Secure Spaces and Spatio-Temporal Weblog Sensors with Temporal Shift and Propagation

Shun Hattori
School of Computer Science
Tokyo University of Technology
1404-1 Katakura, Hachioji, Tokyo 192-0982, Japan
Email: hattori@cs.teu.ac.jp

***Abstract*—Many researches on mining the Web, especially CGM (Consumer Generated Media) such as Weblogs, for knowledge about various phenomena and events in the real world have been done actively, and Web services with the Web-mined knowledge have begun to be developed for the public. However, there is no detailed investigation on how accurately Web-mined data reflect real-world data. It must be problematic to idolatrously utilize the Web-mined data in public Web services without ensuring their accuracy sufficiently. Therefore, this paper defines the basic Weblog Sensor, the temporal-shifted Weblog Sensors, and the temporal-propagated Weblog Sensors, and tries to validate the potential and reliability of these Weblog Sensors' spatio-temporal data by measuring the correlation with weather (precipitation and temperature) statistics of Japan Meteorological Agency as real-world data.**

***Keywords*-Web mining; Web credibility; Web sensors; knowledge extraction; information credibility; search engine indices;**

I. Introduction

In recent years, how to make physical spaces smarter has become one of the hottest topics in the research field of ubiquitous/pervasive computing. Smart Spaces [1] are often physically isolated environments such as rooms, which are made smart by various information communication technologies. They would be much more convenient for information access in the future. Meanwhile, information security has also become very significant in any situation, especially in public places such as indoor work places, educational facilities, healthcare centers and so on. The amount of physical or virtual information resources which should be protected in the physical world grows exponentially.

Physical environments are becoming smart but not always secure. When a virtual (computational) information resource is requested to access by a user via an output device, conventional access control systems make a decision on whether the user should be granted or denied to access the resource based on its access policies and surely enforce the access decision. However, even if the requester is authorized by it, it should not be immediately offered to her via the output device, because there might be its unauthorized users as well as the authorized requester around the output device, especially in public places. A user trying to visit a physical environment might in turn be unexpectedly exposedto her unwanted information access. For example, although she does not want to know about the results of a football game that she had recorded on video to watch later, she unfortunately encounters it in her train. Meanwhile, when a user enters a physical environment, the user might hate its real characteristics (e.g., degrees of dismal and danger) and/or be forced to access her unwanted information resources unexpectedly. This paper proposes a method to extract information for making access or entry decisions in Secure Spaces from very large text corpora such as the Web, especially the Weblog, and also improve my Secure Spaces [1] by adding the concept of the Weblog Sensors, in order to enable users to specify their access policies by keyword-based expressions about their unwanted physical spaces.

The former Web world did not have a familiar relationship with the real world, and it is not too much to say that the former Web world was isolated and independent of the real world. But in recent years, the explosively-growing Web has had more and more familiar relationship with the real world as the use of the Web, especially CGM (Consumer Generated Media) such as Weblogs, WOM (Word of Mouth) sites, SNSs (Social Networking Services), has become more popular with various people without distinction of age/sex.

Many researches on mining the Web, especially CGM (Consumer Generated Media) such as Weblogs, for knowledge about various phenomena and events in the real world have been done very actively. For example, opinion and reputation extraction [2], [3] of various products and services provided in the real world, experience mining [4], [5] of various phenomena and events held in the real world, and concept hierarchy (semantics) extraction such as is-a/has-a relationships [6], [7] and appearance (look and feel) extraction [8], [9] of real objects in the real world. Meanwhile, Web services with the Web-mined knowledge have begun to be developed for the public, and more and more ordinary people actually utilize them as information for choosing better products, services, and actions in the real world.

However, there is no detailed investigation on how accurately Web-mined data about a phenomenon or event held in the real world reflect real-world data. It is not difficult for us to extract some kind of the potential knowledge data

978-1-4244-8581-9/11/$26.00 ©2011 IEEE

from the Web by using various text mining techniques, and it might be not problematic just to enjoy browsing them. But while choosing better products, services, and actions in the real world, it must be problematic to idolatrously utilize the Web-mined data in public Web services without ensuring their accuracy sufficiently.

