Use of Hidden Markov Models to Identify Background States Behind Risks of Cerebral Infarction and Ischemic Heart Disease

Hiroshi Morimoto¹

¹Graduate School of Environmental Study, Nagoya University, Nagoya 464 Japan Correspondence: Hiroshi Morimoto, Graduate School of Environmental Study, Nagoya University, Nagoya 464 Japan. E-mail: hiroshim@info.human.nagoya-u.ac.jp

Received: November 11, 2016	Accepted: December 16, 2016	Online Published: December 30, 2016
doi:10.5539/jmr.v9n1p24	URL: http://dx.doi.org/10.5539/jmr.v9n1p24	

Abstract

Cold exposure is often said to trigger the incidence of cerebral infarctions and ischemic heart disease. This association between weather and human health has attracted considerable interest, and has been explored using standard statistical techniques such as regression models. Meteorological factors, such as temperature, are controlled by background systems, notably weather patterns. Therefore, it is reasonable to posit that the incidence of diseases is similarly influenced by a background system. The aim of this paper was to identify and construct these respective background systems. Possible background states or "hidden states", behind the incidence of diseases were derived using the EM and Viterbi algorithms with in the framework of hidden Markov models (HMM). A self-organizing map (SOM) enabled identification of weather patterns, considered as background states behind meteorological factors. These background states were then compared, and the hidden states behind the incidence of diseases were identified by six weather patterns. This finding indicates new evidence of the links between weather and human health, shedding light on the association between changes in the weather and the onset of disease.

Keywords: cerebral infarction, hidden Markov model, ischemic heart disease, stroke, weather

1. Introduction

Several scholars have used standard statistical techniques such as regression models to study the correlation between low temperature and the incidence of disease (e.g., McDonalda, McDonaldb, Bidae, Kallmesb, & Cloft, 2012; Jimenez-Conde, 2008). These studies have directly examined the variables of meteorological factors and the incidence of cerebral infarction for the use of regression models. However, the mechanism whereby weather changes affect the onset of diseases still remains unclear.

This study investigated the relation between the background states behind the onset of diseases and those behind meteorological factors. An examination of the correlation between these background states revealed the links between the incidence of diseases and weather changes.

EM and Viterbi algorithms were performed to explore the possible background states of the incidence of disease. These algorithms were used to construct a hidden Markov model (HMM). A reasonable assumption was that background states could relate to meteorological data such as temperature entailed in weather charts or weather patterns. A self-organizing map (SOM) was constructed to obtain weather patterns as background states of meteorological data, as this is a versatile method for classifying multidimensional data.

During the process of constructing the HMM, no weather-related information was used. Further, no disease-related information was used to establish a classification of weather patterns. Demonstration of a correlation between these two kinds of background states would, therefore, constitute new evidence of the association between disease and human health, introducing an insight in the relation between incidences of disease and the weather. The existence of a lag period between temperature-related exposure, and its effect on mortality, could elucidate the nature of the onset of diseases. Thus, an assessment of lagged effects was of potential importance in this study which adopted a methodology that enabled their incorporation.

2. Methods

The data on daily numbers of patients for the period 2002 - 2004 were obtained from the Fire Department in the city of Nagoya in Japan. Specifically, the data comprised the patients of all ages, who were first transported by ambulance to a hospital and subsequently diagnosed, at the hospital, with cerebral infarction or ischemic heart disease after their

admission. The meteorological data comprised a selection of daily data that were obtained from the Japan Meteorological Agency. These data included the mean temperature, maximum temperatures, the hours of sunshine and so on.

HMMs are used to express random change of states over a time-based series of observations (Figure 1). In general, HMM is a method to represent random process over sequences of observations. The observation at time *t* is denoted by the variable R(t). In this paper, R(t) represented the number of patients, therefore R(t) is supposed to be non-negative integers. A hidden Markov model consisted of two random processes. First, it supposes that the observation R(t) was generated randomly by some process from a state S(t), which is hidden from the observer. Second, it assumes that the state S(t) is determined randomly from the previous state S(t-1). Both random processes are assumed to be Markov process.

The set of all the hidden states is denoted by $S = \{S_1, ..., S_n\}$. A hidden state in *S* is supposed to change to another state in *S* as time changes from *t* to *t*+1. The probability P_{ij} is defined as the probability that one hidden state S_j changes to another state S_i ,

$$P_{ij} = P(S(t+1) = S_i | S(t) = S_j)$$

This gives us $n \times n$ numbers of probabilities, thus forming (n, n)-matrix $P = (P_{ij})$, which is called a "transition matrix".

