Modeling of Quantification of Sensible Temperature by Analyzing Weather Conditions and Tweet Data

Takashi Maga
Graduate School of Creative Science and Engineering,
Waseda University, Tokyo, Japan
Tel: (+81) 3-5286-3290, Email: kksw.1214-dax@fuji.waseda.jp

Kenta Mikawa
Department of Information Science,
Shonan Institute of Technology, Kanagawa, Japan
Tel: (+81) 466-30-0212, Email: mikawa@info.shonan-it.ac.jp

Masayuki Goto
School of Creative Science and Engineering,
Waseda University, Tokyo, Japan
Tel: (+81) 3-5286-3290, Email: masagoto@waseda.jp

Tomohiro Yoshikai
Japan Weather Association,
Tokyo, Japan
Tel: (+81) 3-5958-8154, Email: yoshikai.tomohiro@jwa.or.jp

Abstract. In the field of industrial management, demand forecasting is important for production planning and inventory management. In the case of retail store chains which sell large-variety of merchandises in various regions, problems such as over-stocking and waste disposal are being still occurred caused by demand fluctuation depending on changes in weather conditions. By analyzing various kinds of big data which can be stored in databases, it is desirable to improve the accuracy of demand forecasting and realize effective inventory management with micro-perspective. However, especially in case of food products, demand fluctuation would be highly influenced by sensible temperature that consumers feel more than absolute weather conditions. In this study, we propose the method of quantifying a sensible temperature index for demand forecasting by analyzing digital text data of Twitter. This method is based on the assumption that information of sensible temperature of public consumers is appeared in Twitter data. We analyze the real data by applying our proposed model for sensible temperature quantification.

Keywords: Sensible Temperature; Demand Forecasting; Big Data; Data analysis

1. INTRODUCTION

In the field of industrial management, demand forecasting is important for production planning and inventory management. In the case of retail store chains which sell large-variety of merchandises in various regions, problems such as over-stocking and waste disposal are being still occurred because demand fluctuation depending on changes in weather conditions. By analyzing various kinds of big data which can be stored in databases, it is desirable to improve the accuracy of demand forecasting and realize effective inventory management with micro-perspective. However, especially in case of food products, demand fluctuation would be highly influenced by sensible temperature that consumers feel more than absolute weather conditions. For example, the demand of cold Chinese noodle depends strongly on sensible temperature felt by consumers. Though the sensible temperature is related with absolute weather condition, it does not perfectly match absolute temperature. If consumers' sensible temperature can be quantified, this information is useful for demand forecasting.

Quantifying the sensible temperature has been studied so far. It has been revealed that sensible temperature is mainly affected by weather conditions such as relative humidity and wind-speed. Missenard's effective temperature, *NET* (Net Effective Temperature), and Discomfort Index are popular examples based on these viewpoints. (Missenard, 1959; Hentschel, 1987; Thom, E. C., 1959). However, the sensible temperature is affected by not only weather conditions but also human conditions such as amount of wearing clothes and metabolic rate. *PMV* (Predicted Mean Vote) and *SET* (Standard Effective Temperature) are the popular indices which consider human conditions (Fanger, 1972; Gagge et al., 1986; Matzarakis, 2001; Spagnolo and de Dear, 2003; Tinz and Jendritzky, 2003).

From the above discussion, it is desirable to create the sensible temperature index which can be useful for demand forecasting. Our study consists of two parts; the first one is to quantify the sensible temperature and the other is to predict it so as to utilize it for demand forecasting. In this paper, we propose the method of quantifying sensible temperature by analyzing digital text data of Twitter. We realize the modeling of temperature sensation of public consumers by utilizing Tweet data. We discuss the effectiveness of our proposed method for the sensible temperature quantification and try to reveal factors that influence sensible temperature.

2. PRE-ANALYSIS

In this study, we deal with both Japanese Tweet data with location information (1/10 sampling) and weather data (AMeDAS Tokyo). The digital text data of Tweet is seemed to have various characteristics which can't be seen in other text data. So we should grasp these characteristics and consider them adequately for later analysis.

About the weather data, we checked seasonal fluctuations of each weather element. Finally, to verify the validity of utilizing Tweet data for this study, we take up factors which can be related to temperature sensation within Tweet data and compared it with weather data.

2.1 Tweet Data

The observed period of Tweet data is from September 1st, 2012 to September 30th, 2015 and total number of tweets is 15,495,108. They include the date and time which are posted and the information of *latitude* and *longitude*. Figure 1 shows the transition of the number of tweets per day. It shows that the number of tweets increases gradually and falls sharply in May, 2015.