This paper defines three kinds of Weblog Sensors [10] to mine the Web, especially CGM such as Weblog documents for spatio-temporal data about a target phenomenon (e.g., precipitation and temperature) in the physical world:

- the basic Weblog Sensor $\mathrm{ws}(s, t, kw)$ based on the frequency of Weblog documents searched by a keyword kw (e.g., a Japanese noun "ame" and a Japanese adjective "atsui") representing the target phenomenon and restricted to a space s (e.g., 47 prefectural capitals) and a time period t (e.g., a day or a month) with a basic assumption that almost all Weblog documents described at a time period t are about the present period t,
- the temporal-shifted Weblog Sensors $\mathrm{ws}_d(s, t, kw)$ with a temporal shift parameter d (days) to utilize Weblog documents described at a time period t about not the present period t but the past or the future period $t + d$,
- the temporal-propagated Weblog Sensors $\mathrm{ws}^{\sigma}(s, t, kw)$ with a single temporal propagation parameter σ, and $\mathrm{ws}^{\sigma_b, \sigma_a}(s, t, kw)$ with double temporal propagation parameters σ_b before a time period t and σ_a after the time period as a standard deviation of the Normal Distribution to utilize Weblog documents described about not only the present but also the past or the future.

And then this paper tries to validate the potential and reliability of these Weblog Sensors' spatio-temporal data by measuring the (average) correlation coefficient with weather (precipitation and temperature) statistics of JMA (Japan Meteorological Agency) [11] as real-world data.

II. Secure Spaces

To build Secure Spaces in the real world by using space entry control based on their dynamically changing contents such as their visitors, physical/virtual information resources via their embedded output devices, each Secure Space requires the following facilities (as shown in Figure 1).

- **Space Management**: is responsible for managing a Secure Space, i.e., for constantly figuring out its contents such as its visitors, its embedded physical information resources and virtual information resources outputted via its embedded output devices and also for ad-hoc making an authorization decision on whether an entry request to enter the Secure Space by a visitor or a physical/virtual information resource should be granted or denied, and for notifying the entry decisions to the Electrically Lockable Doors or enforcing entry control over virtual information resources according to the entry decisions by itself.
- **User/Object Authentication**: is responsible for authenticating what physical entity such as a user or a physical information resource requests to enter or exit the Secure Space (e.g., by using Radio Frequency IDentification or biometrics technologies) and also for notifying it to the Space Management.
- **Electrically Lockable Door**: is responsible for electrically locking or unlocking itself, i.e., for assuredly enforcing entry control over physical entities such as users and physical information resources, according to instructions by the Space Management.
- **Physically Isolating Opaque Wall**: is responsible for physically isolating inside a Secure Space from outside there with regard to information access, i.e., for validating the basic assumption that any user inside a Secure Space can access any resource inside the Secure Space while any user outside the Secure Space can never any resource inside the Secure Space.

To protect us from our unwanted characteristics of physical spaces as well as our unauthorized contents, the following additional facilities are required.

- **Real Sensor**: is responsible for physically sensing inside a Secure Space for its real characteristics to make access decisions in the Secure Space and also for notifying the sensor data stream to the Space Management. For example, thermometers, hygrometers, (security) cameras.
- **Weblog Sensor**: is responsible for logically sensing the Weblog for the approximate characteristics of each Secure Space to make access decisions in the Secure Space and also for notifying the Web-mined data to the Space Management. Note that any Secure Space does not have to equip the extra devices unlike Real Sensors.

