International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 8 (2016) pp. 042–051

Potential Information Maximization: Potentiality-Driven Information Maximization and Its Application to Tweets Classification and Interpretation

Ryozo Kitajima[1], Ryotaro Kamimura[2], Osamu Uchida[3] and Fujio Toriumi[4]

[1]Graduate School of Science and Technology, Tokai University,
4-1-1 Kitakaname, Hiratsuka, Kanagawa 259-1292, Japan
3btad004@mail.tokai-u.jp

[2]IT Education Center, Graduate School of Engineering and Graduate School of Science and Technology, Tokai University,
4-1-1 Kitakaname, Hiratsuka, Kanagawa 259-1292, Japan
ryo@keyaki.cc.u-tokai.ac.jp

[3]Dept. Human and Information Science, Graduate School of Engineering
and Graduate School of Science and Technology, Tokai University,
4-1-1 Kitakaname, Hiratsuka, Kanagawa 259-1292, Japan
o-uchida@tokai.ac.jp

[4]Graduate School of Engineering, The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
tori@sys.t.u-tokyo.ac.jp

Abstract: **The present paper aims to apply a new information-theoretic learning method called "potential information maximization" to the classification and interpretation of tweets. It is well known that social media sites such as Twitter play a crucial role in transmitting important information during natural disasters. In particular, since the Great East Japan Earthquake in 2011, Twitter has been considered as one of the most efficient and convenient communication tools. However, since there is much redundant information contained in tweets, it is critical that methods be developed to extract only the most important information from them. To cope with complex and redundant data, a new neural information-theoretic learning method has been developed for this purpose. The method aims to find neurons with high potential and maximize their information content to reduce redundancy and to focus on important information. The method was applied to real tweet data collected during the earthquake. It was found that the method could classify the tweets as important and unimportant more accurately than other conventional machine learning methods. In addition, the method made it possible to interpret how the tweets could be classified based on the examination of highly potential neurons.**

Keywords: twitter, classification, interpretation, neural network, potential information.

I. Introduction

A. Importance of Tweets

It is well known that social media sites such as Twitter play a very important role in enabling communication during natural disasters. During the Great East Japan Earthquake and the following Tsunami in 2011, it was reported that Twitter was widely used to send valuable information to people suffering from the disasters [1], [2], [3], [4], [5], [6]. However, such tweets contain much redundant information, which has prevented us from efficiently using the information transmitted during the disaster. Thus, it is critical to develop new methods to filter information in the tweets and find new and valuable information to be transmitted to people in need.

In our previous study [7], we developed a model using the backpropagation method. This model could classify tweets with a low error rate, but as is the case with the backpropagation, it was impossible to interpret the knowledge obtained by the model. Certainly, the backpropagation method or machine learning techniques such as support vector machines have been proved to be effective in classifying the tweets into valuable and invaluable ones to find important information. However, though they are good at classifying tweets, they are very weak in explaining why and how the classification is performed. For the methods to be practically used, it is important to know precisely how the method can classify the tweets and explain the inference mechanisms.

B. Potential Information Maximization

As an attempt to explain inference mechanisms, many information-theoretic methods have been developed [8] which condense information on input patterns into a small number of neurons. Because the objective of neural networks is naturally to acquire as much necessary information on input patterns as possible, information-theoretic methods have been applied to neural networks in various ways [9]. Though information-theoretic methods have had much influence on learning in neural networks, they have not been used to their full potential because of complex computations and passive interpretation.

Complex computation refers to the large amount of computations required for information-theoretic methods to obtain information content. Passive interpretation becomes a serious problem when attempting to interpret internal representations obtained by the information maximization method, which is fundamentally a passive way to fire neurons. By maximizing information content, a small number of neurons tend to fire, and all others cease to do so [8]. So, we try to interpret only a small number of these, expecting that information content in input patterns can be compressed into those which are firing. However, one of the main problems is that information maximization cannot specify which neurons should be fired. This means that we have difficulty in interpreting the final firing neurons because of the uncertainty. From this point of view, the conventional information-theoretic methods are passive in terms of neuron firing. This passive firing has prevented us from clearly interpreting final internal representations. To address this concern, we propose that the potentiality of neurons and information should be increased such that only highly potential neurons fire.

C. Neuron Potentiality

Potentiality learning has been previously developed in order to identify those neurons with high potentiality, and improve the real performance of neural networks [10], [11], [12]. Potentiality refers to the general property of neurons to respond to as many different situations as possible. If it is possible to find neurons with high potentiality, they should be enhanced as much as possible and eventually contain the important characteristics of input patterns. However, appropriately defining potentiality can be quite difficult. For the first approximation of potentiality, the variance of neurons is used, as suggested by Linsker [13]. This is because highly potential neurons should respond differently to as many input patterns or neurons as possible. The learning in neural networks is usually initialized with random seeds, and connection weights are trained to decrease errors between targets and outputs. It is certain that neurons with higher variance or potentiality by the random initialization are not necessarily effective in improving the general performance of neural networks. Thus, it is necessary to transform the variance into a more useful form: potentiality.

To transform the variance into potentiality, the present method uses the self-organizing maps, or SOM. For highly potential neurons to be applied to practical problems such as tweet classification, the potentiality is computed with the help of the self-organizing maps. The SOM has been widely used for extracting important characteristics from input patterns [14], [15]. In addition, a number of computational methods have been developed to clarify SOM knowledge [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27]. The SOM can be used to compute highly potential neurons containing the main characteristics of input patterns. However, the method does not simply use the SOM; rather, it actively transforms the representations created by the SOM. Through the highly potential neurons obtained by the help of the SOM, the SOM knowledge is clarified as much as possible. In particular, a few input neurons are forced to be responsible for extracting important characteristics. Thus, the information-theoretic method proposed here can produce highly potential neurons which respond differently and appropriately to as many input patterns as possible.

