathbf{x}_j) = \exp\big(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\big), \tag{1}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are vectors in the input space, and σ is a free parameter. For degree-d polynomials, the polynomial kernel function is defined as:

$$K_{PL}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j + 1)^d, \tag{2}$$

The KL kernel function is defined as:

$$K_{KL}(\mathbf{x}_i, \mathbf{x}_j) = \exp\big(-\beta JS(\mathbf{x}_i, \mathbf{x}_j)\big), \tag{3}$$

$$\begin{aligned} JS(\mathbf{x}_i, \mathbf{x}_j) &= KL(\mathbf{x}_i, \mathbf{x}_j) + KL(\mathbf{x}_j, \mathbf{x}_i) \\ &= \sum_{k=1}^{v}\big\{x_{i,k}\log\frac{x_{i,k}}{x_{j,k}} + x_{j,k}\log\frac{x_{j,k}}{x_{i,k}}\big\}, \end{aligned} \tag{4}$$

where $KL(\mathbf{x}_i, \mathbf{x}_j)$ is the KL divergence, which is the distance between probability distributions, $JS(\mathbf{x}_i, \mathbf{x}_j)$ denotes the Jensen-Shannon divergence, which symmetrizes the KL divergence, and $\beta > 0$ is a scaling parameter.

The basic concept of the kernel SOM is the same as that of the SOM. However, in the kernel SOM, the reference vector is updated in an indirect manner because the reference vector in the mapped space cannot be calculated.

By replacing $\mathbf{x}$ in the updating formula of a reference vector in the standard batch type SOM by a mapped $\phi(\mathbf{x})$, the following updating formula can be obtained:

$$\mathbf{m}_i(t+1) := \gamma\sum_n h_{c(\mathbf{x}_n),i}\phi(\mathbf{x}_n), \tag{5}$$

where t is an iteration step, and γ is a regularization term $\gamma = 1/\sum_n h_{c(\mathbf{x}_n),i}$. However, since $\phi(\mathbf{x}_n)$ cannot be calculated, the i^{th} reference vector is updated using the dissimilarity to all data points $\forall n\ d_{i,n}$, as follows:

$$\begin{aligned} d_{i,n}(t+1) &\equiv ||\phi(\mathbf{x}_n) - \mathbf{m}_i(t+1)||^2 \\ &= K(\mathbf{x}_n, \mathbf{x}_n) - 2\gamma\sum_j h_{c(\mathbf{x}_j),i}K(\mathbf{x}_n, \mathbf{x}_j) \\ &\quad + \gamma^2\sum_k\sum_l h_{c(\mathbf{x}_k),i}h_{c(\mathbf{x}_l),i}K(\mathbf{x}_k, \mathbf{x}_l). \end{aligned} \tag{6}$$

2.3 Proposed method: Sequenced-based kernel SOM

In order to clearly and easily understand the analysis report of a user's sleep, a fine-grained map that depicts the distribution and changing of sleep-related events is necessary. Different from the normal SOM that deals with static data, SbSOM introduces SWF into SOM and can visualize the transition of cluster dynamics since the spatio-temporal neighborhood is converted into the topological neighborhood by the neighborhood function.

In SbSOM, the position of M neurons in the visualization layer be $r_j = (\xi_j, \eta_j), (j = 1, \cdots, M)$, where ξ-direction indicates the temporal dimension. The n^{th} input data are located at the ratio of n/N within the input data sequence, and the j^{th} neuron is located at the ratio of ξ_j/ξ_M on the ξ-direction of cluster map. Let the absolute value of those differences be $\epsilon = |\xi_j/\xi_M - n/N|$. The SWF $\psi(n, \xi_j)$ is defined so as to be able to balance the spatio/temporal resolution; in case where reversal of data order is not allowed, the SWF is given as:

$$\psi(n, \xi_j) = \begin{cases} 1 & if \quad \epsilon < \frac{1}{2K} \\ \infty & otherwise \end{cases}, \tag{7}$$

where K is the number of neurons on ξ-direction. The winner neuron of the input data x_n is determined by spatial distance combined with SWF as follows:

$$c(\mathbf{x}_n) = arg\min_j \psi(n, \xi_j)\|\mathbf{x}_n - \mathbf{m}_j\|. \tag{8}$$