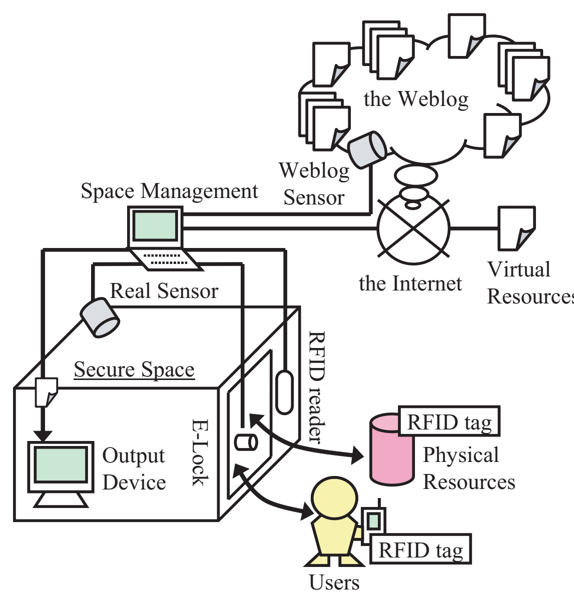

Figure 1. Secure Spaces

III. Method

This section constructs three kinds of Weblog Sensors to mine the Web, especially Weblog documents for spatio-temporal data about a target phenomenon in the physical world (e.g., “precipitation” and “temperature”). First, I define the basic Weblog Sensor with a geographical space s (e.g., 47 prefectural capitals in Japan such as “Tokyo” and “Kyoto”), a time period t (e.g., a day such as “2007/1/1” and “2009/12/31”), and a keyword kw (e.g., “ame” and “atsui” in Japanese that mean “rain” and “hot” in English respectively) for a target phenomenon. Next, I define the temporal-shifted Weblog Sensors with a temporal shift parameter d (days). Last, I define the temporal-propagated Weblog Sensors with a temporal propagation parameter σ or with temporal propagation parameters σ_b and σ_a.

First, the value $\mathrm{ws}(s, t, kw)$ indicated by the basic Weblog Sensor with a geographical space s, a time period t, and a keyword kw for a target phenomenon is defined as follows:

$$\mathrm{ws}(s, t, kw) := \mathrm{bf}_t(\texttt{[}s \texttt{ AND } kw\texttt{]}),$$

where $\mathrm{bf}_t(\texttt{[}q\texttt{]})$ stands for the Frequency of weBlog documents searched by submitting the query q with the custom time range t to Google Blog Search [12].

Next, the value $\mathrm{ws}_d(s, t, kw)$ indicated by the temporal-shifted Weblog Sensor with a temporal shift parameter d (days), a geographical space s, a time period t, and a keyword kw for a target phenomenon is defined as follows:

$$\mathrm{ws}_d(s, t, kw) := \mathrm{ws}(s, t + d, kw).$$

Last, the value $\mathrm{ws}^\sigma(s, t, kw)$ indicated by the temporal-propagated Weblog Sensor with a temporal propagation parameter σ, a geographical space s, a time period t, and a keyword kw for a target phenomenon is defined as follows:

$$\begin{aligned}
\mathrm{ws}^\sigma(s, t, kw) &:= \sum_{\forall d} \mathrm{ws}_d(s, t, kw) \cdot p^\sigma(d),\\
p^\sigma(d) &:= N(0, \sigma, d),\\
N(\mu, \sigma, d) &:= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(d-\mu)^2}{2\sigma^2}\right),
\end{aligned}$$

where $N(\mu, \sigma, d)$ stands for the Normal Distribution with an average μ and a standard deviation σ. In the below-mentioned experiment, $\forall d$ is restricted to $[-30, 30]$.

The another value $\mathrm{ws}^{\sigma_b,\sigma_a}(s, t, kw)$ indicated by the temporal-propagated Weblog Sensor with a temporal propagation parameter σ_b before a time period t, the other σ_a after the time period t, a geographical space s, and a keyword kw for a target phenomenon is defined as follows:

$$\begin{aligned}
\mathrm{ws}^{\sigma_b,\sigma_a}(s, t, kw) := &\sum_{\forall d<0} \mathrm{ws}_d(s, t, kw) \cdot p^{\sigma_b}(d)\\
&+ \mathrm{ws}_0(s, t, kw) \cdot \frac{p^{\sigma_b}(0) + p^{\sigma_a}(0)}{2}\\
&+ \sum_{\forall d>0} \mathrm{ws}_d(s, t, kw) \cdot p^{\sigma_a}(d).
\end{aligned}$$

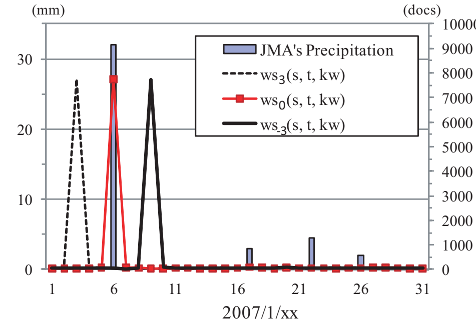

Figure 2. JMA’s Precipitation and Temporal-Shifted Weblog Sensors $\mathrm{ws}_d(s, t, kw)$, s=“Tokyo”, kw=“ame” (“rain” in English).