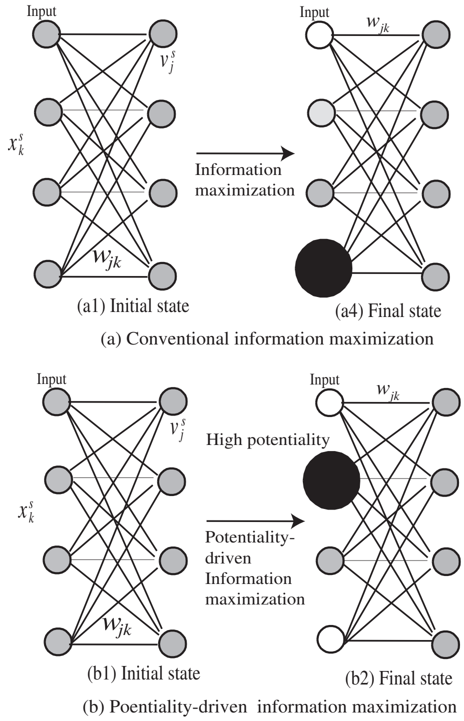


Figure. 1: Concept of conventional (a) and potential (b) information maximization.

D. Paper Outline

In Section 2, we first explain the concept of potential information maximization. Computational procedures to compute potentiality and potential information are given in the following sections with reference to the SOM learning procedures. Then, we show how to update connection weights considering potential information. In Section 3, we present the experimental results for the tweet data collected during the Great East Japan Earthquake. We show that generalization performance was improved, a few important variables were obtained, and the meaning of those variables was naturally interpreted.

II. THEORY AND COMPUTATIONAL METHODS

A. Potential Information Maximization

We have so far tried to maximize information content in hidden neurons [8]. When the information increases, the number of strongly firing neurons becomes smaller and finally only one neuron fires, while all others ceases to do so. This type of information maximization is used to fire a smaller number of neurons, corresponding to input patterns [8].

We here focus on the information content of input neurons and try to extract the most important ones. As mentioned above, one of the major problems is that it is impossible to specify which neurons should be fired by information maximization. As shown in Figure 1(a), a neuron is automatically fired by the procedures of information maximization. Thus, the conventional information maximization is a passive method with respect to the firing of neurons. This means that sometimes there is difficulty in interpreting which neurons actually fire. The new method of potential information maximization aims to fire a neuron with the highest potentiality, as shown in Figure 1(b). In the new method, a neuron to be fired is not determined by the information maximization procedure, but by the criterion of potentiality of the neuron toward input patterns. Information should be maximized under the condition that the potential neurons are forced to fire. Figure 1(b) shows this situation well. By the information maximization procedure, the fourth neuron is fired in Figure 1(a). However, the potentiality of neurons is used to fire the second neuron instead of the fourth neuron in Figure 1(b). Because the second input neuron has the highest potentiality toward input patterns, the neuron has high potentiality for dealing with many input patterns.

B. Learning Procedures

Thus far, it has been supposed that some neurons should have high potentiality in order to improve the general performance of neural networks. Intuitively, highly potential neurons should have the ability to respond appropriately to as many input patterns and neurons as possible. Considering the simplicity, some of this property can be approximated by computing the variance of neurons. Neurons with larger variance should respond differently to many input patterns or neurons. Thus, for the first approximation of the potentiality, neurons with larger variance are considered to be highly potential ones. Naturally, neurons with larger variance are not necessarily effective in extracting main characteristics. Thus, by the help of the self-organizing maps, the variance is forced to represent the main characteristics.

Figure 2 shows the process of potential information maximization learning. In the potential information maximization phase in Figure 2(a), the SOM is applied and the potentiality of input neurons is computed for the first approximation in Figure 2(a1). This approximated potentiality is used to train connection weights by the SOM successively in Figures 2(a2)-(a3). In the final state of the potential information maximization phase in Figure 2(a4), the fourth input neuron's potentiality becomes the largest.

Then, those connection weights obtained in the potential information maximization phase in Figure 2(a) and the potentiality are transferred to the potential information assimilation phase of the supervised learning in Figure 2(b). The potentiality and corresponding connection weights are used as the initial weights for the potential information assimilation phase. Those initial weights and the potentiality guide the supervised learning, and the final weights are obtained by changing the original weights in the potentiality determination phase. Finally, in the potential information adjustment phase in Figure 2(c), connection weights are adjusted to eliminate the over-training effects.

C. Computing Potentiality

As mentioned, the SOM is applied to extract characteristics from input patterns, and the potentiality is determined by using the variance of connection weights. This potentiality is then incorporated into the learning processes to assimilate the potentiality over the connection weights in Figures 2(a1)-a(4). To realize this process, we should define the potentiality of individual input neurons.