Next, as a possible factor which influences the fluctuation of the number of tweets, we take up the day of week. Daily factor like day of week can be influential factor. Average number of tweets of each day of week is shown in Figure 2.

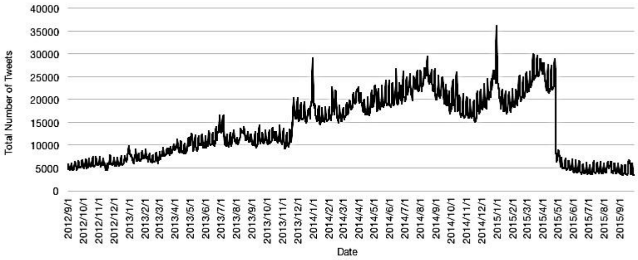


Figure 1: Transition of the number of tweets

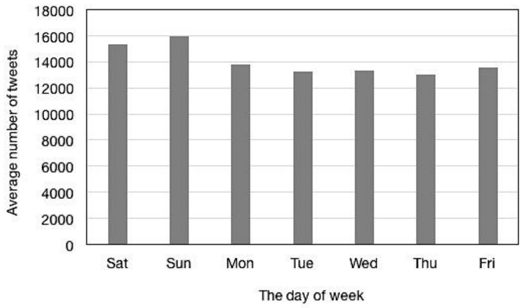


Figure 2: Average number of tweets of each day of week

According to this figure, the averages of tweets posted on weekend such as Saturday and Sunday are likely to be larger than those of weekday. This tendency indicates the fact that the number of tweets is in proportion to the length of spare time. Besides, the reason why the number of Sunday's tweet is larger than that of Saturday's may be that the rate of user who is off in Sunday is higher than that of Saturday. From this analysis,

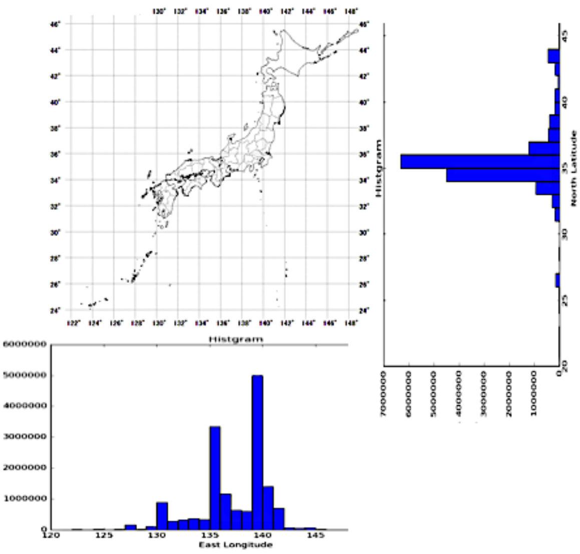


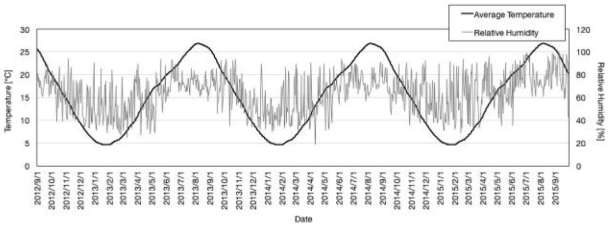
Figure 3. The geographical distribution of the number of tweets. (The source of map of japan is *Geospatial Information Authority of Japan*)

for appropriate analysis, it is necessary to eliminate "the day of week" effect when we utilize factors which are related to the number of tweets. In this study, we apply ratio-to-moving-average method to do so (Tseng et al., 2001).

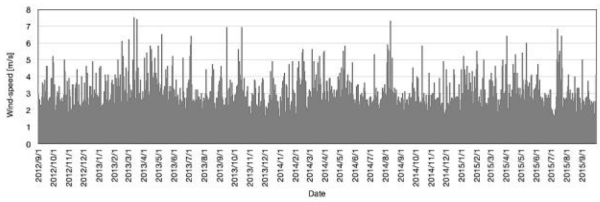
Finally, we analyze the geographical distribution of the number of tweets (shown in Figure 3). It shows that the number of tweets in the center of the Kanto region (e.g. Tokyo) and the Kansai region (e.g. Osaka) have the majority. This tendency is thought to result from concentration of population and abundance of active users of Twitter in urban areas.