Figure 2 shows the JMA’s precipitation, the basic Weblog Sensor $\mathrm{ws}_0(s, t, kw) = \mathrm{ws}(s, t, kw)$, the temporal-shifted Weblog Sensors $\mathrm{ws}_{-3}(s, t, kw)$ and $\mathrm{ws}_3(s, t, kw)$ using a Japanese noun “ame” (“rain” in English) as a representation of the target phenomenon “precipitation” in s = “Tokyo” on t = January 1st to 31st, 2007.

Figure 3 shows examples of the probability $p^\sigma(d)$ with a temporal propagation parameter σ, or temporal propagation parameters σ_b before a time period and σ_a after the time period for the temporal-propagated Weblog Sensors.

Figure 4 shows the JMA’s temperature, the basic Weblog Sensor $\mathrm{ws}(s, t, kw)$, the temporal-propagated Weblog Sensors $\mathrm{ws}^1(s, t, kw) = \mathrm{ws}^{1,1}(s, t, kw)$ and $\mathrm{ws}^{15,0.5}(s, t, kw)$ using a Japanese adjective “atsui” (“hot” in English) as a representation of the target phenomenon “temperature” in a space s = “Kyoto” on t = December 1st to 31st, 2007.

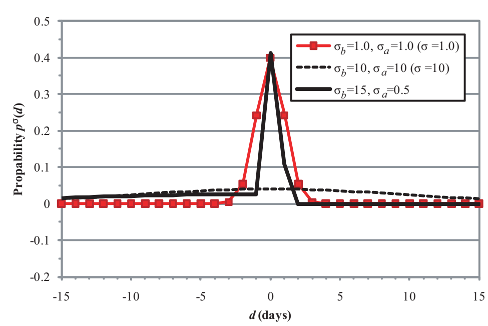

Figure 3. Probability $p^\sigma(d)$ for Temporal-Propagated Weblog Sensors.

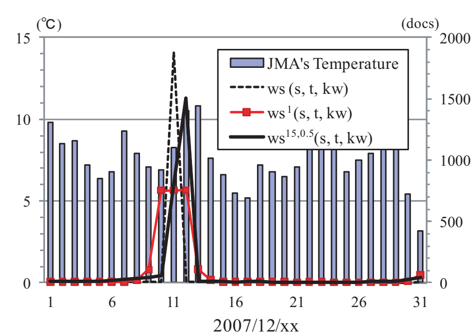

Figure 4. JMA’s Temperature and Temporal-Propagated Weblog Sensors $\mathrm{ws}^{\sigma_b,\sigma_a}(s, t, kw)$, s=“Kyoto”, kw=“atsui” (“hot” in English).

IV. Experiment

This section shows several experimental results to validate the basic Weblog Sensor, the temporal-shifted Weblog Sensors, and the temporal-propagated Weblog Sensors in the case of the target phenomenon "precipitation" or "temperature" in the real world. The whole experiments evaluate the correlation coefficient between real statistics by JMA (Japan Meteorological Agency) [11] as a physical sensor and spatio-temporal data mined by the proposed Weblog Sensors.

Figures 5 to 8 show the average of correlation coefficient between the JMA's precipitation and the temporal-shifted Weblog Sensor $\mathrm{ws}_d(s, t, kw)$ with a temporal shift parameter d (days) using a Japanese noun "ame" ("rain" in English) as a positive correlated keyword to the target phenomenon "precipitation", a Japanese noun "hare" ("shine" in English) as a negative correlated keyword to the target phenomenon "precipitation", and Japanese adjectives "samui" and "atsui" ("chill" and "hot" in English) as noncorrelated keywords to the target phenomenon "precipitation" for each space $s \in$ 47 prefectural capitals in Japan and each day $t \in$ January 1st, 2007 to December 31st, 2009.

In only Figure 5 from among these four figures, the temporal-shifted Weblog Sensor with only $d = 0$ which is equivalent to the basic Weblog Sensor shows positive correlation with JMA's precipitation to some extent, the temporal-shifted Weblog Sensors with $d \in \{-1, 1\}$ show a little positive correlation, i.e., the temporal shift parameter d is ineffective to improve the correlation of the Weblog Sensor at least in the case of the target phenomenon "precipitation".