Let w_{jk} denote the connection weights from the kth input neuron to the jth output neuron in Figure 2. Then, the variance of the kth input neuron is defined by

$$V_k = \sum_{j=1}^{M} (w_{jk} - w_k)^2, \tag{1}$$

where M is the number of hidden neurons and the weight w_k denotes the averaged weights computed by

$$w_k = \frac{1}{M} \sum_{j=1}^{M} w_{jk}. \tag{2}$$

Then, the input potentiality is defined by dividing the variance by its maximum variance

$$\phi_k = \left(\frac{V_k}{V_{\max}} \right)^r, \tag{3}$$

where r is the potentiality parameter and $r > 0$. The input potentiality ranges between zero and one, and individual variances or potentialities are changed easily by changing the parameter r. When the parameter r becomes larger, the majority of individual potentialities become smaller and only a small number of individual neurons have higher potentiality.

D. Potential Information

In the method, it is supposed that a few input neurons tend to have higher potentiality. To describe this selectivity of potentiality, we need to introduce the concept of potential information. For this, first, the potentiality is normalized by

$$\phi_k^{nrm} = \frac{\phi_k}{\sum_{k=1}^{L} \phi_k}, \tag{4}$$

where L is the number of input neurons. Then, we can compute the entropy H by

$$H = - \sum_{k=1}^{L} \phi_k^{nrm} \log \phi_k^{nrm}. \tag{5}$$

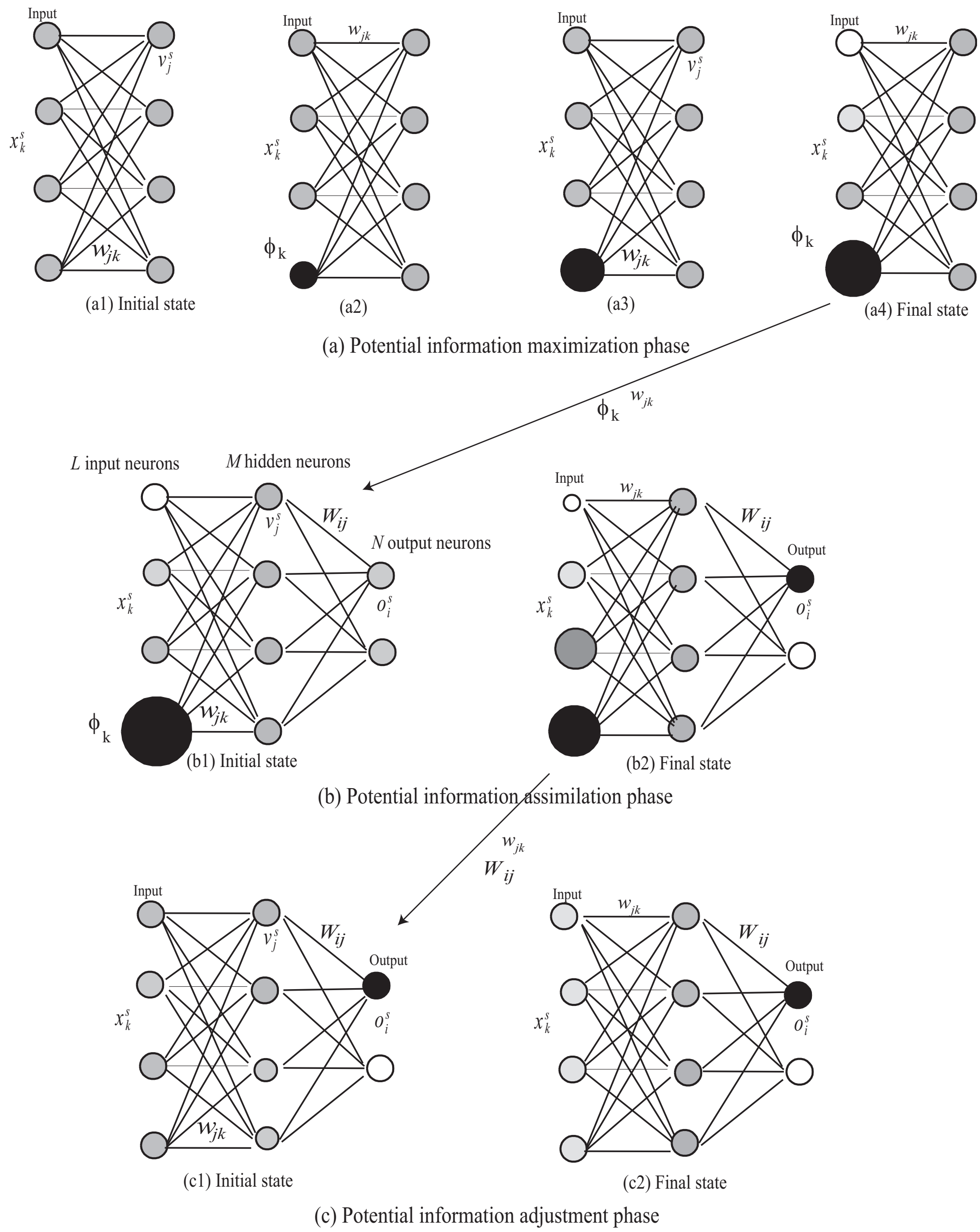


Figure. 2: Concept of potential information maximization learning, where the potential information is increased in the potential information maximization phase (a); the knowledge by this phase is transferred to the potential information assimilation phase (b); and finally some adjustment is performed, in particular, to eliminate the over-training effects in the potential information adjustment phase (c).

Then, the potential information is defined by

$$PI = \frac{H_{\max} - H}{H_{\max}}. \tag{6}$$

When this potential information becomes larger, only a small number of input neurons tend to have higher potentiality.