We denote the probability of the occurrence R(t) for given state S(t) by P(R(t) | S(t)). Let *m* be the maximum number of R(t) during the period considered. Then R(t) takes the value from 0 to *m*. Let *R* be the set of numbers from 0 to *m*. The above probability forms together (m+1,n)- matrix $Q = (Q_{ij})$,

$$Q_{ii} = P(R(t)=i-1 | S(t) = S_i),$$

where i = 1, ..., m+1 and j = 1, ..., n. Each column of the above matrix gives the distribution of the observed data for given hidden state.

As a consequence, we have a set {S(0), S, R, P, Q}, where S(0) represents the initial state. This set is called a Hidden Markov Model. In this paper, for given R(t), these two matrices P, Q and the initial state S(0) were calculated by EM algorithm. A possible sequence of S(t) (t > 0) was calculated by Viterbi algorithm.

In this paper, "observation data" were the daily data on numbers with patients of cerebral infarction or ischemic heart disease described above. For basic elements of HMM, see Ghahramani (2001) and see Morimoto (2015; 2016a) for applications to this field.

A sequence of "states" in HMM was considered to represent a "background" process that is occurring behind "observed data". We can posit that a "background" also exists for the incidence of diseases. To investigate this, the time series of these background states was calculated using EM (expectation-maximization) and Viterbi algorithms.

The EM algorithm is an estimation procedure aimed at identifying the parameters of a HMM (Ghahramani, 2001). The Viterbi algorithm is a formal technique used for finding the single best state sequence. It provides an optimal solution for the problem of estimating the state sequence of HMM (Lou, 1995). For this study, the parameters of the HMM were first defined using an EM algorithm. Subsequently, a Viterbi algorithm was applied to calculate the sequence of hidden states, using "R" software.

A self-organizing Map (SOM) is a useful data mining technique. It was first introduced by Kohonen (1982). It aims to produce overviews of multivariate data sets (the input layer), and to visualize these on a planar two-dimensional lattice in plane (the target layer). Using a principle of artificial neural networks, the SOM algorithm tries to find prototype vectors that represent the input data set and at the same time realize a continuous mapping from input space to a lattice. This lattice consists of a finite number of "neurons" (or "units") and forms a two-dimensional lattice. The important factor of the SOM algorithm is that the "weight" vectors of neurons which are first initialized randomly, come to represent a number of original structure of vectors during an recursive data input process. As a result of SOM, it holds that if the points in original data are "near" (or "distant"), then they were mapped to "near" (or "distant") units in plane. Thus a SOM can be used to make visible the cluster structure of the data.

In this study, a SOM was applied to the daily data for weather elements (e.g., maximum daily temperature) in Nagoya city. Here we used the a "standard" SOM, based on unsupervised neural learning algorithms.

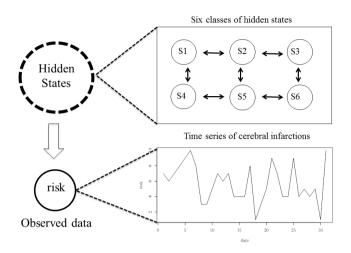


Figure 1. A scheme of hidden Markov models

3. Results

We extracted the data of patients with cerebral infarction and ischemic heart disease in both winter and summer seasons for the period 2002-2005 from the data provided by the Nagoya City Fire Department. As a result, these data comprised the time series (daily data) of the number of patients during a period of about 90 days each for winter (December, January and February) and summer (June, July, August for summer). The time series of the number of patients, which was called the "observed data", was expressed as r(t) according to the following equation:

r(t), t=1,...,T, (T=92 for the summer season, or 90 for the winter season).

The EM algorithm was used to identify the parameters of the HMM, where r(t) was the given observed data and the number of hidden states was set at six. The value "six" was chosen as this was the number of appropriate weather patterns. After applying the EM algorithm, The Viterbi algorithm was next applied to identify the time series of six hidden states:

$$S1(t), S2(t), ..., S6(t), t=1,..., T$$

Figure 2 provides an illustrative example of r(t) and S2(t) for data relating to the incidence of ischemic heart disease during the 2004 winter season. The weather effect is assumed to be delayed by 2 days. In this figure, the time series r(t) was expressed as a polygonal line and S2(t) was expressed as a vertical graduation of gray. Here S2(t) denoted the probability of the hidden state at time t becoming S2. A higher probability corresponded to a darker graduation. Thus the variability of both S2(t) and r(t) could be simultaneously observed. Peaks of r(t) evidently corresponded, on the whole, to the states S2(t).