Figures 9 to 12 show the average of correlation coefficient between the JMA's temperature and the temporal-shifted Weblog Sensor $\mathrm{ws}_d(s, t, kw)$ with a temporal shift d (days) using a Japanese adjective "atsui" ("hot" in English) as a positive correlated keyword to the target phenomenon "temperature", a Japanese adjective "samui" ("chill" in English) as a negative correlated keyword to the target phenomenon "temperature", and "ame" and "hare" ("rain" and "shine" in English) as noncorrelated keywords to the target phenomenon "temperature" for each space $s \in$ 47 prefectural capitals in Japan and each day $t \in$ January 1st, 2007 to December 31st, 2009.

In Figure 9 among these four figures, the temporal-shifted Weblog Sensor with any d shows much more positive correlation with JMA's temperature than in the case of the target phenomenon "precipitation", but gives the best performance when $d = 0$. Meanwhile, in Figure 10 among these four figures, the temporal-shifted Weblog Sensor with any d shows negative correlation with JMA's temperature to some extent, and gives the best performance when $d = -15$. So, the temporal shift parameter d has a possibility to improve the correlation of the Weblog Sensor at least in the case of the target phenomenon "temperature".

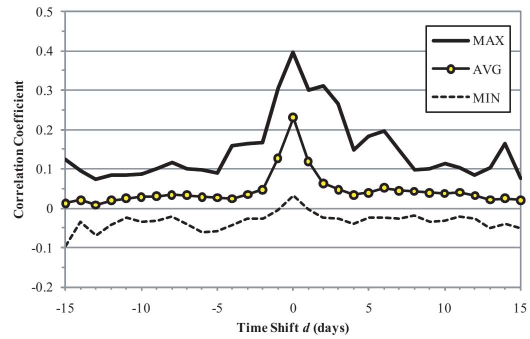

Figure 5. Average Correlation between JMA's Precipitation and Temporal-Shifted Weblog Sensor $\mathrm{ws}_d(s, t, kw)$, kw="ame" ("rain" in English).

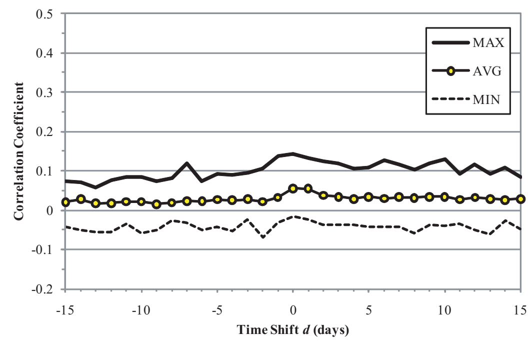

Figure 6. Average Correlation between JMA's Precipitation and Temporal-Shifted Weblog Sensor $\mathrm{ws}_d(s, t, kw)$, kw="hare" ("shine" in English).

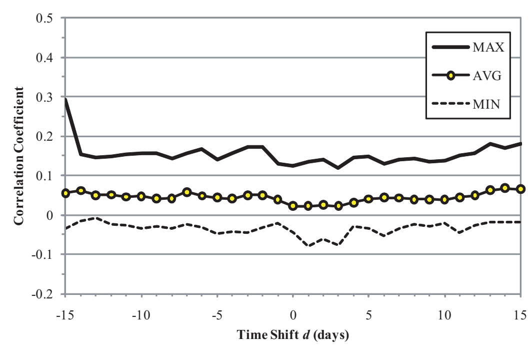

Figure 7. Average Correlation between JMA's Precipitation and Temporal-Shifted Weblog Sensor $\mathrm{ws}_d(s, t, kw)$, kw="atsui" ("hot" in English).

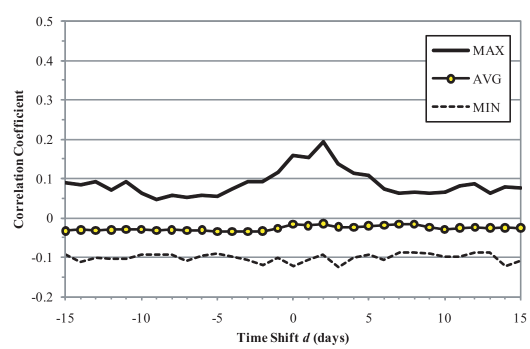

Figure 8. Average Correlation between JMA's Precipitation and Temporal-Shifted Weblog Sensor $\mathrm{ws}_d(s, t, kw)$, kw="samui" ("chill" in English).