2.2 Weather Data

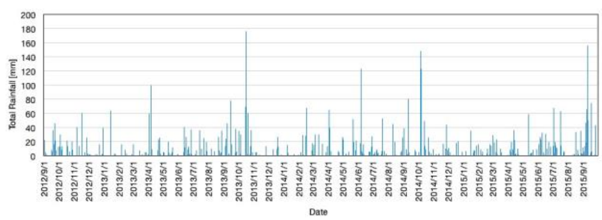
This weather data is observed at AMeDAS Tokyo (located in Tokyo) and its period is the same as that of Tweet data. They include weather elements such as Average Temperature T[℃], Minimum Temperature $T^{\min}$ [℃], Maximum Temperature $T^{\max}$[℃], Relative Humidity H[%], Wind-speed W[m/s], Total Rainfall TRF[mm], Maximum Snowfall MSF[cm] Solar Time ST[hour] and Total Solar Radiation TSR[MJ/m^2]. Hereinafter, weather element with the subscript t denotes the weather element of term t.



(*a*) Average Temperature and Relative Humidity



(*b*) Wind-speed



(*c*) Total Rainfall

Figure 4: Transition of Average Temperature, Wind-speed, Relative Humidity and Total Rainfall

Among weather elements which we state previously, we take up Average Temperature, Relative Humidity, Wind-speed and Total Rainfall and visualize the transitions of them in Figure 4.

In Figure 4(*a*), the time series of Average Temperature has cyclic ups and downs periodically. In addition, the Relative Humidity repeats the up-and-down little by little and the degree of its movement is smaller within the period from June to September than other periods. Figure 4(*b*) shows the tendency that strong wind is likely to arise in August, September and October and occasionally in other periods. Figure 4(*c*) shows that Total Rainfall is likely to be high in October. In addition, it snows only in January or February in observed regions though it is not indicated in Figure 4. The transitions of $T^{\min}$, $T^{\max}$, ST and TSR have similar tendency to the Average Temperature T.

2.3 Comparison of Tweet Data and Weather Data

In order to utilize tweet data for quantifying temperature sensation, we apply the morphological analysis to Tweet data and calculate the rate of tweets including the words "*hot*" and "*cold*" per day respectively. The relationship between each rate and Average Temperature is shown in Figure 5.

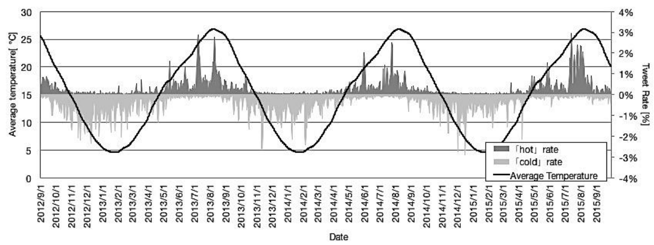


Figure 5: Transitions of Average Temperature and *hot*/*cold tweet* rate (plus area is *hot's* and minus area is *cold's*)

Figure 5 shows that there is a measurable correlation between the Average Temperature and the *hot*/*cold tweet* rate, which declare the validity of utilizing *hot*/*cold tweet* rate for quantifying temperature sensation. On the other hand, we can find several days with a gap between Average Temperature and *hot*/*cold tweet* rate. On these dates, it can be considered that the temperature sensation felt by people deviates from one based on Average Temperature. In next section, we assume that *hot*/*cold tweets* are generated by the basis of sensible temperature and give the definition of sensible temperature by the *hot*/*cold tweet* rate.

3. PRELIMINARIES

In this section, we explain two types of conventional prediction models which are important for our study. One is a method which tries to reveal how the inputs affect the output. For this purpose, Linear Regression Model and Logistic Regression Model (Trevor, 2009) can be applied. The other is time series analysis such as Autoregressive model (Shumway, 2010).

3.1 Linear Regression (*Lin*R) Model

Linear Regression Models are used to predict one dependent variable from some given independent variables. Let $y_i \in \mathbb{R}$ be the dependent variable of the i-th case and $\boldsymbol{x}_i \in \mathbb{R}^{d+1}$ be the corresponding independent variable vector (i = 1, 2, …, n). The regression model is indicated below.

$$y_i = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i + \varepsilon_i \tag{1}$$

$$\varepsilon_i \sim N(0,\sigma^2) \tag{2}$$

Here $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_d)^{\mathrm{T}}$ is the partial regression coefficients and ε_i is the error term which follows a normal distribution with the mean equals to 0 and the variance equals to σ^2. When $d \geq 2$, it is called *multiple linear regression* and in case of $d = 1$, it is called *simple linear regression*. Partial regression coefficient vector $\boldsymbol{\beta}$ is mainly estimated based on least squares method. This is the method to determine parameters so as to minimize the summation of square errors. Optimization problem is as follows:

$$\min_{\boldsymbol{\beta}} \quad \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i\right)^2 \tag{3}$$