E. SOM-based Potentiality

In the potential information maximization phase, we use the self-organizing maps (SOM). Let w_{jk} denote the connection weights between the kth input and the jth hidden neuron, and x_k^s represent the kth element of the sth input pattern. Then, the distance between the sth input pattern and the jth neuron is defined by

$$d_j^s \quad = \quad \sum_{k=1}^{L} \phi_k^r (x_k^s - w_{jk})^2, \tag{7}$$

where L is the number of input neurons or elements of an input neuron. The winner j for the sth input pattern is given by

$$j^* = \mathrm{argmin}_j d_j^s. \tag{8}$$

The connection weights w_{jk} are obtained by

$$w_{jk} = \frac{\sum_{s=1}^{S} h_{jj^*} x_k^s}{\sum_{s=1}^{S} h_{jj^*}}, \tag{9}$$

where h denotes the neighborhood function and is defined by

$$h_{jj^*} = \exp\left(-\frac{\parallel \mathbf{j} - \mathbf{j}^* \parallel}{\sigma}\right) \tag{10}$$

where **j** denotes the position of the jth neuron on the output map.

F. Computational Procedures

After the potential information maximization phase (Figure 2(a4)), training is moved to the potential information assimilation phase (Figure 2(b1)). This phase is based on the back-propagation (BP) training method. First, initial weights are transferred from the potentiality information maximization phase. Then, potentiality is used to modify the connection weights according to the magnitude of the potentiality. Finally, it is applied to the connection weights from the input to hidden neurons. For the method to be practically applicable and to avoid over-training, the connection weights multiplied by the potentiality in the potential information maximization phase are given as the initial connection weights

$$^{new}w_{jk} = {}^{old}w_{jk}\phi_k. \tag{11}$$

Supervised learning is performed using these new connection weights in the potential information assimilation phase in Figure 2(b). Actually, the errors between the targets and outputs are minimized by the conventional BP method[1]. Usually, initial connection weights play the most important role in learning, and it is necessary to train connection weights longer than expected to assimilate the obtained connection weights with a given potentiality. Thus, over-training becomes a serious problem. To eliminate the effect of longer assimilation processes, the final potential information adjustment phase is introduced in Figure 2(c), where early stopping is used to eliminate the effects of over-training. This is a way to stop learning when the error of the validation dataset is over the error of the training dataset.

III. Results and Discussion

A. Tweet data

The method was applied to the tweet data collected during the Great East Japan Earthquake. The data was composed of 600 tweets (294 necessary and 306 unnecessary tweets). To analyze the tweets, the text data were decomposed into morphemes and transformed into quantitative data. The morphological analysis was performed using the Japanese morphological analysis software "JUMAN" [28]. Through the morphological analysis, 25 variables were created related to the words shown in Table 1. Variables No.1 to No. 24 represent the corresponding word frequencies (variable No.1, for example, is the word frequency related to people), while variable No.25 represented the number of characters used in a tweet. For the experiment, the number of hidden neurons was 30. Then, the data set was divided randomly into the training (70%), validation (15%) and testing (15%) with ten runs.

[1]The Matlab neural networks package was used with all default parameter values for easy reproduction.

Table 1: Variables of the dataset

No	Name	No	Name
1	People	16	Culture
2	Organization	17	Family
3	Artificial-vehicle	18	Food
4	Artificial-other	19	Transportation
5	Nature	20	Education
6	Location-institution	21	Business
7	Location-Nature	22	Media
8	Location-function	23	Politics
9	Location-other	24	Transportation-railroad(line)-Tokyo
10	Abstraction	25	Number of characters
11	Shape		
12	Quantity		
13	Time		
14	Location name:Japan		
15	URL link		

B. Evaluating Generalization Performance

1) Potential Information and Generalization

Figure 3(a) shows the potential information when the parameter r increased from 0.1 to 1. As can be seen in the figure, the potential information increased gradually when the parameter increased, meaning that the number of input neurons with higher potentiality became smaller when the parameter r increased.

Figure 3(b) shows the generalization errors as a function of the parameter r. When the parameter increased from 0.1 to 0.2, the errors first increased, then decreased to a minimum point of 0.207 when the parameter r was 0.5. Though the errors did not necessarily decrease in proportion to the increase in the potential information, by changing the parameter r and correspondingly the potential information, generalization could be changed. Compared with the generalization errors of 0.257 and 0.250 by the conventional BP and the method without potentiality, the generalization error by the present method was much lower.

2) Generalization Comparison

Table 2 shows the summary of the results by the present method and three other methods, namely, the method without the potentiality (without), the conventional back-propagation (BP) and the fine-tuned support vector machines (SVM). The conventional BP was used with the early stopping criteria to eliminate the over-training effects. The supposed vector machines were tuned by changing the parameters, and the best possible results were shown in the table.

As shown in the table, the best possible values were obtained by the potential information method in terms of the average, minimum and maximum values. The second best value was obtained by the method without the potentiality, meaning that the SOM and BP were simply combined. The support vector machines showed the third best value but had the best standard deviation. This means that the support vector machines could produce relatively stable results. Finally, the worst case was obtained by the standard BP with the early stopping.