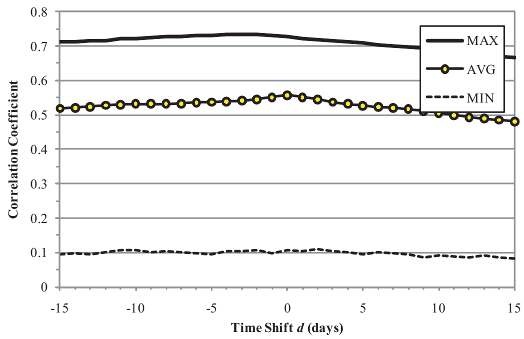

Figure 9. Average Correlation between JMA's Temperature and Temporal-Shifted Weblog Sensor $\mathrm{ws}_d(s,t,kw)$, kw="atsui" ("hot" in English).

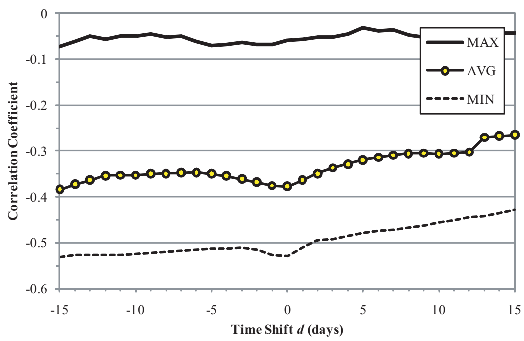

Figure 10. Avg. Correlation between JMA's Temperature and Temporal-Shifted Weblog Sensor $\mathrm{ws}_d(s,t,kw)$, kw="samui" ("chill" in English).

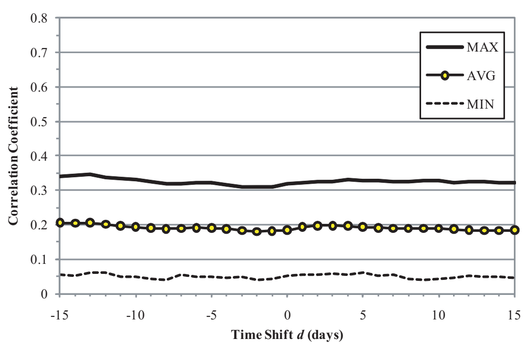

Figure 11. Avg. Correlation between JMA's Temperature and Temporal-Shifted Weblog Sensor $\mathrm{ws}_d(s,t,kw)$, kw="ame" ("rain" in English).

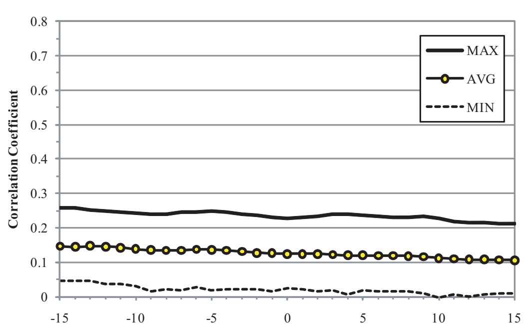

Figure 12. Avg. Correlation between JMA's Temperature and Temporal-Shifted Weblog Sensor $\mathrm{ws}_d(s,t,kw)$, kw="hare" ("shine" in English).

Figure 13 shows the average of correlation coefficient between the JMA's precipitation and the temporal-propagated Weblog Sensor $\mathrm{ws}^{\sigma}(s,t,kw)$ with a temporal propagation parameter σ using a Japanese noun "ame" ("rain" in English) as a keyword representing the target phenomenon "precipitation" for each space $s \in 47$ prefectural capitals in Japan and each day $t \in$ January 1st, 2007 to December 31st, 2009.

Figure 14 shows the average of correlation coefficient between the JMA's precipitation and the temporal-propagated Weblog Sensor $\mathrm{ws}^{\sigma_b,\sigma_a}(s,t,kw)$ with temporal propagation parameters σ_b before a time period t and σ_a after the time period t using a Japanese noun "ame" ("rain" in English) as a keyword representing the target phenomenon "precipitation" for each space $s \in 47$ prefectural capitals in Japan and each day $t \in$ January 1st, 2007 to December 31st, 2009.