3.2 Logistic Regression Model

Logistic Regression Model is the appropriate regression model to conduct when the dependent variable is binary or rate value. Let y_i be the dependent variable of the i-th case and $\boldsymbol{x}_i$ be the corresponding independent variable. The model has the form

$$\operatorname{logit}(y_i) = \log\left(\frac{y_i}{1-y_i}\right) = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{x}_i \tag{4}$$

$\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_d)$ is a partial regression coefficient vector. The parameter $\boldsymbol{\alpha}$ is usually estimated by the maximum log-likelihood estimator.

3.3 Autoregressive(AR) Model

The autoregressive model is the most basic model for time-series analysis. This model is based on the idea that the current value of series can be expressed as a function of some past values. Given time-series data $y_1, y_2, \ldots, y_n$, the autoregressive model is expressed as follows:

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + v_t \tag{5}$$

Here, p defines the parameter dimension which determine how many previous terms are considered and corresponding model is expressed as AR(p). p is likely to be determined based on AIC criteria. The parameter a_i is the autoregressive coefficient and v_t is a white noise which follows a normal distribution with the mean 0 and the variance σ_v^2. The parameter a_i is determined based on the maximum likelihood estimation.

3.4 Vector Autoregressive(VAR) Model

The Vector Autoregressive (VAR) model is an extended model of AR model to multivariate data. Let $\boldsymbol{Y}_t \in \mathbb{R}^D$ be the D dimensional data of term t. Then, given time-series data $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n$, the Vector Autoregressive model is given as follows:

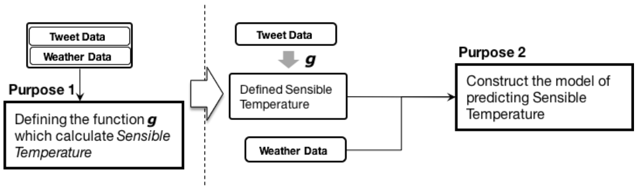
$$\boldsymbol{Y}_t = \sum_{i=1}^{p'} \boldsymbol{A}_i \boldsymbol{Y}_{t-i} + \boldsymbol{e}_t \tag{6}$$

The matrix $\boldsymbol{A}_i \in \mathbb{R}^{D\times D}$ is the autoregressive coefficients matrix. $\boldsymbol{e}_t$ is D dimensional white noise which follows a normal distribution with mean $\mathbf{0}_D$ (zero vector in D-dimensions) and variance-covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{D\times D}$.

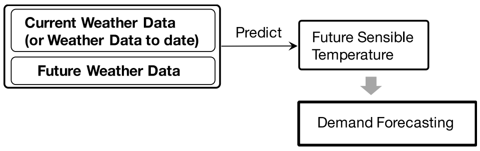
In case of conducting time-series analysis, we should confirm the stationarity of the data. The stationarity of a time series data is the characteristic that probability distribution doesn't change when shifted in time and the required factor when applying time-series analysis. In brief, the parameters such as the mean and variance don't change and don't follow any trend over time. If a time-series data doesn't satisfy the stationarity, it needs to take appropriate preprocessing such as differential conversion and log conversion so as to satisfy stationarity (Shumway, 2010).

4. UTLIZING THE SENSIBLE TEMPERATURE

To utilize the sensible temperature for demand forecasting, it is necessary to quantify the sensible temperature felt by consumers. In this study, we realize the modeling of temperature sensation by utilizing Tweet data. Besides, looking towards demand forecasting, the prediction of future sensible temperature is necessary. For example, if the lead-time is one day, the shop keeper should predict the tomorrow demand by using the predicted sensible temperature and make an order. The predicting experiment of sensible temperature is, therefore, conducted in section 4.2. Perspective of our study is illustrated in Figure 6.



(a) Purposes of this study



(b) Practical usage

Figure 6: Perspective of this study

4.1 Quantifying the Sensible Temperature

In this section, we will explain the method of quantifying sensible temperature. In this method, we utilize the *hot/cold tweet* rate (stated at section 2.3) and the observed average temperature.