C. Interpreting Internal Representation

1) Change in Potentiality

Figure 4 shows the potentialities ϕ_k when the parameter r increased from 0.1 (a) to 1 (j). When the parameter r was 0.1,

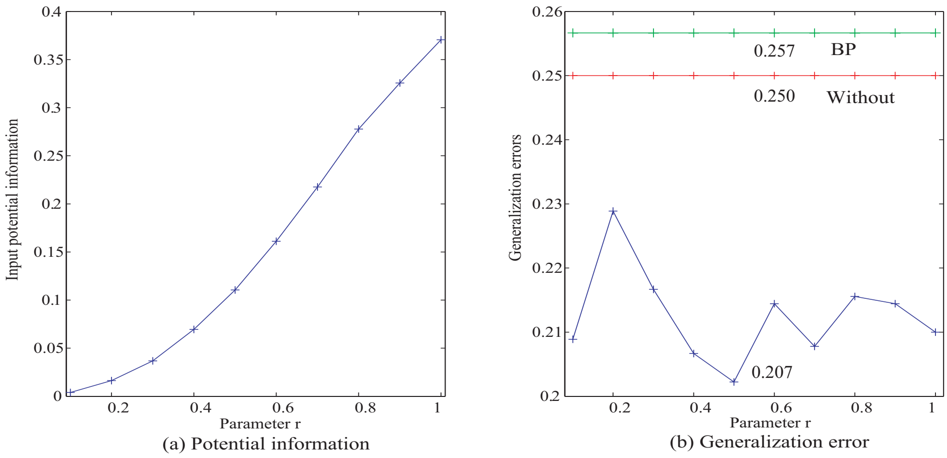


Figure. 3: Potential information (a) and generalization errors (b) for the tweet data.

Table 2: Summary of the experimental results (generalization errors) by the four methods for the Twitter data set. The "without" method represents the one without the potential information maximization phase. The bold face numbers represent the best values.

Method	Avg	Std dev	Min	Max
Potential	**0.207**	0.052	**0.111**	**0.289**
Without	0.250	0.031	0.200	0.300
BP	0.257	0.038	0.178	0.322
SVM	0.253	**0.030**	0.200	0.311

all input potentialities were almost equal with small fluctuations. Then, when the parameter r increased gradually from 0.2 to 0.9, the input potentialities became differentiated, and some input neurons tended to have higher potentiality. Finally, when the parameter r was 1.0 in Figure 4(j), only the 19th, 3rd and 17th input neurons tended to have higher potentiality. When those input neurons were examined, it was found that the 3rd and 19th were related to the category "transportation", and the 17th input neuron represented "family". This means that the "transportation" and "family" categories were most important in classifying the tweets.

2) Interpreting Connection Weights

Here we present our interpretation of connection weights by the new method and conventional BP. Figure 5(c) shows connection weights by the conventional back-propagation. As can be seen in the figure, the connection weights were random and it was impossible to see any regularity.
On the other hand, Figure 5(a) shows the connection weights by the potential information maximization phase, namely, by repeating the SOM procedures many times. As can be seen in the figure, regular and symmetric patterns were observed with help from the SOM. Figure 5(b) shows the connection weights for the potential information adjustment phase. By the potential information adjustment phase, only a small number of connection weights remained strong, for example, connection weights from the 19th (transportation), 17th (family) and 3rd (vehicle) input neurons. This

Table 3: Top five important words related to three highly potential neurons (variables).

Rank	Transportation	Family	Vehicle
1	traffic congestion	traffic congestion	cars
2	cars	cars	trains
3	roads	returning home	buses
4	driving	power failure	taxes
5	streets	roads	motorbikes

suggests that connection weights from those three input neurons should play an important role in learning.
We then tried to interpret the meaning of the variables by examining all related words. Table 3 shows the list of detected words related to the three important variables. The variable "transportation" was closely related to words such as "traffic congestion", "cars", "roads", "driving" and "street". The variable "vehicle" was naturally related to "cars", "trains", "buses", "taxes" and "motorbikes". In addition, the variable "family" was closely connected with traffic conditions such as "traffic congestion", "cars", "returning home", "power failure" and "roads". Thus, it can be said that all three variables are basically represented by words related to traffic conditions. During the Great East Japan Earthquake, it was reported that a number of people were unable to return home because of paralyzed roads and public transportation systems [29]. These paralyzed situations were well described in the several studies on the difficult traffic conditions during disasters [30], [31], [32]. Though the neural networks had no knowledge of these traffic conditions, they could discover their importance in this situation. Thus, the experimental results certainly show the possibility of the present method to extract important information from complex data.

D. Discussion

1) Explaining the Improved Performance

The potential information maximization method produced the best generalization performance, which can be explained by two facts, namely, SOM knowledge and information compression. First, to extract important characteristics, the potential method is based on knowledge obtained by the self-organizing maps, or SOM knowledge. It has been well