These figures show that the basic Weblog Sensor $\mathrm{ws}(s,t,kw)$ is superior to any temporal-propagated Weblog Sensor $\mathrm{ws}^{\sigma_b,\sigma_a}(s,t,kw)$ in the case of the target phenomenon "precipitation" and that the basic Weblog Sensor is followed by the temporal-propagated Weblog Sensor $\mathrm{ws}^{0.4,0.5}(s,t,kw)$, $\mathrm{ws}^{0.3,0.4}(s,t,kw)$, and $\mathrm{ws}^{0.5,0.6}(s,t,kw)$ (i.e., the performance of the temporal-propagated Weblog Sensors is not monotonic decreasing dependent on σ, σ_a, or σ_b).

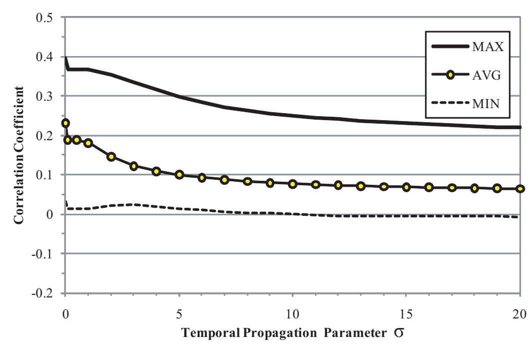

Figure 13. Average Correlation between JMA's Precipitation and Spatio-Temporal Weblog Sensor with Single Temporal Propagation Parameter σ, $\mathrm{ws}^{\sigma}(s,t,kw)$, kw="ame" ("rain" in English).

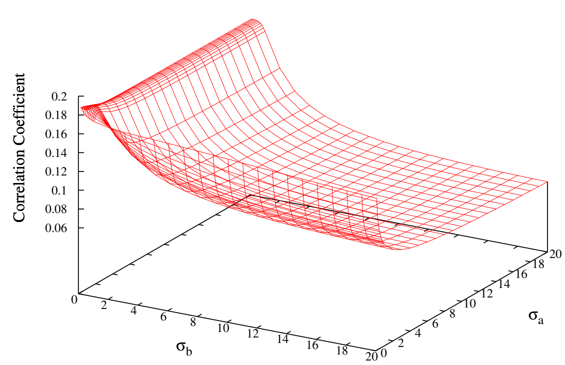

Figure 14. Average Correlation between JMA's Precipitation and Spatio-Temporal Weblog Sensor with Temporal Propagation Parameters σ_b & σ_a, $\mathrm{ws}^{\sigma_b,\sigma_a}(s,t,kw)$, kw="ame" ("rain" in English).

Figure 15 shows the average of correlation coefficient between the JMA's temperature and the temporal-propagated Weblog Sensor $\mathrm{ws}^{\sigma}(s,t,kw)$ with a temporal propagation parameter σ using a Japanese adjective "atsui" ("hot" in English) as a representation of the target phenomenon "temperature" for each space $s \in 47$ prefectural capitals in Japan and each day $t \in$ January 1st, 2007 to December 31st, 2009.

Figure 16 shows the average of correlation coefficient between the JMA's temperature and the temporal-propagated Weblog Sensor $\mathrm{ws}^{\sigma_b,\sigma_a}(s,t,kw)$ with temporal propagation parameters σ_b before a time period t and σ_a after the time period t using a Japanese adjective "atsui" ("hot" in English) as a representation of the target phenomenon "temperature" for each space $s \in 47$ prefectural capitals in Japan and each day $t \in$ January 1st, 2007 to December 31st, 2009.

These figures show that any temporal-propagated Weblog Sensor $\mathrm{ws}^{\sigma_b,\sigma_a}(s,t,kw)$ is superior to the basic Weblog Sensor $\mathrm{ws}(s,t,kw)$ in the case of the target phenomenon "temperature" unlike the case of the target phenomenon "precipitation", that the temporal-propagated Weblog Sensor $\mathrm{ws}^{20,20}(s,t,kw) = \mathrm{ws}^{20}(s,t,kw)$ gives the best performance and gains about 25% of the basic Weblog Sensor, and that the performance of the temporal-propagated Weblog Sensors is monotonic increasing dependent on σ but not always monotonic increasing dependent on σ_a, or σ_b.