4.1.1 Overview of Quantifying Method

Let r_t^{hot} and r_t^{cold} be the rate of tweets including the word "*hot*" and "*cold*" at the time t (day) respectively. Besides, let N be the number of terms (t = 1, 2, …, N). Then, the *hot* tweet rate r_t^{hot} can be regarded as the index showing the degree of temperature sensation of hot that people are feeling and the cold tweet rate r_t^{cold} is the same for cold sensation. Therefore, the relation model between the rate $r_t^{hot}(r_t^{cold})$ and the average temperature T_t can lead to quantifying the sensible temperature.

To quantify the relationship between r_t^{hot} and T_t, we use the regression model. In this method, we set r_t^{hot} as the dependent variable and T_t as the independent variable though the opposite case can be considered. This setting is based on the assumption that the hot rate r_t^{hot} includes a specific noise and has a mean conditioned by T_t. By constructing this regression model, we can obtain the regression formula $\hat{r}_t^{hot} = f(T_t|\hat{\alpha}_0, \hat{\alpha}_1)$. Then, f can be regarded as the function that can calculate an expected value of r_t^{hot} conditioned by the temperature T_t. Here, the inverse function $g(r_t^{hot}|\hat{\alpha}_0,\hat{\alpha}_1) = f^{-1}(T_t|\hat{\alpha}_0,\hat{\alpha}_1)$ can be defined. Therefore, $g(r_t^{hot}|\hat{\alpha}_0,\hat{\alpha}_1)$, which is obtained by conducting the formula deformation so as to form like " $T_t =$ ", can be regarded as the function to calculate the most common value of temperature when the value of r_t^{hot} is given. Since the temperature calculated by the function g is the reasonable value based on people' sensation, we define this calculated temperature as *sensible temperature* based on hot sensation.

Though we only explained the case of r_t^{hot} so far, the same processing can be conducted for r_t^{cold}. Namely, two sensible temperatures can be obtained per t. Let S_t^{hot} be the sensible temperature calculated from r_t^{hot} and S_t^{cold} be one based on r_t^{cold}. The ultimate sensible temperature S_t is calculated by the weighted average of S_t^{hot} and S_t^{cold}. The schematic illustration of this method is shown in Figure 7.

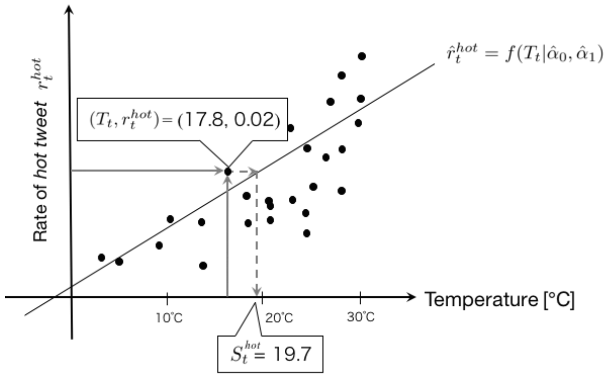


Figure 7: Schematic illustration of quantifying the sensible temperature

4.1.2 Algorithm

In this subsection, we state the algorithm of quantifying the sensible temperature. Since the dependent variable is rate value, we select the logistic regression model to obtain formula which expresses relations between the tweet rate and the temperature.

Step1) Estimation of Coefficient

The partial regression coefficients of logistic regression model defined as Eq. (7) is estimated.

$$\mathrm{logit}(\hat{r}_t^{hot}) = \log\left(\frac{\hat{r}_t^{hot}}{1-\hat{r}_t^{hot}}\right) = \hat{\alpha}_0 + \hat{\alpha}_1 T_t \qquad (7)$$

Step2) Obtaining the Transformation Equation

By utilizing the partial regression coefficients which are obtained at **Step1**, r_t^{hot} can be converted into S_t^{hot}.

$$S_t^{hot} = g\left(r_t^{hot} | \hat{\alpha}_0, \hat{\alpha}_1\right) \tag{8}$$

$$g\left(r_t^{hot} | \hat{\alpha}_0, \hat{\alpha}_1\right) = \frac{1}{\hat{\alpha}_1} \log \frac{r_t^{hot}}{e^{\hat{\alpha}_0}(1 - r_t^{hot})} \tag{9}$$

Step3)

Step1-2 are conducted for r_t^{cold} and S_t^{cold} is obtained.

