33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

# Analysis of Time-Series Correlation Between Weighted Lifestyle Data and Health Data

Hiroshi Takeuchi, *Senior Member*, *IEEE*, Yuuki Mayuzumi, and Naoki Kodama

*Abstract*— **The time-series data analysis described here is based on the simple idea that the accumulation of the effects of lifestyle events, such as ingestion and exercise, could affect personal health with some delay. The delay may reflect complex bio-reactions such as those of metabolism in a human body. In the analysis, the accumulation of the effects of lifestyle events is represented by a summation of daily lifestyle data whose time-series correlation to variations of health data is examined (healthcare- data-mining). The concept of weighting is introduced for the summation of daily lifestyle data. As a result, it is suggested that the nature of personal health could be represented by a weighting pattern characterized by a small number of parameters.**

## I. Introduction

Application of the Internet to healthcare such as m-health and p-health are current topics of interest [1]- [3]. In particular, the demand for personalized healthcare systems to prevent diseases and improve health has been increasing recently [4]. Within this context we have developed a personal dynamic healthcare system (PDHS) utilizing the Internet [5]. It enables time-series of daily- health and lifestyle data to be stored in a database by utilizing a mobile phone. In addition, it can extract personally useful information such as rules and patterns concerning lifestyles and health conditions embedded in daily time-series personal health and lifestyle data. We call this 'healthcare data mining'.

In the healthcare- data-mining process [6], first we check the correlation between variations of the time-series health data and summations of the time-series lifestyle data. Then, if the correlation coefficient is larger than a certain threshold value, the lifestyle is selected as an independent variable in the data-mining process relevant to the health condition. In this study, we introduced the concept of weighting characterized by a small number of parameters in the summation process of daily time-series lifestyle data.

## II. Materials and Methods

### A. Analysis Method

The time-series data analysis described here is based on the simple idea that the accumulation of the effects of lifestyle events, such as ingestion and exercise, could affect personal health with some delay [6]. The delay may reflect complex bio-reactions such as those of metabolism in a human body. In the analysis, the accumulation of the effects of lifestyle events is represented by a summation of energy supply or expenditure data (calories) due to ingestion, exercise, etc. The accumulation of the effects may cause variation of health data, such as body- mass- index and body- fat percentage, with some delay.

In the analysis, we examine the correlation coefficient, *r*, described as:

$$r(\Delta h_{nm}, e^t{}_{ij}) = \frac{Cov(\Delta h_{nm}, e^t{}_{ij})}{SD(\Delta h_{nm})SD(e^t{}_{ij})} \qquad (1)$$

Here,

$$\Delta h_{nm} = h_n - h_m \qquad (2)$$

is the difference of time-series health data $h$ , representing the variation of health condition, and

$$e^t_{ij} = e_i + e_{i-1} + .... + e_j \qquad (3)$$

is the summation of time-series lifestyle data $e$ during a certain period, representing the accumulation of the effects of lifestyle events. The delay is represented by retardation, $s = n - i \geq 1$ (Fig.1). In Eq.(1), $SD(\Delta h_{nm})$ and $SD(e^t{}_{ij})$ are the standard deviation of $\Delta h_{nm}$ and $e^t_{ij}$ , respectively, and $Cov(\Delta h_{nm}, e^t{}_{ij})$ is the covariance.

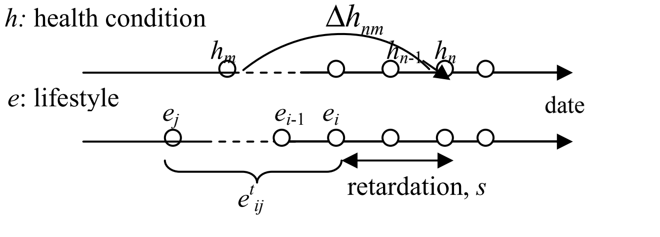


Fig.1 Reference figure for checking time-series correlation between variation of health condition and accumulation of of effects of lifestyle events.

Manuscript received March 26, 2011. This work was supported in part by Grants-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

Hiroshi Takeuchi is with the Department of Healthcare Informatics, Takasaki University of Health and Welfare, 37-1, Nakaorui-machi, Takasaki-shi, Gunma, 370-0033, Japan (phone: +81-27-352-1290; fax: +81-27-353-2055; e-mail: htakeuchi@ takasaki-u.ac.jp).

Yuuki Mayuzumi is a PhD student, Graduate School of Takasaki University of Health and Welfare, 37-1, Nakaorui-machi, Takasaki-shi, Gunma, 370-0033, Japan (e-mail: 0910404@takasaki-u.ac.jp).