To identify the backgrounds behind the meteorological data, an SOM was applied to the daily data of eight weather elements in Nagoya: maximum temperature, minimum temperature, precipitation, humidity, local pressure, wind velocity, the hours of daylight and solar radiation. The number of units was set at six. We used the so-called "standard" SOM rather than bi-directional SOM.

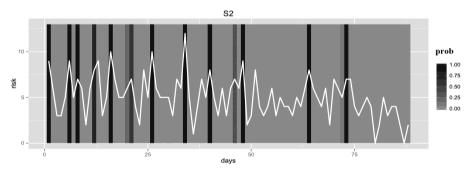


Figure 2. A graph depicting the risk r(t) of ischemic heart disease and hidden state S2(t).

Note. The horizontal line denotes the days in winter in 2004. The delay had a value of 2

As a target layer of the SOM, we selected lattices of 3×2 units (a total of 6 units). Thus, SOM resulted in six classes of

"weather states". These classes, shown in Figure 3, were denoted by (W1) –(W6) and comprised two groups. The first group {W1, W2} entailed high air pressure of the local atmosphere, and the second group {W3, ... W6} entailed low air pressure. The following weather classes were included: warm weather with high air pressure (W1), cold weather with high air pressure (W2), cold and windy weather (W3), rainy weather (W4), warm weather with low air pressure (W5) and humid weather (W6). These six classes, shown in Figure 3, were abbreviated as warm high pressure (W1), cold high pressure (W2), cold low pressure (W3), rain (W4), warm low pressure (W5), and humid (W6) (Morimoto, 2016b).

The hidden states obtained using the EM and Viterbi algorithms were compared with the weather patterns mapped with the SOM. For this comparison, correlations between hidden states and weather states were calculated.

Lagged effects of cold exposure were observed for the incidence of diseases (Morimoto, 2016b) [8]. To include these lag effects, the following lagged time series of weather patterns were considered:

W1(t-d), W2(t-d), ..., W6(t-d) instead of W1(t), W2(t), ..., W6(t),

where d represented a lag.

To sum up, the following three types of time series were defined, for a given season comprising three months of a certain year:

hidden states of HMM: S1(t), S2(t), ..., S6(t)

weather states: W1(t-d), W2(t-d), ..., W6(t-d),

risk series: r(t),

where d denoted lag.

The series S1(t), ..., S6(t) was derived from the series of risks r(t) obtained using the EM and Viterbi-algorithm. The series W1(t-d),...,W6(t-d) was given by the SOM using the meteorological data.

To investigate the relations between hidden states of the HMM and weather states, the correlations of these series were calculated. For the given integers, i, j=1, ..., 6, we defined a matrix (d_{ij}) , wherein d_{ij} was the correlation of Wi(t-d) with Sj(t) as follows:

$$d_{ij} = \operatorname{cor}(Wi(t-d), Sj(t)), i, j=1, ..., 6.$$

Thus (d_{ii}) formed a matrix entailing 6×6 dimensions.

Moreover, the correlation of r(t) with Sj(t), j=1, ..., 6 was calculated to investigate which Sj(t) represented the risk r(t). Thus a vector with six elements was formed.

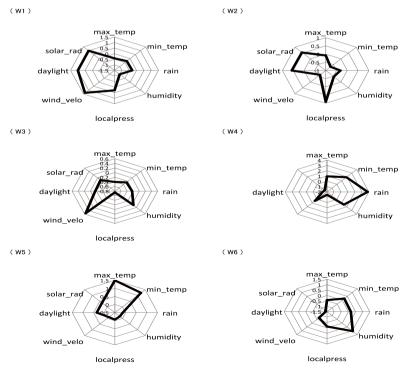


Figure 3. The six weather patterns classified by SOM.

Table 1 depicts an example of this matrix and the vector. The first six rows expressed the matrix of correlations and the last row the vector of correlation with r(t). The period of calculation covered days of December 2004 and January and February 2005 (the winter season of 2004). The concerned disease was ischemic heart disease and the lag *d* had a value of 2. In Table 1, a relatively high value of 0.31 was obtained for the correlation of W6 and S2. The value of the correlation of S2 and risk was 0.66, indicating that the risk was well expressed by the hidden state S2.