In this study, the comparison of the clustering results between standard and kernel SOM demonstrated that KL divergence as kernel function exhibits better performance. Based on this premises, a novel algorithm, Sb-KSOM, is proposed, which is an extension of SbSOM. The proposed Sb-KSOM kernelized the SbSOM by replacing the Euclidean distance with KL divergence to enable it to handle the frequency spectrum data. In the proposed Sb-KSOM, we replaced the normal Euclidean distance calculation in Eq. (8) with the KL kernel function.

3. Experiment

The data we used in the paper is prepared by Graduate School of Dentistry in Osaka University. The study protocol was approved by the Clinical Research Ethics Committee of the Osaka University Graduate School of Dentistry. Written informed consent was obtained from all subjects.

We first applied the standard SOM and three kinds of kernel SOM to the extracted sound data, and compared the wPF. Then we used Sb-KSOM on the data to obtain the spatio-temporal dimensional cluster map, and discussed the relation between the transition of sleep stages and cluster dynamics of sound events. All of the experimental subjects are university students from Osaka University, and hence, their age was mostly around 20-24. The male to female ratio was balanced.

Table 1: Comparison of wPF between standard SOM and kernel SOM clustering results

Subject id	$SOM_{spectrum}$		SOM_{MFCC}		$KSOM_{KL}$		$KSOM_{RBF}$		$KSOM_{PL}$	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	0.537	0.033	0.509	0.041	**0.604**	0.037	0.593	0.051	0.504	0.047
2	0.521	0.041	0.497	0.043	0.573	0.038	**0.577**	0.038	0.482	0.025
3	0.506	0.031	0.493	0.035	**0.551**	0.031	0.532	0.033	0.535	0.042
4	0.559	0.040	0.482	0.032	**0.592**	0.037	0.561	0.035	0.567	0.026
5	0.602	0.039	0.594	0.039	**0.629**	0.039	0.624	0.041	0.608	0.054
6	0.543	0.033	0.549	0.045	**0.600**	0.035	0.562	0.038	0.557	0.037
7	0.483	0.042	0.501	0.035	0.523	0.047	0.531	0.051	**0.537**	0.036
Mean	0.535	0.037	0.518	0.039	**0.581**	0.037	0.568	0.041	0.541	0.038

3.1 Event extraction

We selected seven nights of sound data. Based on the burst extraction method, we obtained a total of 6775 sound events, which included sleep disorder and other sound events such as outdoor traffic noise. FFT was applied to the extracted sound data to obtain the frequency power spectrum. From 24 Hz to 20 kHz, at intervals of 4 Hz, 4995 discretized points as an input for SOM were obtained for every sound data.

3.2 Quantitative comparison between standard and kernelized clustering

In the first part of this experiment, we used the sound data from each subject as a respective dataset and compared the wPF values for each subject between standard and kernelized algorithms, including standard SOM based on frequency spectrum or MFCC similarity, kernel SOM with KL, RBF or polynomial kernel. In order to avoid initial value dependency, the experiments were executed 50 times and the average values were computed. The hyper parameter of the kernel functions were tuned by a linear search. The mean wPF values and standard deviation are shown in Table 1. The average of wPF shows that MFCC feature does not performs well on this kind of sound data, and the KL kernel SOM has the best performance, which improved by about 10% from standard SOM.