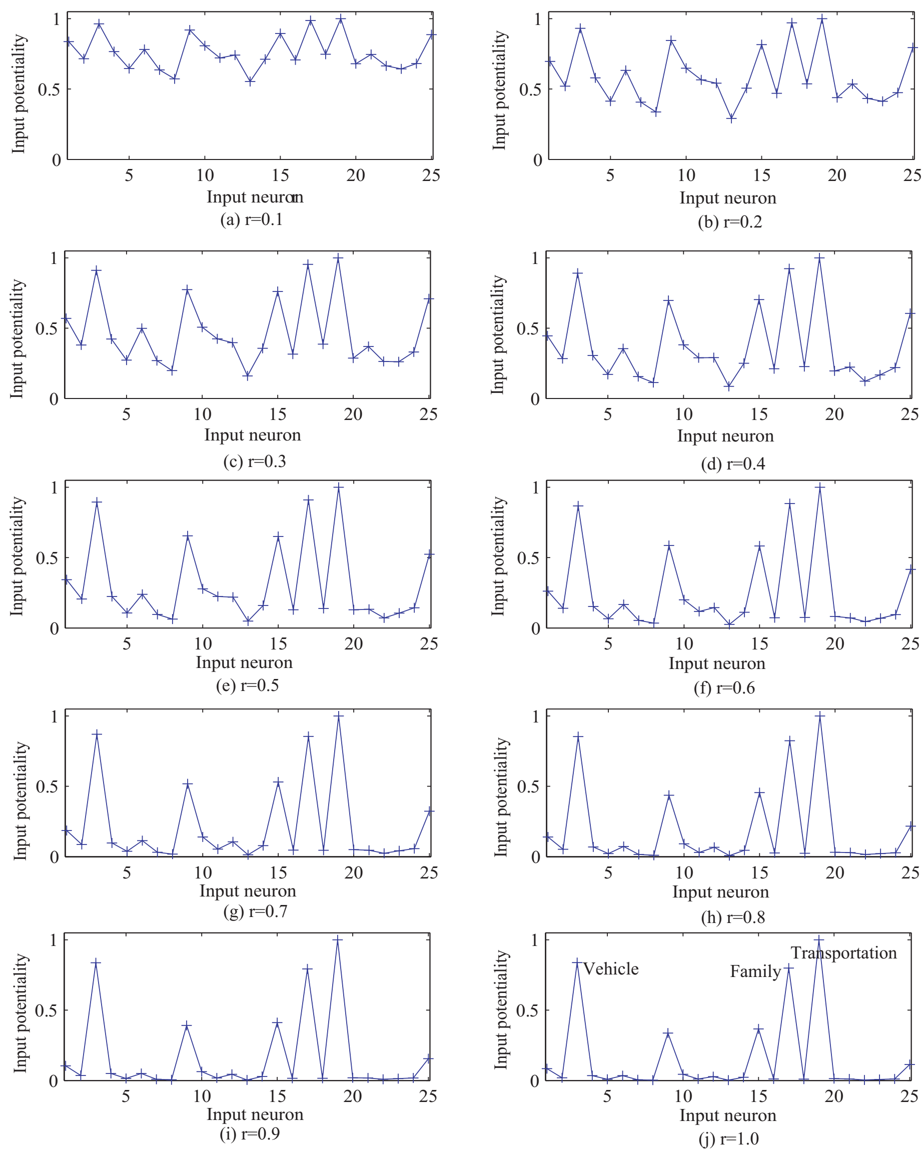
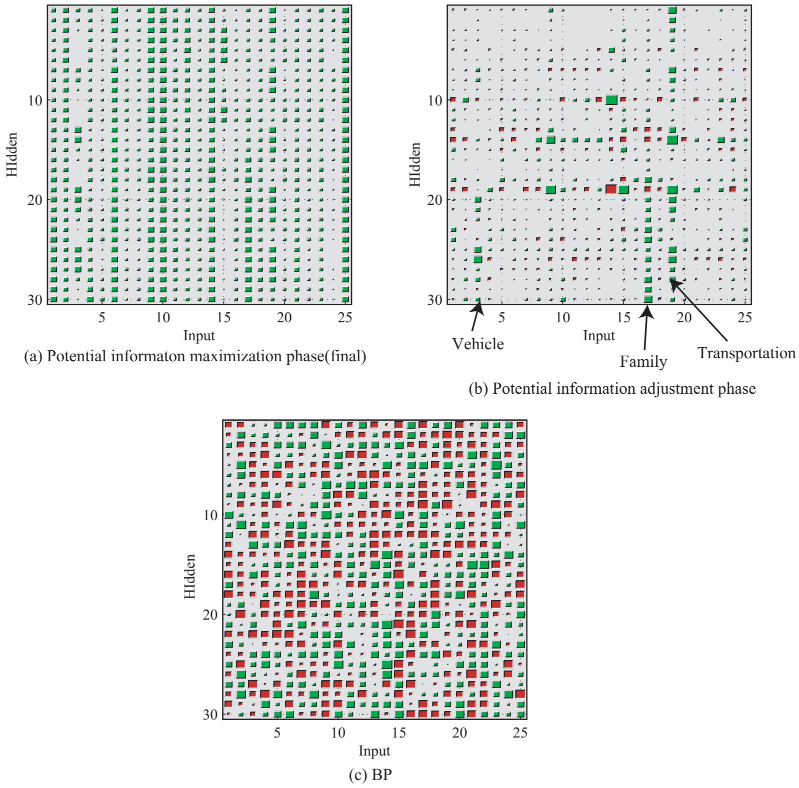


Figure. 4: Input potentialities for Twitter data set.



(a) Potential informaton maximization phase(final)

(b) Potential information adjustment phase

(c) BP

Figure. 5: Connection weights in the potential information maximization phase (a), the potential information adjustment phase (b) and the conventional back-propagation (c). Weights in green and red represent positive and negative ones.

known that the SOM can extract important features from complex input patterns. As shown in Figure 5(a), some symmetric and regular patterns were observed over connection weights by the SOM or in the potential information maximization phase. These regular patterns contributed to improved generalization. Because the SOM knowledge is sometimes very ambiguous and redundant, it is quite difficult to clarify it and use it to improve the general performance of neural networks [18], [19], [20], [21], [22], [23], [24], [25], [26]. The potential information maximization method aims to focus on some important parts of connection weights to improve the performance. The potential method uses the variance of input neurons as a measure of the potentiality. Though neurons with higher variance do not necessarily show better generalization performance, variance with the help of SOM knowledge was a good indicator of important neurons.