To investigate the stochastic significance of these correlation values, the Pearson's test was performed for all of the correlations. Thus p-values were obtained for all of the correlations corresponding to Table 1. The results for p-values were listed in Table 2 comprising a 6×6 matrix and a vector in the last row.

Table 1. An example of the correlations of hidden states with weather states during the winter season 2004–2005 with a lag value of 2 for ischemic heart disease.

	S 1	S2	S 3	S4	S5	S6
W1	0.03	0.07	0.18	-0.03	-0.27	0.05
W2	0.05	-0.18	0.14	-0.07	0.09	-0.02
W3	-0.11	-0.15	-0.1	-0.07	0.36	-0.02
W4	0.03	-0.06	-0.06	0.12	0	0.03
W5	0.07	-0.09	-0.09	0.15	-0.08	0.13
W6	-0.05	0.31	-0.18	0.05	-0.07	-0.08
risk	0.09	0.66	0.19	0.31	-0.42	-0.71

Table 2. The p-values obtained for the Pearson's test to identify the correlations shown in Table 1.

	S 1	S2	S 3	S4	S5	S6
W1	0.76	0.51	0.09	0.78	0.01	0.63
W2	0.67	0.09	0.19	0.5	0.42	0.83
W3	0.33	0.16	0.36	0.53	0	0.84
W4	0.81	0.55	0.59	0.27	0.98	0.78
W5	0.52	0.39	0.39	0.17	0.47	0.23
W6	0.67	0	0.09	0.66	0.5	0.47
risk	0.41	0	0.07	0	0	0

The results shown in Table 2 enabled an assessment of whether or not the correlations shown in Table 1 were stochastically significant. In Table 2, p-values at the coordinates (W6, S2) and (risk, S2) were very small (almost zero). This indicated that the correlations of S2 with W6 was stochastically significant and that S2 was strongly correlated to the risk of disease. Therefore the W6 (humid low pressure) weather pattern was identified among the hidden states of the HMM, showing a strong correlation to the risk of ischemic heart disease.

Figure 4 shows the variability of three time series, S2(t), W6(t), and risk r(t). The hidden state, S2, was found to be well expressed by the W6 weather pattern, and the risk was found to be related to W6

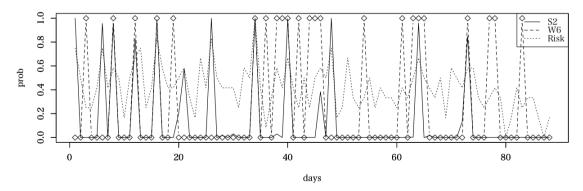


Figure 4. Combined plotting of the hidden state (S2) of the HMM, the W6 weather state, and the risk of ischemic heart disease. *Note*. The duration was the 2004–2005 winter season, and the lag had a value of 2.

The same procedure was performed for all of the possible combinations of the following three sets:

disease ={cerebral infarction, ischemic heart disease}

period = $\{2002 \text{ winter}, 2003 \text{ winter}, 2004 \text{ winter}, \}$

2002 summer, 2003 summer, 2004 summer, 2005 summer}

lag $d = \{0, 1, 2, \dots, 8\}.$

This procedure enabled a detailed investigation of the links between weather and the diseases under study. Tables 3 and 4 show the weather patterns that were found to have stochastically significant correlations with hidden states of the HMM

Table 3. Identified weather patterns for cerebral infarction

	small d	large d
winter	W1, W2	W4
summer	W2,W3,W4	W1,W2,W3,W5,W6

Table 3 shows that for cerebral infarctions, the W2 weather pattern featured in most of entries except during winter seasons with a large d (lag) value. For summer seasons, the W3 weather pattern (cold low pressure) was observed for d values that were both small and large. The W5 and W6 weather patterns (warm low pressure and humid low pressure, respectively) were observed during summer periods with large lag values. The W1 (warm high pressure) and W4 (rain) patterns were also observed.

Table 4. Identified weather patterns for Ischemic heart disease

	small d	large d
winter	W2,W6	W2
summer	W1,W2,W4	W2,W5,W6

For ischemic heart disease, the W2 weather pattern (cold high pressure) was a common feature for entries in Table 4. The W4 (rain) weather pattern was identified during the summer season with a small d value. The W5 (warm low pressure) and W6 (humid low pressure) weather patterns were observed as hidden states.