3.3 Sleep pattern analysis

In this experiment, we made a comparative analysis between cluster maps generated by Sb-KSOM and sleep stage sequences to reveal the relation between them. We analyzed all the subject respectively. One of the clustering results is shown in this section. Subject 2 was chosen since the tooth grinding or snoring activities are frequent, and generated more related sound events than the others. Fig. 1 upper part shows the result when Sb-KSOM was applied to the sound data from Subject 2; the number of neurons was set to 50×10 with a two-dimensional grid. Subjects' sleep stages were scored by a medical specialist based on PSG data from the same night, with a time window size of 30s. The sleep stage sequence of Subject 2 is shown in lower part of Fig. 1, where REM stage is shown as "R", awake stage is shown as "W". We defined the period that contain continuous N3 stages with intervals of other stages that less than 3 min as a deep sleep period, and periods except deep sleep periods ,awakening stages and REM stages as light sleep periods. Since the REM stage is a unique phase in the sleep process, we will discuss it separately.

The sleep periods of Subject 2 were interpreted as follows:

Deep sleep periods (0:13:30 - 1:01:30), (1:39:30 - 1:52:30), (2:00:30 - 02:11:00), (2:20:30 - 02:51:00), (4:04:30 - 4:20:30), (6:00:30 - 6:18:30): There were many snoring events during these periods, few body movements, and tooth grindings. We found out that a cluster center of snoring event is usually associated with a deep sleep period.

REM stages (2:53:00 - 3:05:00), (4:42:00 - 5:39:30), (6:56:00 - 7:29:30): Compared with other stages, REM stages have a stronger association with clusters of body movement and a weaker association with those of snoring or tooth grinding.

Light sleep periods: In each light sleep period, there were some clusters of tooth grinding and body movement event but only a few snoring events.

In this experiment, we found that the distribution of sound event clusters changed simultaneously with the sleep stage change, for not only Subject 2, but also the other subjects. Even though our analysis includes other subjects who have different primary sleep disorders and varying pattern of sleep stages, the finding led to similar conclusions. For example, on Subject 4, the deep sleep periods were also obviously associated with the clusters of snoring, the number of body movements was notably more in the light periods and REM stages than in the deep periods, and no snoring clusters were found in REM stages.

From this experiment, we found that the transition of cluster dynamics and the changing of sleep stage are related. Since the sleep stage sequence is an important tool in the study of sleep pattern, its relation provides the possibility of discovering sleep patterns based on the cluster map of sleep-related sound data from Sb-KSOM.

4. Conclusion

This proposed method combined the advantages of kernelization and sequence-based technologies, and obtained a fine-grained map that depicts the distribution and changes of sleep-related events. According to the experiment results, we can find out that MFCC feature does not performs well on this kind of sound data, and the KL kernel SOM has the best performance among KL, polynomial and RBF kernels, which improved by about 10% from standard SOM.