Naoki Kodama is with the Department of Healthcare Informatics, Takasaki University of Health and Welfare, 37-1, Nakaorui-machi, Takasaki-shi, Gunma, 370-0033, Japan (e-mail: kodama@takasaki-u.ac.jp).

978-1-4244-4122-8/11/$26.00 ©2011 IEEE

The correlation coefficient, $r$ is estimated for the time-series health and lifestyle data in a certain period by changing $n - m$, $i - j$, and $s$ as parameters. If the maximum value of $r$ is larger than a certain threshold value, the lifestyle is selected as an independent variable in the data-mining process relevant to the health condition.

From this analysis, we have found that the parameter set ($n$-$m$, $i$-$j$, $s$) at which $r$ becomes largest significantly depends on the person [7]. It might be possible to characterize the nature of personal health by summation-day number, $(i\text{-}j+1)_{\max}$, and retardation, $s_{\max}$, at which $r$ becomes largest.

### B. *Introduction of Weighting*

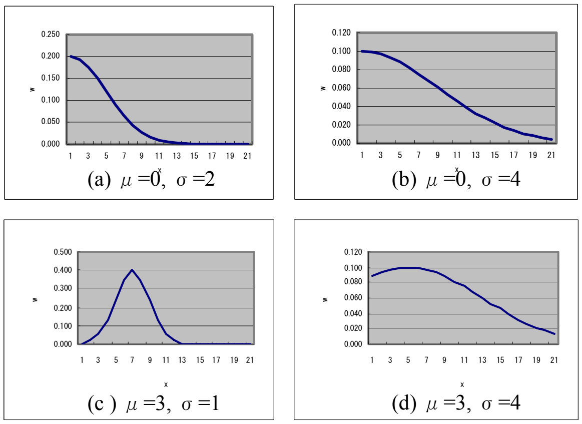
Equation (3) is a simple summation of daily time-series lifestyle data in a certain period. This may be too simplified to represent the accumulation of the effects of lifestyle events. Thus, we introduced a weighting in the summation of daily time-series lifestyle data. By carefully examining previous results, we proposed a weighting method utilizing the normal distribution function, which can be characterized by two parameters, $\mu$ and $\sigma$. The weighting function is described as:

$$w(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \qquad (\sigma\,,\ \mu > 0) \qquad (4)$$

When we use Eq. (4) as a weighting function in the summation of daily time-series lifestyle data, the values of $x$ are 0, 1, 2, …., $m$, and corresponding $w(0)$, $w(1)$, $w(2)$,….., $w$(m) are assigned as weighting coefficients descendent in date. For example, when $m$ = 10, the weighted summation of daily time-series lifestyle data is expressed as,

$$w(0)e_{n-1} + w(1)e_{n-2} + w(2)e_{n-3} + \ldots.. + w(10)e_{n-11} \qquad (5)$$

Here, we assume that the cursor date is $n$ in the time-series health data (Fig.1). We calculated the correlation coefficient, $r$, between Eq. (5) and $\Delta h_{nm}$ to find a suitable weighting pattern characterized by $\mu$ and $\sigma$ so that $r$ becomes largest. Weighting patterns for some ( $\mu$ , $\sigma$ ) sets are shown in Fig. 2. If $r$ becomes largest for the (a) pattern, an accumulation of the effects of quite recent lifestyle events affects the health condition without delay. If $r$ becomes largest for the (b) or (d) pattern, an accumulation of the effects for a long term affects the health condition. If $r$ becomes largest for the (c) pattern, an accumulation of the effects for a short term affects the health condition with some delay.



(a) $\mu$ =0, $\sigma$ =2

(b) $\mu$ =0, $\sigma$ =4

(c ) $\mu$ =3, $\sigma$ =1

(d) $\mu$ =3, $\sigma$ =4

Fig.2 Weighting patterns for some ( $\mu$ , $\sigma$ ) sets (horizontal : $x$, vertical: $w$).

## III. Results

### A. *Acquisition of Time-series Data*

Time-series data of body- fat, energy expenditure due to exercise, and energy supply due to ingestion were obtained from a 59-year-old male almost every day for about a half year. Body-fat percentage was measured with a reactance meter (Tanita, Japan). Energy expenditure due to exercise was measured with a wearable monitor (Omron, Japan) and energy supply was estimated from each day's breakfast, lunch and dinner contents.