Step4) Calculation of Sensible Temperature

Sensible temperature S_t is calculated in the following formula.

$$S_t = \frac{r_t^{hot}}{r_t^{hot} + r_t^{cold}} S_t^{hot} + \frac{r_t^{cold}}{r_t^{hot} + r_t^{cold}} S_t^{cold} \tag{10}$$

4.2 Predicting the Sensible Temperature

In order to utilize the sensible temperature for demand forecasting, the future value of sensible temperature is required when an order should be made. Therefore, it is necessary to predict the sensible temperature. In this section, we conduct prediction experiment by applying existing prediction methods. Besides, we try to reveal factors which influence sensible temperature. For this analysis, we set four prediction patterns indicated in **Table 1**. Hereinafter, let $\boldsymbol{w}_t \in \mathbb{R}^V$ be the weather elements vector of the term t and $\boldsymbol{w}_t^f \in \mathbb{R}^V$ be the estimated value of $\boldsymbol{w}_t$ by weather forecast (elements will be described later). $\hat{\boldsymbol{w}}_t$ in Table 1 is the predicted weather elements vector of term t obtained by applying VAR model.

Table 1. Prediction Patterns

Pattern	Independent Variables	Dependent Variables	Applied Method
1	$\boldsymbol{w}_t^f$	$\hat{S}_t$	*Lin*R
2	$\hat{S}_{t-1}$ predicted from $\boldsymbol{w}_{t-1}$		*Lin*R + AR
3	Combined Vector $(\hat{S}_{t-1}, \boldsymbol{w}_{t-1})$	$(\hat{S}_t, \hat{\boldsymbol{w}}_t)$	*Lin*R + VAR
4	$\tilde{S}_t$ (predicted from $\hat{S}_{t-1}$) and $\boldsymbol{w}_t^f$	$\hat{S}_t$	*Lin*R + AR + *Lin*R

To predict the sensible temperature of the next day, it is not appropriate to utilize the current actual sensible temperature. This is because it is necessary to calculate it by using huge tweet day every day. Considering the practical use, it is not reasonable to take a labor of counting the number of tweets every day. Then, we construct a model to predict the sensible temperature without any actual value but in the form of an estimated value for prediction.

To conduct the prediction experiment based on four patterns indicated in Table 1, we select existing prediction methods for each pattern. In *Pattern* 1, the multiple linear regression method is applied. In *Pattern* 2, after the multiple regression at the term (t-1), the AR model is applied in order to predict the sensible temperature of the next day from the previous day. In *Pattern* 3, we apply the VAR model whose observations are vectors composed of the estimated sensible temperature and weather conditions at the term t. Finally, *Pattern* 4 includes the additional multiple regression whose independent variables vector is the combined form such as $(\tilde{S}_t, \boldsymbol{w}_t^f)$.

5. EXPERIMENT

We apply our method of quantifying the sensible temperature to actual data and try to verify its validity. Besides, we try to extract factors which would influence the sensible temperature through prediction experiment.

5.1 Applying the Quantifying Algorithm

First, we apply the proposed algorithm of quantifying the sensible temperature that stated in chapter 4 to real data. The applied period is from September 1st, 2012 to September 30th, 2015 (1125 days). Before the analysis, we eliminate "the day of the week" effect by applying ration-to-moving-average method to *hot/cold tweet* rate (Tseng et al., 2001).

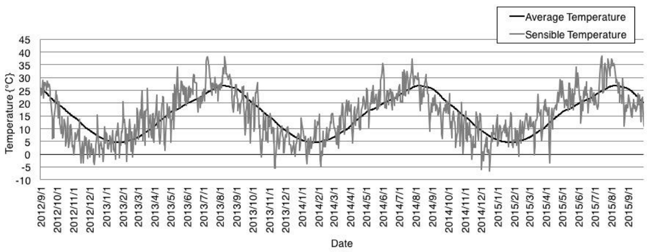


Figure 8: Transitions of S_t and T_t

Figure 8 shows the calculated sensible temperature and the observed average temperature. It shows that though rough tendency of the sensible temperature is similar to that of the average temperature, the variability of sensible temperature is much larger than that of average temperature. Besides, there exist dates whose gaps between average temperature and sensible temperature are extremely large. There is room for further consideration about this result. In addition, to analyze relationship between the average temperature and the sensible temperature per month, we visualize the mean values and the standard deviations of them in Figure 9.