In addition, a relatively small number of strong connection weights made it possible to interpret the main mechanism of neural networks. In the case of the tweets analysis, their usefulness could be classified into categories such as “transportation” and “family”. It is also possible to completely interpret the main inference mechanism of neural networks by carefully examining the small number of connection weights. This improved interpretation performance is effective in the practical analysis of tweets.

2) Relations to the SOM

Potential information maximization method can be considered as a method to improve the performance of the SOM in two ways, namely, clarifying the knowledge and allowing extension to supervised learning. First, the potential method could help clarify SOM knowledge. Though the SOM has a good reputation for extracting important characteristics from input patterns, it is not so easy to explicitly represent the SOM knowledge. As mentioned, this has led to the development of a number of different types of computational methods for clarification. The potential method tries to make each input neuron’s role more explicit, and this eventually leads to the clarification of the SOM knowledge.

Second, the present method aimed to use the SOM knowledge to train supervised learning. Though several attempts were made to extend the SOM to supervised learning [14], [15], they were not necessarily successful, compared with conventional supervised learning methods such as the back-propagation method. Intuitively, if it is possible to use the SOM knowledge to train conventional supervised learning, such training can be made more efficient due to the rich SOM knowledge. The present method aims to include the SOM knowledge in supervised learning by focusing the small number of neurons.

IV. Conclusion

The present paper applied a new neural learning method called “potential information maximization” to the classification and interpretation of tweets collected during the Japan East Earthquake. It is well known that Twitter plays an important role as a major communication tool during disasters. However, because there is much redundant information in tweets, the raw data cannot necessarily give valuable information to people in need. It is thus critical to extract the most

important information from abundant, redundant data.
For such complex tweet data, conventional machine learning methods such as support vector machines have proved to be effective in terms of generalization performance. However, the conventional methods have some difficulty in interpreting final results. The present method aimed not only to improve generalization performance but also to improve interpretation. The interpretation of final results was realized by focusing on some important or highly potential input neurons. These highly potential neurons were selected with help from the SOM by compressing the characteristics of the input patterns into a few highly potential neutrons.
Naturally, several problems should be considered to improve performance. For example, in the present method, it was necessary to eliminate the effects of over-training in the potential information adjustment phase, transforming the learning procedure into a 3-stage process. Thus, in the future, the method should be simplified by taking into account the over-training effects in higher phases Though the method can and should be applied to larger tweet datasets to reaffirm the validity of the method, the present study showed the possibility for the new method to be applied to the extraction of important information from redundant and complex data.

References

[1] Y. Yoshitsugu, "Roles of social media at the time of major disasters observed in the great east japan earthquake:twitter as an example (in Japanese)," *The NHK Monthly Report on Broadcast Research*, vol. 7, pp. 16–23, 2011.

[2] H. Wilensky, "Twitter as a Navigator for Stranded Commuters during the Great East Japan Earthquake," *Proc. of 11th Int'l ISCRAM Conference*, pp. 695–704, 2014.

[3] F. Toriumi, T. Sasaki, K. Shinoda, K. Kazama, S. Kurihara, I. Nodao, "Information Sharing on Twitter during the 2011 Catastrophic Earthquake," *Proc. 22nd Int'l Conference on World Wide Web Companion*, pp. 1025–1028, 2013.

[4] M. Kaigo, "Social Media Usage during Disasters and Social Capital," *Keio Communication Review*, no. 34, pp. 19–35, 2012.

[5] S. Doan, B.-K. H. Vo, N. Collier, "An Analysis of Twitter Messages in the 2011 Tohoku Earthquake," *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol. 91, pp. 58–66, 2012.

[6] A. Acar, Y. Muraki, "Twitter for Crisis Communication: Lessons Learned from Japan's Tsunami Disaster," *Int. J. Web Based Communities*, vol. 7, no. 3, pp. 392–402, 2011.

[7] R. Kitajima, R. Kamimura, O. Uchida, F. Toriumi, "A Study on Tweet Classification by Neural Networks (in Japanese)," *The 29th Annual Conference of the Japanese Society for Artificial Intelligence*, 2015.

[8] R. Kamimura, "Information-theoretic competitive learning with inverse Euclidean distance output units," *Neural Processing Letters*, vol. 18, pp. 163–184, 2003.

[9] R. Kamimura, "Selective information enhancement learning for creating interpretable representations in competitive learning," *Neural Networks*, vol. 24, no. 4, pp. 387–405, 2011.

[10] R. Kamimura, R. Kitajima, "Selective Potentiality Maximization for Input Neuron Selection in Self-Organizing Maps," *2015 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2015.

[11] R. Kamimura, "Self-Organizing Selective Potentiality Learning to Detect Important Input Neurons," *(To appear in The 2015 IEEE International Conference on Systems, Man, and Cybernetics)*, 2015.

[12] R. Kitajima, R. Kamimura, O. Uchida, F. Toriumi, "Neural potential learning for tweets classification and interpretation," *The 7th International Conference on Soft Computing and Pattern Recognition*, 2015.

[13] R. Linsker, "Self-organization in a perceptual network," *Computer*, vol. 21, pp. 105–117, 1988.

[14] T. Kohonen, *Self-Organization and Associative Memory*. New York: Springer-Verlag, 1988.

[15] T. Kohonen, *Self-Organizing Maps*. Springer-Verlag, 1995.

[16] J. Vesanto, "SOM-based data visualization methods," *Intelligent Data Analysis*, vol. 3, pp. 111–126, 1999.

[17] S. Kaski, J. Nikkila, and T. Kohonen, "Methods for interpreting a self-organized map in data analysis," in *Proceedings of European Symposium on Artificial Neural Networks*, (Bruges, Belgium), 1998.