### B. *Correlation Between Body-Fat and Energy Expenditure*

The weighting pattern used in the analysis was Eq. (4) with combinations of $\mu$ = 0, 1, 2, 3, and $\sigma$ = 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0 (32 patterns in total). The weighting coefficients obtained from Eq. (4), $w(0)$, $w(1)$, $w(2)$,…., $w(10)$, were normalized as follows:

$$\sum_{i=0}^{10} w(i) = 1 \qquad (6)$$

Time-series data of body-fat percentage and energy expenditure were analyzed for 2 different target terms,: (I) from 01/03/2005 to 31/05/2005 (for 3 months), and (II) from 01/03/2005 to 31/08/2005 (for 6 months). It was found that the correlation coefficient, $r$, became largest in the case of $\mu$ = 2 and $\sigma$ = 1.0 for both target terms. Scatter plots for the respective target terms are shown in Figs. 3 and 4.

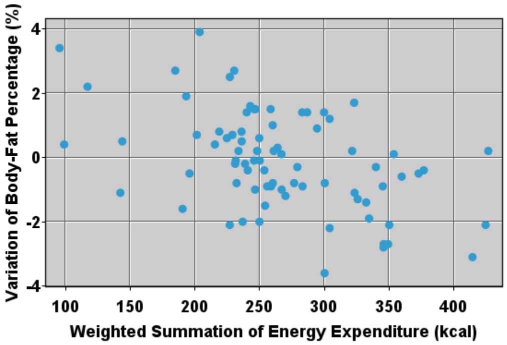


Fig.3 Scatter plots of variation of body-fat percentage vs. weighted summation of energy expenditure (target term I: 3 months) ($n$ = 79, $r$ = -0.482).

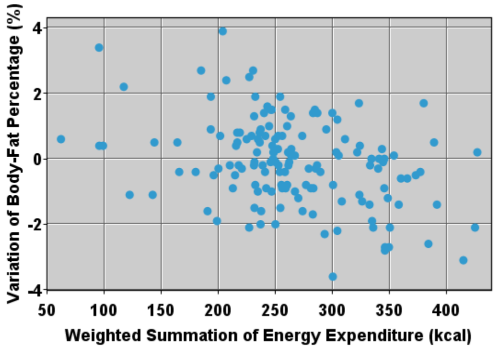


Fig.4 Scatter plots of variation of body-fat percentage vs. weighted summation of energy expenditure (target term II: 6 months) ($n$ = 148, $r$ = -0.367).

Negative correlations between variations of body-fat percentage and weighted summations of energy expenditure were observed for both target terms ($p < 0.01$).

For comparison, the time-series correlation based on the previous method (simple summation with a possible retardation) was analyzed. Maximum correlations were found when $i$ - $j$ = 3 and $s$ = 2 for both target terms. Scatter plots for the respective terms are shown in Figs. 5 and 6.

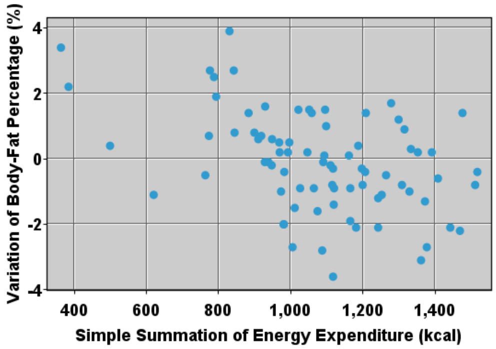


Fig.5 Scatter plots of variation of body-fat percentage vs. simple summation of energy expenditure with delay (target term I: 3 months) ($n$ = 79, $r$ = -0.452).

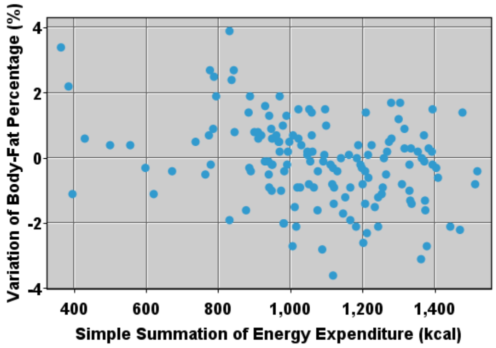


Fig.6 Scatter plots of variation of body-fat percentage vs. simple summation of energy expenditure with delay (target term II: 6 months) ($n$ = 148, $r$ = -0.323).

The maximum correlation coefficients were smaller than in the weighted summation cases, although the differences did not affect the results of a correlation test.

### C. *Correlation Between Body-Fat and Energy Supply*

Time-series data of body-fat percentage and energy supply were analyzed for two different target terms,: (I) from 01/03/2005 to 31/05/2005 (for 3 months), and (II) from 01/03/2005 to 31/08/2005 (for 6 months). The weighting patterns used were the same as before (32 patterns in total).