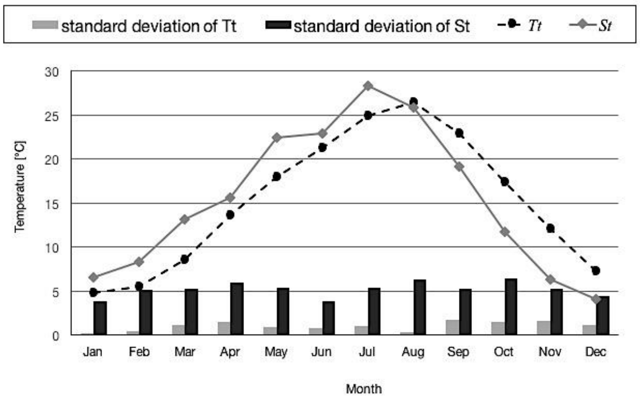


Figure 9: S_t and T_t per month

From Figure 9, we can identify that the sensible temperature tends to be higher than the average temperature from January to July and lower from August to December. Besides, especially in March and May (easy to change temperature in short term), July (temperature rises sharply) and October (temperature falls sharply), the gap between the average temperature and the sensible temperature is large. This result suggests the measurable effect of sharp changes in temperature. In addition, the standard deviation of the sensible temperature is much larger than that of the average temperature especially in February and August, which imply the effect of other factors in these months.

5.2 Prediction Experiment

We state the prediction of the sensible temperature. The assumed situations are the same as indicated in the section 4.2. There are two purposes of this experiment. One is to utilize the future value of sensible temperature for demand forecasting. The other is to identify factors which would influence the sensible temperature and its validities.

5.2.1 Experimental Conditions

We set the data during the period from September 1, 2012 to September 30, 2014 as training data (N = 760) and the data from October 1, 2014 to September 30, 2015 as test data (N_{test} = 365). In this experiment, the unit of term (i.e. t) is denoted that to correspond one day. Evaluation indices are the mean average error (MAE) and coefficient of determination (expressed as R^2). These are calculated in Eq. (11) and (12).

The independent variables of first linear regression model in all patterns are T_t, T_t - $T_{t\text{-}1}$, H_t, W_t, TRF_t, TSR_t, DV_t^1, DV_t^2, DV_t^3 and DV_t^4, whose DV_t^1, DV_t^2, DV_t^3 and DV_t^4 are the dummy variables for T_t which conform to the rules: $DV_t^1 = 1$ if $T_t \geq 25$, otherwise $DV_t^1 = 0$ (and so are others), $DV_t^2 = 1$ if $25 > T_t \geq 20$, $DV_t^3 = 1$ if $20 > T_t \geq 15$, and $DV_t^4 = 1$ if $15 > T_t \geq 10$.

$$MAE = \frac{1}{N_{test}} \sum_{t=1}^{N_{test}} |\hat{S}_t - S_t| \tag{11}$$

$$R^2 = \frac{\left(\sum_{t=1}^{N_{test}} (y_t - \bar{y})\left(\hat{y}_t - \bar{\hat{y}}\right)\right)^2}{\sum_{t=1}^{N_{test}} (y_t - \bar{y})^2 \sum_{t=1}^{N_{test}} \left(\hat{y}_t - \bar{\hat{y}}\right)^2} \tag{12}$$

The dimension parameters p for the autoregressive model is set to 1. To convert into stationary data, we apply difference conversion and the neglog transformation to the sensible temperature. For the linear regression after AR model in *Pattern* 4 (indicated in **Table 1**), we select 9 variables: $\tilde{S}_t$, H_t, W_t, TRF_t , TSR_t, DV_t^1, DV_t^2, DV_t^3, DV_t^4.

5.2.2 Results and Considerations

Table 2. Relationship between parameter p and evaluation indices (MAE and R^2)

Pattern	*MAE*	R^2
1	3.532	0.779
2	4.033	0.706
3	3.950	0.714
4	3.455	0.786

Table 2 shows the result of each prediction pattern. By comparison between *Pattern* 1 and *Pattern* 2, it turns out that applying the autoregressive model to estimated value leads to deterioration of performance. Moreover, the weather elements may affect the sensible temperature of the next day directly by comparison with *Pattern* 2 and 3. Besides, utilizing the weather data of the same day lead to improvement of the prediction performance (by comparison with *Pattern* 2 and 4). Nevertheless, in this experiment, the weather data of the same day $\boldsymbol{w}_t^f$ is equal to $\boldsymbol{w}_t$. So, for more strict analysis, it is desirable to conduct an experiment with actual forecasted value.