[18] I. Mao and A. K. Jain, "Artificial neural networks for feature extraction and multivariate data projection," *IEEE Transactions on Neural Networks*, vol. 6, no. 2, pp. 296–317, 1995.

[19] C. De Runz, E. Desjardin, and M. Herbin, "Unsupervised visual data mining using self-organizing maps and a data-driven color mapping," in *Information Visualisation (IV), 2012 16th International Conference on*, pp. 241–245, IEEE, 2012.

[20] S. Shieh and I. Liao, "A new approach for data clustering and visualization using self-organizing maps," *Expert Systems with Applications*, 2012.

[21] H. Yin, "ViSOM-a novel method for multivariate data projection and structure visualization," *IEEE Transactions on Neural Networks*, vol. 13, no. 1, pp. 237–243, 2002.

[22] M.-C. Su and H.-T. Chang, "A new model of self-organizing neural networks and its application in data projection," *IEEE Transactions on Neural Networks*, vol. 123, no. 1, pp. 153–158, 2001.

[23] S. Wu and T. Chow, "Prsom: A new visualization method by hybridizing multidimensional scaling and self-organizing map," *Neural Networks, IEEE Transactions on*, vol. 16, no. 6, pp. 1362–1380, 2005.

[24] L. Xu, Y. Xu, and T. W. Chow, "PolSOM-a new method for multidimentional data visualization," *Pattern Recognition*, vol. 43, pp. 1668–1675, 2010.

[25] Y. Xu, L. Xu, and T. Chow, "Pposom: A new variant of polsom by using probabilistic assignment for multidimensional data visualization," *Neurocomputing*, vol. 74, no. 11, pp. 2018–2027, 2011.

[26] L. Xu and T. Chow, "Multivariate data classification using polsom," in *Prognostics and System Health Management Conference (PHM-Shenzhen), 2011*, pp. 1–4, IEEE, 2011.

[27] S. Van Gassen, B. Callebaut, M. J. Van Helden, B. N. Lambrecht, P. Demeester, T. Dhaene, and Y. Saeys, "FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data,"*Cytometry Part A*, 2015.

[28] Kyoto University, Kurohashi & Kawahara Lab, "Juman (a user-extensible morphological analyzer for japanese) ver.7.0." `http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN`, 2012.

[29] K. Uri "March 11: How the Great East Japan Earthquake News were Viewed (in Japanese)," *The NHK Monthly Report on Broadcast Research*, vol. 7, pp. 2–15, 2011.

[30] Y. Hara "Returning-Home Analysis in Tokyo Metropolitan Area at the Time of the Great East Japan Earthquake using Twitter Data (in Japanese)," *Journal of Natural Language Processing*, vol. 20, no. 3, pp. 315–334, 2013.

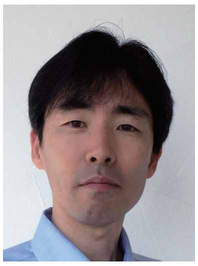
[31] H. Takayanagi, H. Oneyama, T. Ishikura and S. Shikata "Returning Home of Stranded Commuters in Tokyo in the East Japan Great Earthquake -Analysis of the Situation on Twitter- (in Japanese)," *Papers of Research Meeting on Civil Engineering Planning*, vol. 45, pp. 1–7, 2012.

[32] T. Rokuse, O. Uchida and F. Toriumi "Tweets Classification Method to Provide Useful Information at the Time of Large Scale Disaster (in Japanese)," *Forum on Information Technology*, vol. 13, no. 4, pp. 313–316, 2014.

Author Biographies

Ryozo Kitajima is a student of doctoral program at graduate school of science and technology in Tokai University. He received his B.A. degree in 2006 and M.Eng. degree in 2008 from Tokai University, Kanagawa, Japan. His research interests are neural networks, human behavior, market research and data mining. He is a member of the International Neural Network Society, Japan Society for Fuzzy Theory and Intelligent Informatics and Japan Association for Consumer Studies.

Ryotaro Kamimura obtained his Master's degree in French Literature from Tsukuba University in 1981 and Doctoral degree in Electrical Engineering from Tokai University in 1996. He is currently a professor of ICT Education Center and School of Science and Technology of Tokai University. His research interests covers areas related to information-theoretic neural networks.



Osamu Uchida received his B.E. degree from Meiji University, Kawasaki, Japan in 1995, M. Info. Sci. degree from Japan Advanced Institute of Science and Technology, Ishikawa, Japan in 1997, and Ph.D. from University of Electro-Communications, Chofu, Japan in 2000. From 2000 to 2002, he was a research associate at Kanagawa Institute of Technology, Atsugi, Japan. He joined Tokai University in 2002, and since 2007, he has been an associate professor at the Department of Human and Information Science, Tokai University, Hiratsuka, Japan. He was a visiting researcher at the Department of Information and Computer Sciences, University of Hawaii at Manoa in 2014. His research interests include information theory, image processing, Internet technology, information security, natural language processing, and disaster mitigation information systems. Since 2012, he has been a vice editor in chief of the Institute of Image Electronics Engineers of Japan.

Fujio Toriumi received his Ph.D degree from Tokyo Institute of Technology, Japan in 2004. He is an Associate Professor of School of Engineering, The University of Tokyo. His research interests are Agent based simulation, Artificial Market, and Social Data Mining. He is a member of the Institute of Electronics, Information and Communication Engineers of Japan, The Japanese Society for Artificial Intelligence and The Society of Instrument and Control Engineers.