It was found that $r$ became largest when $\mu$ = 3 and $\sigma$ = 2.0 for target term I, and when $\mu$ = 3 and $\sigma$ = 1.0 for target term II. Scatter plots for the respective terms are shown in Figs 7 and 8.

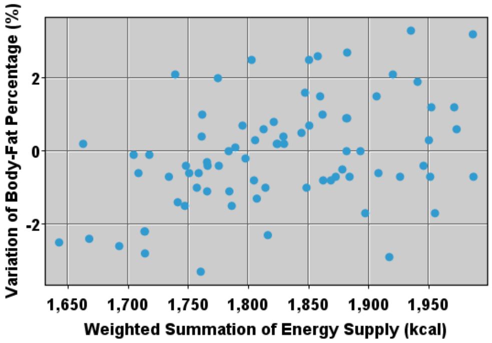


Fig.7 Scatter plots of variation of body-fat percentage vs. weighted summation of energy supply (target term I: 3 months) ($n$ = 79, $r$ = 0.400).

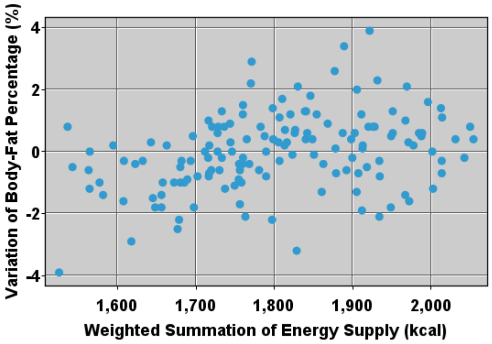


Fig.8 Scatter plots of variation of body-fat percentage vs. weighted summation of energy supply (target term II: 6 months) ($n$ = 148, $r$ = 0.342).

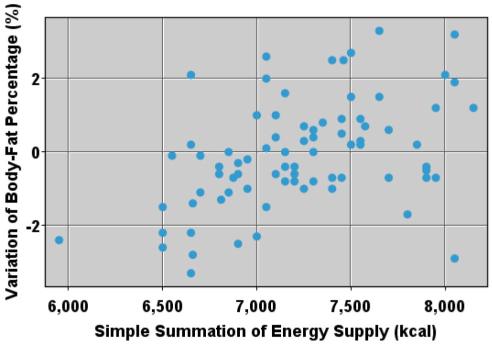


Fig.9 Scatter plots of variation of body-fat percentage vs. simple summation of energy supply with delay (target term I: 3 months) ($n$ = 78, $r$ = 0.446).

Positive correlations between variations of body-fat percentage and weighted summations of energy supply were observed for both target terms ($p < 0.01$).

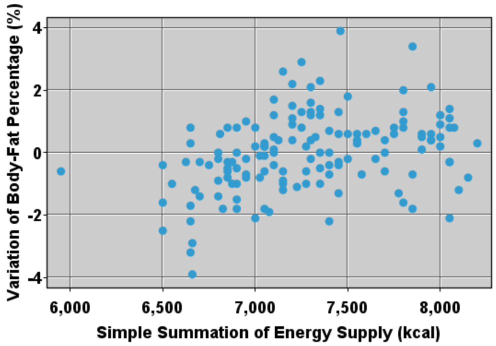


Fig.10 Scatter plots of variation of body-fat percentage vs. simple summation of energy supply with delay (target term II: 6 months) ($n$ = 147, $r$ = 0.352).

For comparison, the time-series correlation based on the previous method was analyzed. Maximum correlations were found when $i$ - $j$ = 3 and $s$ = 3 for both target terms. Scatter plots for the respective terms are shown in Figs. 9 and 10.

## IV. Discussion

### A. Comparison to Previous Method

The maximum correlation between variations of body-fat percentage and weighted summation of energy expenditure was obtained when $\mu$ = 2 and $\sigma$ = 1.0. The weighting pattern at this condition is shown in Fig. 11. The maximum correlation was also obtained when $i - j$ = 3 and s = 2 by the previous method. This condition corresponds to weighting with $w(0) = 0$, $w(1) = w(2) = w(3) = w(4)$, $w(5)$~$w(10)$ = 0. Thus, this weighting pattern is also shown in the figure. The figure shows that the weighting pattern with $\mu$ = 2 and $\sigma$ = 1.0 was reasonably selected from 32 patterns.