4. Discussion

For this study, two kinds of background systems were constructed. The first system for the HMM, was derived from the time series of the numbers of patients. The backgrounds were called "hidden states". The hidden states, which were constructed using the EM and Viterbi algorithms were intended to control the observed data (i.e., the time series of the numbers of patients shown in Figure 1). The second background system was derived by SOM from the time series of meteorological data including daily mean temperatures. This yielded six classes of weather patterns (W1–W6).

The first system (HMM) was derived using only the information on disease, and the second was constructed using only the information of meteorological data. Therefore, identifying the correlations between these backgrounds was critical for assessing the links between the incidence of diseases and weather changes.

Table 1 illustrates an example of the results was shown in. The hidden states of HMM and the weather states were denoted as follows:

W1(t-d), W2(t-d), ..., W6(t-d).

Here d denoted the lag. If d had a value of 0, this meant that there was no delay.

The study was restricted to the to winter and summer seasons. Therefore, during each three-month period, t varied from 1 to T (about 90 days), In this example, the winter season of 2004–2005 (i.e., December of 2004 and January and February of 2005) was considered, and the concerned disease was ischemic heart disease.

Table 1 showed all of the correlations of $S_j(t)$ with $W_i(t-d)$, together with the correlation of $S_j(t)$ and risk r(t). The stochastic significance, according to Pearson's test, was shown in Table 2. From this example, we can conclude that the hidden state S2 could be identified with W6 (humid weather of low pressure). It suggests that a connection exists between the incidence of ischemic heart disease and weather changes through hidden states.

The same procedure was applied for all winter and summer seasons from 2002 to 2005 for all possible lag values of d (0, ... 8), and for cerebral infarction and ischemic heart disease. Table 3 shows the associations of hidden states and weather patterns for cerebral infarction, and Table 4 shows these associations for ischemic heart disease. Thus, hidden states that dominated the incidence of particular diseases, by weather patterns, were identified.

It should be noted that the effects of weather varied for small or large d values. The W2 weather pattern (cold high pressure) was observed for most of categories in Tables 3 and 4. This indicated that exposure to cold during conditions of high pressure had a significant effect on the incidence of both cerebral infarction and ischemic heart disease.

Other patterns showed complex differences with respect to delays and diseases. It is noteworthy that the W5 weather pattern (warm low pressure) had an effect with a large lag for both cerebral infarction and ischemic heart disease. While cold exposure has been considered a main trigger for the onset of disease, recent findings show that warm conditions can also trigger the incidence of disease (Morimoto, 2016c).

For incidences of cerebral infarction that occurred during summer months, the W3 weather pattern (cold low pressure) was associated with both small and large lags. This indicates that cold exposure during the winter season influenced the onset of cerebral infarction.

For ischemic heart disease, the influence of the W1 weather pattern (warm high pressure) was observed during the summer. The W6 weather pattern (humid low pressure) could also be a trigger during summer months with a large lag.

At a molecular level, it is known that cold exposure results in extensive changes in the expression of genes (Dong et al. (2013). It has been reported that cold-triggered food-intake-independent lipolysis significantly increases plasma levels of small low-density lipoprotein (LDL) remnants, and stimulates atherosclerotic plaque growth. It has been also reported that the vascularization is directly caused by a cold-induced increase in sympathetic stimulation of the adipose tissues, which results in increased VEGF gene expression (Xue, et al. 2009). VEGF consequently leads to increased angiogenesis. These findings relate to both white adipose tissues (WAT) and brown adipose tissues (BAT). Acute exposure to cold could result in the stimulation of the sympathetic nerve system and cause an increase of noradrenaline.

Following the increase of noradrenalin stimulated by cold exposure, there may be two pathways leading to the incidence of disease, for both cerebral infarction and ischemic heart disease. One has a directive influence, causing shrinkage of blood vessels. The other may have a delayed effect on human health. Noradrenaline directly stimulates adipose tissues through beta-adrenergic activation of brown adipocytes (Foster & Frydman, 1978; Marcher et al. 2015). Therefore acute cold exposure stimulates the activity of adipose tissues and the expression of UCP1 (a thermogenic gene). This results in the increased LDL and plaque levels in the blood, and could lead to the onset of cerebral infarction and ischemic heart disease. This could be associated with time lags in cold exposure relating to the onset of diseases. Therefore, there were pathways with both small lags and large lags. These findings suggest that there a metabolism exists that bridges the gap between weather changes and the incidence of diseases.

5. Conclusion

Based on the observation of the variability of the daily series of the number of patients with cerebral infarction and ischemic heart disease, it is reasonable to suppose the existence of a background system that brings about the observed series of patients. Such a background system was realized by HMM, and the hidden states of HMM were found by the use of the EM and Viterbi algorithms. On the other hand, the daily series of the meteorological elements, such as mean temperature, must have a background system like "weather patterns". A SOM was carried out to find six classes of weather patterns.

We compared these two kinds of backgrounds, and identified association of hidden states of HMM with the six weather patterns. Thus, the study has provided a new method for conducting an in-depth investigation of the relation between weather and health. It sheds light on the links between weather and human health.

References

- Dong, M., Yang, X., Lim, S., Cao, Z., Honek, J., Lu, H., ...Cao,Y. (2013). Cold Exposure Promotes Atherosclerotic Plaque Growth and Instability via UCP1-Dependent Lipolysis. *Cell Metabolism*, 18, 118–129. http://dx.doi.org/10.1016/j.cmet.2013.06.003
- Foster, D. O., & Frydman, M. L. (1978). Nonshivering thermogenesis in the rat. II. Measurements of blood flow with microspheres point to brown adipose tissue as the dominant site of the calorigenesis induced by noradrenaline. Can. *J. Physiol. Pharmacol*, 56, 110–122. http://dx.doi.org/10.1139/y78-015
- Ghahramani, Z. (2001). An Introduction to Hidden Markov Models and Bayesian Networks. *Journal of Pattern Recog*nition and Artificial Intelligence, 15, 9-42. http://dx.doi.org/10.1142/S0218001401000836
- Jimenez-Conde, J., Ois A., Gomis, M., Rodriguez-Campello, A., Cuadrado-Godia, E., Subirana, I., & Roquer, J. (2008). Weather as a Trigger of Stroke. Cerebrovasc. 26(4), 348–354. doi:http://dx.doi.org/10.1159/000151637
- Kohonen, T. (1982). Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, 43, 59-69. http://dx.doi.org/10.1007/BF00337288

- Lou, H.L. (1995). Implementing the Viterbi algorithm. *IEEE Signal Processing Magazine*, *12*(5). http://dx.doi.org/10.1109/79.410439
- McDonalda, R. J., McDonaldb, J. S., Bidae, J. P., Kallmesb, D. F., & Cloft, H. J. (2012). Subarachnoid Hemorrhage Incidence in the United States Does Not Vary with Season or Temperature. Am. J. Neuroradiol, 33, 1663-1668. doi:http://dx.doi.org/10.3174/ajnr.A305
- Marcher, A. B., Loft, A., Nielsen, R., Sysi-Aho, M., Ekroos, K., & Mandrup, S. (2015). RNA-Seq and Mass-Spectrometry-Based Lipidomics Reveal Extensive Changes of Glycerolipid Pathways in Brown Adipose Tissue in Response to Cold. *Cell Reports*, 13, 2000–2013. http://dx.doi.org/10.1016/j.celrep.2015.10.069
- Morimoto, H. (2015). Study on Links between Cerebral Infarction and Climate Change Based on Hidden Markov Models. *International Journal of Social Science Studies*, *3*, 180-186. http://dx.doi.org/10.11114/ijsss.v3i5.1045
- Morimoto, H. (2016a). Hidden Markov Models and Self-Organizing Maps Applied to Stroke Incidence. *Open Journal of Applied Sciences*, 6, 158-168. http://dx.doi.org/10.4236/ojapps.2016.63017
- Morimoto, H. (2016b). Hidden Markov models to estimate the lagged effects of weather on stroke and ischemic heart disease. *Applied Mathematics*, 7, 1415-1425. http://dx.doi.org/10.4236/am.2016.71312
- Morimoto, H. (2016c). Association Analysis Identifies Risk of Ischemic Heart Disease When Temperature Increases. *International Journal of Social Science Studies*, 4(7), 55-62. http://dx.doi.org/10.11114/ijsss.v4i7.1650
- Xue, Y., Petrovic, N., Cao, R., Larsson, O., Lim, S., Chen, S., Feldmann, H. M., ... Cao, Y. (2009). Hypoxia-Independent Angiogenesis in Adipose Tissues during Cold Acclimation Cell Metabolism 9, 99–109. http://dx.doi.org/10.1016/j.cmet.2008.11.009

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).