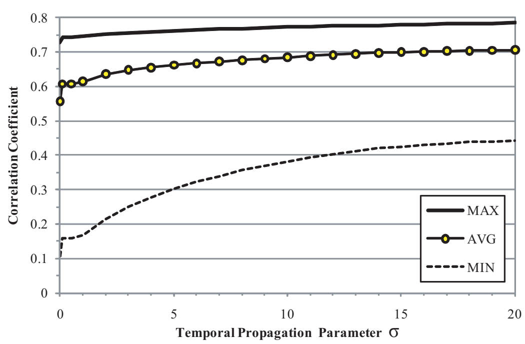

Figure 15. Average Correlation between JMA's Temperature and Spatio-Temporal Weblog Sensor with Single Temporal Propagation Parameter σ, $\mathrm{ws}^{\sigma}(s,t,kw)$, kw="atsui" ("hot" in English).

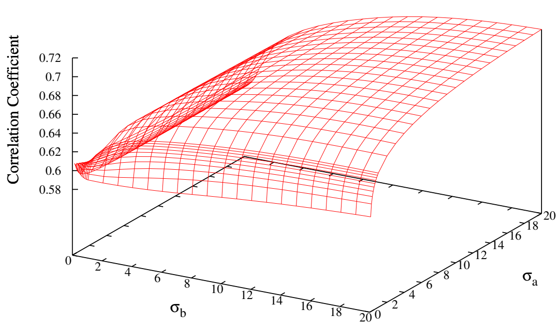

Figure 16. Average Correlation between JMA's Temperature and Spatio-Temporal Weblog Sensor with Temporal Propagation Parameters σ_b & σ_a, $\mathrm{WeblogSensor}_{\sigma_b,\sigma_a}(s,t,kw)$, kw="atsui" ("hot" in English).

V. Conclusion

This paper has defined three kinds of Weblog Sensors to mine the Weblog documents as CGM (Consumer Generated Media) for spatio-temporal data about a target phenomenon (e.g., precipitation and temperature) in the real world: the basic Weblog Sensor, the temporal-shifted Weblog Sensors, and the temporal-propagated Weblog Sensors. And also this paper has validated some potential and reliability of these Weblog Sensors' spatio-temporal data by measuring the correlation coefficient with weather (precipitation and temperature) statistics of Japan Meteorological Agency as real-world data.

References

[1] Hattori, S. and Tanaka, K.: "Towards Building Secure Smart Spaces for Information Security in the Physical World," Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII), vol.11, no.8, pp.1023–1029 (2007).

[2] Dave, K., Lawrence, S., and Pennock, D.M.: "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews," Proc. of WWW'03, pp.519–528 (2003).

[3] Fujimura, S., Toyoda, M., and Kitsuregawa, M.: "A Reputation Extraction Method Considering Structure of Sentence," Proc. of DEWS'05, 6C-i8 (2005).

[4] Tezuka, T., Kurashima, T., and Tanaka, K.: "Toward Tighter Integration of Web Search with a Geographic Information System," Proc. of WWW'06, pp.277–286 (2006).

[5] Inui, K., Abe, S., Morita, H., Eguchi, M., Sumida, A., Sao, C., Hara, K., Murakami, K., and Matsuyoshi, S.: "Experience Mining: Building a Large-Scale Database of Personal Experiences and Opinions from Web Documents," Proc. of WI'08, pp.314–321 (2008).

[6] Hattori, S., Ohshima, H., Oyama, S., and Tanaka, K.: "Mining the Web for Hyponymy Relations based on Property Inheritance," Proc. of APWeb'08, LNCS Vol.4976, pp.99–110 (2008).

[7] Hattori, S., and Tanaka, K.: "Extracting Concept Hierarchy Knowledge from the Web based on Property Inheritance and Aggregation," Proc. of WI'08, pp.432–437 (2008).

[8] Hattori, S., Tezuka, T., and Tanaka, K.: "Mining the Web for Appearance Description," Proc. of DEXA'07, LNCS Vol.4653, pp.790–800 (2007).

[9] Hattori, S.: "Peculiar Image Search by Web-extracted Appearance Descriptions," Proc. of SoCPaR'10, pp.127–132 (2010).

[10] Hattori, S. and Tanaka, K.: "Mining the Web for Access Decision-Making in Secure Spaces," Proc. of SCIS&ISIS'08, TH-G3-4 (2008).

[11] Japan Meteorological Agency, http://www.jma.go.jp/jma/indexe.html (2010).

[12] Google Blog Search, http://blogsearch.google.co.jp/ (2010).