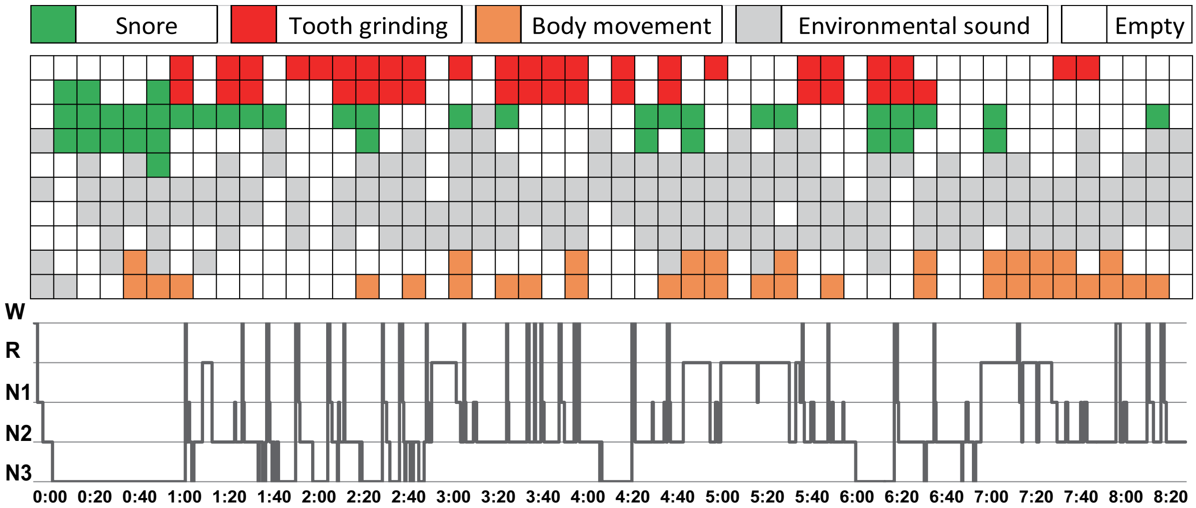

Fig 1: Cluster map generated by Sb-KSOM on Subject 2

Since the final objective of our research is to make the assessment of personal sleep quality more economical, more practical and more reliable, the correlation between sound data and sleep stages provides a new train of thought for studying the sleep pattern. In the future work, we will proceed to develop a predictive model for personal sleep quality scoring, the sleep-related sound data will be a main part of input data, and the relationship between sleep stages and sound events will play a key role in the algorithm development.

Acknowledgments

This research is partially supported by the Center of Innovation Program from Japan Science and Technology Agency, JST, the Grant-in-Aid for Scientific Research (B)(#25293393) from the JSPS, and Challenge to Intractable Oral Diseases from Osaka University Graduate School of Dentistry.

References

[Andras 02] Andras, P.: Kernel-kohonen networks, *International Journal of Neural Systems*, Vol. 12, No. 02, pp. 117–135 (2002)

[Behar 13] Behar, J., Roebuck, A., Domingos, J. S., Gederi, E., and Clifford, G. D.: A review of current sleep screening applications for smartphones, *Physiological measurement*, Vol. 34, No. 7, pp. R29–R46 (2013)

[Boulet 08] Boulet, R., Jouve, B., Rossi, F., and Villa, N.: Batch kernel SOM and related Laplacian methods for social network analysis, *Neurocomputing*, Vol. 71, No. 7, pp. 1257–1273 (2008)

[Chen 13] Chen, Z., Lin, M., Chen, F., Lane, N. D., Cardone, G., Wang, R., Li, T., Chen, Y., Choudhury, T., and Campbell, A. T.: Unobtrusive sleep monitoring using smartphones, in *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, pp. 145–152IEEE (2013)

[Choe 10] Choe, E. K., Kientz, J. A., Halko, S., Fonville, A., Sakaguchi, D., and Watson, N. F.: Opportunities for computing to support healthy sleep behavior, in *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pp. 3661–3666ACM (2010)

[Davis 80] Davis, S. and Mermelstein, P.: Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, *IEEE transactions on acoustics, speech, and signal processing*, Vol. 28, No. 4, pp. 357–366 (1980)

[Fukui 11] Fukui, K., Akasaki, S., Sato, K., Mizusaki, J., Moriyama, K., Kurihara, S., and Numao, M.: Visualization of damage progress in solid oxide fuel cells, *Journal of Environment and Engineering*, Vol. 6, No. 3, pp. 499–511 (2011)

[Kohonen 95] Kohonen, T.: Self-organisation maps (1995)

[Mantua 16] Mantua, J., Gravel, N., and Spencer, R.: Reliability of Sleep Measures from Four Personal Health Monitoring Devices Compared to Research-Based Actigraphy and Polysomnography, *Sensors*, Vol. 16, No. 5, p. 646 (2016)

[Wu 16] Wu, H., Kato, T., Yamada, T., Numao, M., and Fukui, K.: Sleep Pattern Discovery via Visualizing Cluster Dynamics of Sound Data, in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pp. 460–471Springer (2016)