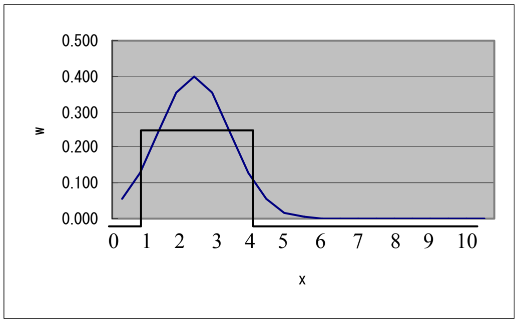


Fig. 11 Weighting pattern at which maximum correlation occurred between variations of body-fat percentage and weighted summation of energy expenditure ($\mu$ = 2, $\sigma$ = 1.0) (corresponding weighting pattern in previous method also shown).

Maximum correlation between variations of body-fat percentage and weighted summation of energy supply was obtained when $\mu$ = 3 and $\sigma$ = 2.0 for target term I, and when $\sigma$ = 1.0 for target term II. The weighting patterns at these conditions are shown in Fig. 12. Maximum correlation was also obtained when $i - j$ = 3 and s = 3 by the previous method. This condition corresponds to weighting with $w(0) = w(1) = 0$, $w(2) = w(3) = w(4) = w(5)$, $w(6)$~$w(10)$ = 0. Thus, this weighting pattern is also shown in the figure. The figure shows that the weighting pattern with $\mu$ = 2 and $\sigma$ = 1.0 or 2.0 was also reasonably selected from 32 patterns.

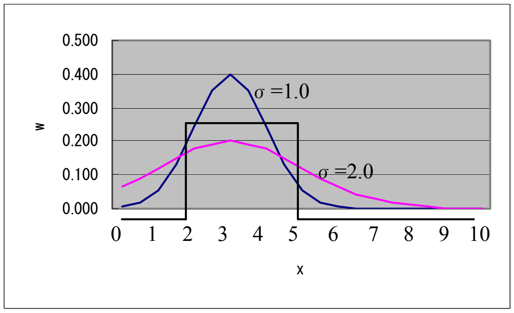


Fig. 12 Weighting pattern at which maximum correlation occurred between variations of body-fat percentage and weighted summation of energy supply ($\mu$ = 3, $\sigma$ = 1.0, 2.0) (corresponding weighting pattern in previous method also shown).

### B. Characterization of Personal Health Nature

The effect of the weighting was not so dramatic as we expected. It was suggested, however, in this study, that correlations between time-series lifestyle and health data can be reasonably extracted using weighting patterns characterized by two parameters, $\mu$ and $\sigma$. These parameters are more visible compared to previous ones (summation-day number, $i$-$j$+1, and retardation, $s$) for characterizing the nature of personal health in the correlation between lifestyles and health conditions. The combination of $\mu$ and $\sigma$ might be a useful parameters for clustering the nature of personal health.

## References

[1] E. C. Kyriacou, C. S. Pattichis, and M. S. Pattichis, "An overview of recent health care support system for eEmergency and mHealth applications," *Proc. 31st Annual International Conference of the IEEE EMBS*, 2009, pp. 1246-1249..

[2] R. S. H. Istepanian, A. Sungoor, and K. A. Earle, "Technical and compliance considerations for mobile health self-monitoring of glucose and blood pressure for patients with diabetes," *Proc. 31st Annual International Conference of the IEEE EMBS*, 2009, pp. 5130-5133..

[3] H. Kumpusch, D. Hayn, K. Kreiner, M. Falgenhauer, J. Mor, and G. Schreier, "A mobile phone based telemonitoring concept for the simultaneous acquisition of biosignals and physiological parameters, *Proc. 13trd World Congress on Medical and Health Informatics*, 2010, pp.1344-1348.

[4] I. Korhonen, E. Mattila, A. Ahtinen, J. Salminen, L. Hopsu, R. Lappalainen, and T. Leino, "Personal health promotion through personalized health technologies – Nuadu experience," *Proc. 31st Annual International Conference of the IEEE EMBS*, 2009, pp. 316-319.

[5] H. Takeuchi, N. Kodama, T. Hashiguchi, and N. Mitsui, "Healthcare data mining based on a personal dynamic healthcare system," *Proc. 2nd Int. Conf. on Computational Intelligence in Medicine and Healthcare,* 2005, pp. 37 -43.

[6] H. Takeuchi, N. Kodama, T. Hashiguchi, and D. Hayashi, "Automated healthcare data mining based on a personal dynamic healthcare system," *Proc. 28th IEEE EMBS Annual Int. Conf.* 2006, pp.3604 -3607.

[7] H. Takeuchi, Y. Ikeda, and N. Kodama, "Time-series data analyses for healthcare-data-mining based on a personal dynamic healthcare system," *Proc. 12th World Congress on Medical Informatics* (MEDINFO), 2007, P309.