Table 3. Parameter p and MAE

p	*MAE*		
	Pattern 2	*Pattern* 3	*Pattern* 4
1	4.0332	3.9469	3.4553
2	4.0061	3.9050	**3.4431**
3	**4.0024**	**3.9040**	3.4464
4	4.0322	3.9179	3.4547
5	4.0282	3.9200	3.4474
6	4.0133	3.9339	3.4508
7	4.0359	3.9353	3.4647

Table 3 is the result of performance of each value of the dimension parameter p of the AR model. Here, note that the result of *Pattern* 4 in Table 3 is obtained by using the result of *Pattern* 2. It shows that the autoregressive model performs best in case of $p = 2$ or 3 for prediction of the sensible temperature. Namely, the weather data from three days ago to the previous day may affect the sensible temperature of the predicted day (*Pattern* 2 and 3). In addition, the weather data of the past two days and that of predicted day mainly affect the sensible temperature of the predicted day (*Pattern* 4).

5.2.3 Abstracting the Influential Factors

In order to reveal factors which are especially influential and how it affects, we indicate the partial coefficients of *Pattern* 4 (best performance pattern).

Table 4. Standardized partial regression coefficients of *Pattern* 4

Intercept	$\tilde{S}_t$	H	W	TRF	TSR	DV_t^1	DV_t^2	DV_t^3	DV_t^4
15.703	4.614	3.393	-0.440	-0.371	3.962	0.571	-0.196	-1.053	-1.691

Table 4 shows that $\tilde{S}_t$, Relative Humidity H and Total Solar Radiation TSR are effective factors from the largeness of its absolute value. In addition, the fact that the coefficient of Wind-speed is negative value and that of Relative Humidity is positive value could indicate the consistency with the facts revealed in previous studies.

In a previous study in Japan, it is well known that 1[m/s] down of Wind-speed leads to 1[℃] down of the sensible temperature (Kamiyama, 1961). Missenard suggested that the relative humidity rises the sensible temperature in case that the temperature is 10 or more. The validity of the degree of effect of each element should be considered more.

6. Conclusion and Future Works

In this study, we proposed the method of quantifying the sensible temperature and discuss its validity. In addition, we try to predict the sensible temperature by using existing prediction methods. From now on, we should verify the validity of the sensible temperature through experiments of demand forecasting with sensible temperature data. Besides, it is necessary to take account of differences between regions and past values of weather data.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to all members of Goto laboratory, Waseda University, and members of Japan Weather Association who support us for their helpful comments in this research. A part of this study was supported by JSPS KAKENHI Grant Numbers 26282090 and 26560167.

REFERENCES

Fanger, P. O. (1972) Thermal comfort. McGraw-Hill, New York

Gagge, A. P., Fobelets, A. P., Bergluad, L. G. (1986) A standard predictive index of human response to the thermal environment. *ASHRAE Trans* 92, 709–731.

Gregorczuk, M., Cena, K. (1967) Distribution of Effective Temperature over the surface of the Earth, pp. 111–123.

Hentschel, G. (1987) A human biometeorology classification of climate for large and local scales. In *Proc. WMO/HMO/UNEP Symposium on Climate and Human Health, Leningrad 1986*, Vol. I, WCPA - No. 1, WMO.

Honjo, T. (2009) Thermal comfort in outdoor environment, Global environmental research 13.2009: 43-47.

Kamiyama, K. (1961) Sensible Temperature. *Kisho-Kenkyu Note* (*The Meteorological Society of Japan*) vol. 12(3), 214-248. (in Japanese)

Matzarakis, A. (2001) Die thermische Komponente des Stadtklimas, Wiss. Ber. Meteorologisches Institut der Universität Freiburg No. 6.

Missenard, A. (1959) On thermally equivalent environments. JIHVE (Building Services: now the Journal of CIBSE), UK, vol 27, pp 231-237.

Shumway, R. H. and Stoffer, D. S. (2010). Time series analysis and its applications: with R examples. Springer Science & Business Media.

Spagnolo, J., de Dear, R. (2003) A field study of thermal comfort in outdoor and semi-outdoor environments in subtropical Sydney Australia. Build Environ 38, 721–738.

Taylor, J. W., & Buizza, R. (2003). Using weather ensemble predictions in electricity demand forecasting. International Journal of Forecasting, 19(1), 57-70.

Thom, E. C. (1959). The discomfort index. Weatherwise, 12(2), 57-61.

Tinz, B., Jendritzky, G. (2003) Europa- und Weltkarten der gefühlten Temperatur. In: Chmielewski, F.-M., Foken, Th. (eds.) Beiträge zur Klima- und Meeresforschung, Berlin und Bayreuth, 111-123.

Tseng, F. M., Yu, H. C., & Tzeng, G. H. (2001). Applied hybrid grey model to forecast seasonal time series. Technological Forecasting and Social Change, 67(2), 291-302.

Trevor, H., Robert, T. and Jerome, F. (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction.