Climate Tracking for Farmland Using Open Data from the Central Weather Bureau in Taiwan

Chin-Shun Hsu, Yung-Hsing Peng, and Po-Chuang Huang
Innovative DigiTech-Enabled Applications and Services Institute, Institute for Information Industry, Kaohsiung City, 80661 Taiwan
Email: {ling9082, pengyh, arvinpchuang}@iii.org.tw

***Abstract*—In agri-business, to keep track of the potential harvest is an important issue to production managers. For this purpose, field observation is necessary for collecting the information of crop growth and climate changes. In Taiwan, the central weather bureau (CWB) is the national authority that provides weather reports of national scale, which is estimated from weather stations established by Taiwan government. The estimated weather of cities or counties in Taiwan can be found on the official website of CWB. To our knowledge, however, there is no official tool for farmers to track and to record the climate data of his/her farmland. In this paper, we utilize the open data of 29 weather stations published by CWB, developing a farmland climate tracking tool for farmers. In this system, we implement four different algorithms for climate estimation, which are the nearest neighbor (NN), the inverse distance weighting (IDW), the Kriging with partial least square regression (KPLS), and the IDW with altitude filter (IDWAF). We compare the performance of these four algorithms by the leave-one-out cross validation on the 29 weather stations with observed climates as open data. The considered types of climate are temperature, humidity, insolation, and wind speed. According to the experimental results, the temperature and insolation are of lower errors in each algorithm, which indicates that field temperature and insolation could be tracked with open data. In addition, IDWAF achieves better accuracies than IDW. Therefore, altitude filtering is a valuable approach to be combined with inverse distance.**

***Index Terms*—climate tracking, farmland, open data.**

I. Introduction

In agriculture, tracking the climate of farmland is a crucial task in production management. To experienced producers, the climate data serves as important information for decision support. For example, the suitable timing for planting, watering, fertilizing, and harvest could be totally different under different climates, even if the species of crop remains unchanged. From the perspective of data science, therefore, the historical climate data and cultivation data on a farmland are of great value for analyzing the potential harvest in the future. By examining such potential harvest from various farmlands, crop production scheduling becomes more systematic to managers, which offers a scientific approach to check if the supply meets the demand. Therefore, many previous researches focus on finding the relationship between climate and harvest [1][2][3][4]. For finding linear and non-linear relationships, the partial least square (PLS) regression [5], the multi-degree polynomial model [6][7], and the artificial neural network [8][9] are approaches commonly adopted.

To obtain the required climate data, there are two possible approaches. The first is to establish a weather station close to the farmland, by which the climate can be accurately observed and recorded. However, this approach is often research-driven, and is definitely limited by the cost. In Taiwan, most farmers do not have such station on his/her farm. Besides, to a common farmer the cost for owning such weather station is not acceptable. Therefore, for real application, it is recommended to obtain the required data by the second approach: to estimate the farmland climate by interpolation. For climate interpolation, the nearest neighbor (NN) [10], the inverse distance weighting (IDW) [11], and the Kriging method [12][13] are well-known techniques applied in previous research. In Taiwan, the central weather bureau (CWB) is the national authority that provides weather reports, and currently there are 29 official weather stations with their historical climates recorded as open data. Inspired by these resources, we would like to estimate the climate of some specified farmland by interpolations with the open data of CWB, developing a new management tool for agricultural decision support. To achieve this goal, therefore, the feasibilities of various interpolation algorithms are worth of investigation.

In this paper, we implement and compare the performance of four interpolation algorithms, which are the nearest neighbor (NN), the inverse distance weighting (IDW), the Kriging with partial least square regression (KPLS), and the IDW with altitude filter (IDWAF). The considered climates are temperature(℃), humidity(%), insolation(MJ/m^2), and windspeed(m/s). The open data collector and interpolation algorithms are implemented on the Smart Agri-management Platform (S.A.M.P.), which is a system offering agricultural cloud services developed by the Institute for Information Industry (III). In our experiment, we perform the leave-one-out cross validation on the 29 official weather stations, comparing the accuracies of different algorithms to different types of

Manuscript received June 15, 2015; accepted September 1, 2015.

© 2015 J. Adv. Inf. Technol.
doi: 10.12720/jait.6.4.227-232

climate. The organization of this paper is as follows. In Section II, we provide preliminary explanations to notations and algorithms. In Section III, we propose the experimental results obtained from S.A.M.P. Finally, we give conclusions and some future studies in Section IV.

II. METHODOLOGY

In this section, we first give the required notations for representing climate data. After that, the adopted four algorithms for field climate estimation are briefly described, which are the nearest neighbor (NN), the inverse distance weighting (IDW), the Kriging with partial least square regression (KPLS), and the IDW with altitude filter (IDWAF).

A. Notations

Let $s_1, s_2, ..., s_n$ be the n weather stations considered in this paper. For each weather station, we assume the number of types of the observed climate is m. In this paper, we have $m = 4$ because there are four types of climate considered, which are temperature, humidity, insolation, and windspeed. In addition, let C_t be an $n \times m$ matrix that stores the climate data of each weather station at time t. In C_t, each element $c_{t,i,j}$ denotes the jth climate value observed from the ith weather station at time t. For identifying the location of stations, let lon_i, lat_i, and alt_i be the longitude, the latitude, and the altitude of s_i, respectively. In addition, we assume the target field F (whose climate is to be estimated) is located at longitude F_{lon}, latitude F_{lat}, and altitude F_{alt}. From here on, we use the triplet (F_{lon}, F_{lat}, F_{alt}) to denote the location of F. That is, the distance $D(s_i, F)$ between s_i and F is computed by the Euclidean distance between (lon_i, lat_i, alt_i) and (F_{lon}, F_{lat}, F_{alt}). Finally, we use $\hat{c}_{t,F,j}$ to denote the jth estimated climate of the target field F at time t.

B. The Nearest Neighbor (NN)

The nearest neighbor [10] is an intuitive approach that estimates the field climate by the observed climate data of the weather station closest to the field. To estimate the climate of F, the NN algorithm can be implemented as follows.

Step 1: Compute the distance $D(s_i, F)$ between s_i and F for $1 \le i \le n$.

Step 2: Determine the station s_u that is closest to F (with the minimal $D(s_u, F)$).

Step 3: Apply $\hat{c}_{t,F,j} = c_{t,u,j}$ for $1 \le j \le m$.

C. The Inverse Distance Weighting (IDW)

The inverse distance weighting technique [11] is a well known approach that puts different weights to stations according to their distance the target field. Different from the NN algorithm, the IDW algorithm estimates the field climate by interpolation of many stations, with a reasonable weighting concept that stations close to the field are of greater influences. Let $I(s_i, F)$ be the inverse distance weight for station s_i when estimating the climate of field F. The IDW algorithm can be implemented as follows.

Step 1: Compute the distance $D(s_i, F)$ between s_i and F for $1 \le i \le n$.

Step 2: Compute $I(s_i, F) = \dfrac{1}{D(s_i, F)^p}$.

Step 3: Compute $\hat{c}_{t,F,j} = \dfrac{\sum_{i=1}^{n} I(s_i, F) \times C_{t,i,j}}{\sum_{i=1}^{n} I(s_i, F)}$ for $1 \le j \le m$.

In the IDW algorithm, the variable p is used to control the inverse weight, which is often set from 1 to 3.

D. The Kriging with Partial Least Square (KPLS)

In geostatistics, the Kriging method [12][13] gives the best linear unbiased prediction for an unmeasured point (such as the field F in this paper) by optimizing the weights of stations with the technique of regression. Therefore, the climate of F can be expressed as

$$\hat{c}_{t,F,j} = w_{j,F,1} \times c_{t,1,j} + w_{j,F,2} \times c_{t,2,j} + \ldots + w_{j,F,1} \times c_{t,n,j}, \quad (1)$$

where each $w_{j,F,i}$ is the optimized weight for station s_i, which is determined by regression over historical data. Since there could be no historical data on F, an alternative approach is to treat each s_i as F, solving the following n regressions.

$$c_{t,1,j} = 0 \times c_{t,1,j} + w_{1,2,j} \times c_{t,2,j} + w_{1,3,j} \times c_{t,3,j} + \ldots + w_{1,n,j} \times c_{t,n,j},$$
$$c_{t,2,j} = w_{2,1,j} \times c_{t,1,j} + 0 \times c_{t,2,j} + w_{2,3,j} \times c_{t,3,j} + \ldots + w_{2,n,j} \times c_{t,n,j},$$
$$c_{t,3,j} = w_{3,1,j} \times c_{t,1,j} + w_{3,2,j} \times c_{t,2,j} + 0 \times c_{t,3,j} + \ldots + w_{3,n,j} \times c_{t,n,j},$$
$$\vdots$$
$$c_{t,n,j} = w_{n,1,j} \times c_{t,1,j} + w_{n,2,j} \times c_{t,2,j} + w_{n,3,j} \times c_{t,3,j} + \ldots + 0 \times c_{t,n,j}.$$

Note that the historical climate data from stations at various time t can be used as the sample data for regression. By the above formula, we can derive n weighting vectors, each of size n, for estimating $c_{t,1,j}$, $c_{t,2,j}$, ..., $c_{t,n,j}$. To estimate $c_{t,i,j}$ from stations other than s_i, the weight $c_{t,i,j}$ of is set to zero. There are total $n \times m$ such weighting vectors for regression because $1 \le j \le m$. Here we adopt the partial least square technique (PLS) [5] to accomplish the regression, because of its good performance in previous work [14][15]. The detailed explanation of PLS is omitted, because it is beyond the scope of this paper. For ease of implementation, we utilize the Weka package [16] to accomplish the PLS

© 2015 J. Adv. Inf. Technol.

regression. The obtained $n \times m$ weighting vectors can be realized as the relationship of climates between stations. With these vectors, the climate of some field F can be estimated by first computing the $1 \times n$ weighting vector W_F for F by spacial interpolation, and then multiply W_F with observed climates from all weather stations.

E. The IDW with Altitude Filter (IDWAF)

Compared with the Kriging method, the IDW algorithm has both merits and shortcomings. In general, IDW is a much simpler algorithm for implementation, which also offers a reasonable weighting scheme with inverse distances. However, unlike the Kriging method, the weights derived from IDW may fail to reflect the true relationships between stations. For example, when applying the Euclidean distance to compute $D(s_i, F)$ between s_i (located at (lon_i, lat_i, alt_i)) and F (located at $(F_{lon}, F_{lat}, F_{alt})$), the longitude and the latitude usually dominates $D(s_i, F)$, and the difference between alt_i and F_{alt} becomes insignificant. This phenominen would cause an improper ignorance of the environmental lapse rate (ELR). To reduce such problem but keep the efficiency of IDW, we propose an enhanced IDW algorithm with an altitude filter, which ignores the stations with altitude much higher or lower than F. Let σ_{alt} be the standard deviation for the altitudes of all stations, and α be the parameter for controling the altitude filter. The newly proposed algorithm is implemented as follows, in which some stations are ignored by setting the distance to infinity.

Step 1: Compute the distance $D(s_i, F)$ between s_i and F for $1 \le i \le n$.

Step 2: Set $D(s_i, F) = \infty$ if $|alt_i - F_{alt}| > \alpha\sigma_{alt}$.

Step 3: Compute $I(s_i, F) = \dfrac{1}{D(s_i, F)^p}$.

Step 4: Compute $\hat{c}_{t,F,j} = \dfrac{\sum_{i=1}^{n} I(s_i, F) \times C_{t,i,j}}{\sum_{i=1}^{n} I(s_i, F)}$ for $1 \le j \le m$.

If Step 2 happens to ignore all stations, the station with altitude closest to F will be picked as the climate indicator, whose observed data is treated as the climate of F. In this paper, we set $\alpha = 1$.

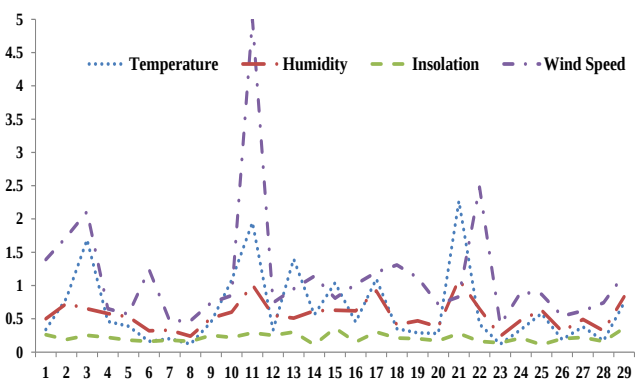
III. Experiments

In our experiment, we adopt the hourly records from 29 official stations of CWB as the source of climate data. The information of these 29 stations is shown in Table I.

We perform the leave-one-out cross validation on these 29 stations, and the period is from January 1st, 2015 to March 31st, 2015. Note that the KPLS algorithm requires historical data for optimizing the weights of stations. Hence, we utilize the climate data from January 1st, 2014 to March 31st, 2014 as the training data. The error curves of the four algorithms over 29 stations are summarized in Fig. 1, in which the errors of estimation is further averaged and quantified as a ratio to the standard deviation of the real observed climate. From Fig. 1, one can see that the errors of temperature and insolation are less than those of other climates. In addition, the KPLS and IDWAF achieve better results.

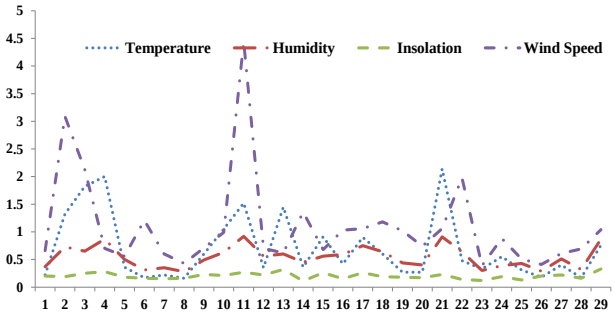
TABLE I. The Profile of 29 Weather Stations.

No.	Name in English	Longitude	Latitude	Altitude
1	Dawu	120.895	22.357	8.1
2	Sun Moon Lake	120.908	23.881	1017.5
3	Yushan	120.951	23.489	3844.8
4	Chenggong	121.365	23.099	33.5
5	Zhuzihu	121.536	25.165	607.1
6	Yilan	121.748	24.765	7.2
7	Dongjidao	119.659	23.258	43
8	Banqiao	121.433	24.999	9.7
9	Hualien	121.604	23.976	16
10	Kinmen	118.289	24.407	47.88
11	Alishan	120.804	23.51	2413.4
12	Hengchun	120.738	22.005	22.1
13	Matsu	119.923	26.169	97.8
14	Kaohsiung	120.308	22.567	2.3
15	Keelung	121.732	25.134	26.7
16	Wuqi	120.515	24.258	31.73
17	Tamsui	121.44	25.165	19
18	Pengjiayu	122.071	25.629	101.7
19	Hsinchu	121.006	24.83	26.9
20	Xinwu	121.047	25.006	20.6
21	Chiayi	120.424	23.497	26.9
22	Taichung	120.675	24.147	84.04
23	Taipei	121.506	25.04	5.3
24	Taitung	121.146	22.754	9
25	Tainan	120.204	22.993	40.8
26	Penghu	119.555	23.567	10.7
27	Anbu	121.52	25.186	825.8
28	Su-ao	121.86	24.601	24.9
29	Lanyu	121.55	22.038	324

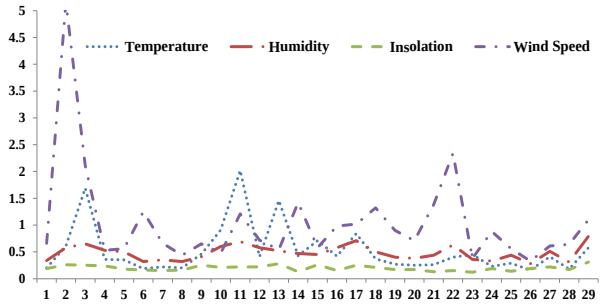


(a) The error of NN over 29 stations.

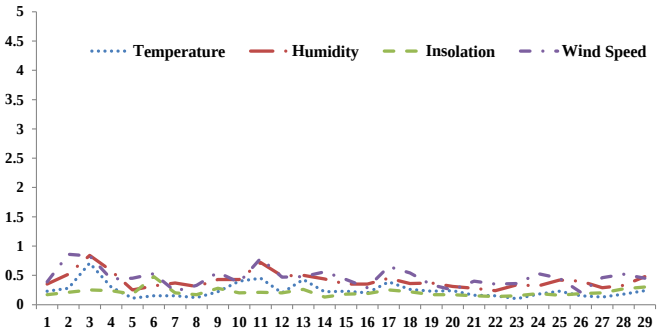
© 2015 J. Adv. Inf. Technol.



(b) The error of IDW over 29 stations.



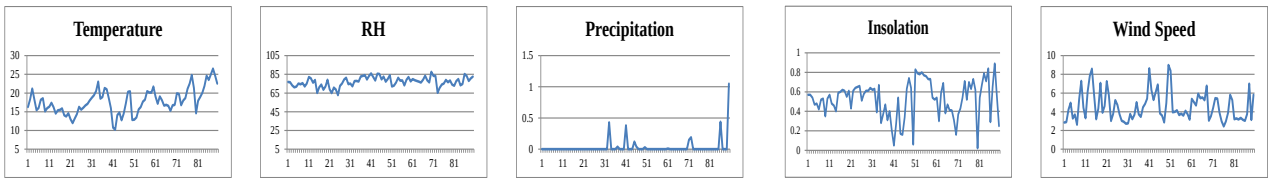
(c) The error of IDWAF over 29 stations.



(d) The error of KPLS over 29 stations.

Figure 1. The quantified errors of four algorithms.

(a) The 29 weather stations and the example farmland under management.



(b) The estimated climates of the example farmland with 29 stations from January 1st, 2015 to March 31st, 2015.

Figure 2. The farmland managment with climate estimations using S.A.M.P.

In Fig. 2, we provide an example for managing a farmland located in the county of Chiayi using the S.A.M.P. With the cloud service of S.A.M.P., the locations of weather stations and farmlands are visualized for ease of management, as shown in Fig. 2(a). In addition, we adopt the KPLS algorithm to estimate the climate of

© 2015 J. Adv. Inf. Technol.

the target farmland, as shown in Fig. 2(b). From our experiment, we notice that insolation and temperature are the two types of climate most suitable to be estimated by interpolations. To illustrate the detailed performance of the four interpolation algorithms, in Fig. 3 and Fig. 4 we show the comparison of IDWAF and KPLS for the temperature and insolation of Chiayi station, which is the station closest to the example farmland. One can see that KPLS is the most reliable algorithm, and IDWAF is an alternative algorithm for reducing the complexity of computation.

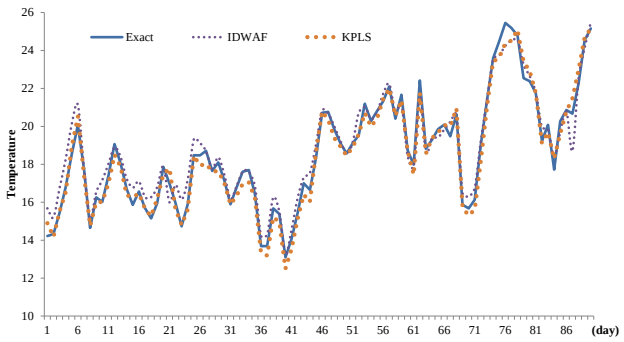


Figure 3. The temperature of Chiayi (No. 21) station.

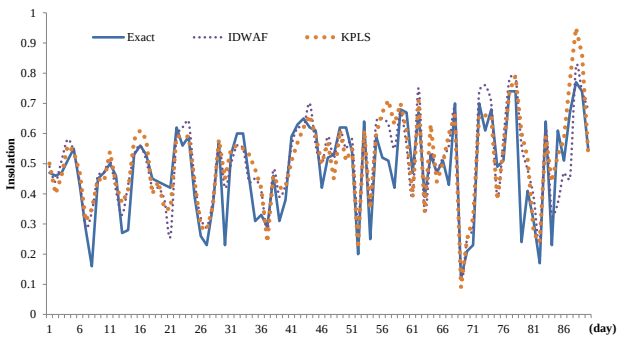


Figure 4. The insolation of Chiayi (No. 21) station.

IV. Conclusion

In this paper, we devise an agri-management tool for tracking the climate of farmland, which is implemented on the S.A.M.P. There are four interpolation algorithms available in this tool, and the performances of them are verified by the open data obtained from 29 official stations of CWB in Taiwan. According to our experimental results, the temperature and insolation are suitable climates to be estimated by interpolation algorithms, which help to reduce the cost of building weather stations. In addition, the KPLS algorithm is recommended if the historical climate data is available. At the same time, the IDWAF algorithm is a good alternative without consideration to historical data. For future study, we will utilize non-linear techniques [17] to improve the Kriging approach. We will also apply our result to find the relationship between climates and yields, designing a decision support system for production management.

Acknowledgment

This study is conducted under the "Online and Offline integrated Smart Commerce Platform (2/4) Project" of the Institute for Information Industry which is subsidized by the Ministry of Economy Affairs of the Republic of China. The authors would like to express their appreciation to the staff of central weather bureau for their support on maintaining and publishing the open data of climates in Taiwan, which benefits the development of the Smart Agri-management Platform (S.A.M.P.).

References

[1] F. L. Tao, M. Yokozawa, J. Y. Liu, and Z. Zhang, "Climate–crop yield relationships at provincial scales in China and the impacts of recent climate trends," *Climate Research*, vol. 38, pp. 83-94, 2008.

[2] G. Koufos, T. Mavromatis, S. Koundouras, N. M. Fyllas, and G. V. Jones, "Viticulture–climate relationships in Greece: The impacts of recent climate trends on harvest date variation," *International Journal of Climatology*, vol. 34, pp. 1445-1459, 2014.

[3] F. L. Tao, Z. Zhang, D. P. Xiao, S. Zhang, R. P. Rötter, W. J. Shi, Y. J. Liu, M. Wang, F. S. Liu, and H. Zhang, "Responses of wheat growth and yield to climate change in different climate zones of China, 1981–2009," *Agricultural and Forest Meteorology*, vol. 189-190, pp. 91-104, 2014.

[4] Y. H. Peng, C. S. Hsu, and P. C. Huang, "Harvest forecasting with environmental information for cucumbers cultivated in net houses," *Journal of Advanced Agricultural Technologies*, vol. 2, pp. 83-87, 2015.

[5] R. Rosipal and N. Krämer, "Overview and recent advances in partial least squares," *Subspace, Latent Structure and Feature Selection, Lecture Notes in Computer Science*, vol. 3940, pp. 34-51, 2006.

[6] D. C. Montgomery, *Design and Analysis of Experiments*, 2009, New Jersey: John Wiley & Sons, Inc.

[7] T. Yoshida, S. Tsubaki, Y. Teramoto, and J. I. Azuma, "Optimization of microwave-assisted extraction of carbohydrates from industrial waste of corn starch production using response surface methodology," *Bioresource Technology*, vol. 101, pp. 7820-7826, 2010.

[8] D. L. Ehret, B. D. Hill, T. Helmer, and D. R. Edwards, "Neural network modeling of greenhouse tomato yield, growth and water use from automated crop monitoring data," *Computers and Electronics in Agriculture*, vol. 79, pp. 82-89, 2011.

[9] D. L. Ehret, B. D. Hill, D. A. Raworth, and B. Estergaard, "Artificial neural network modelling to predict cuticle cracking in greenhouse peppers and tomatoes," *Computers and Electronics in Agriculture*, vol. 61, pp. 108-116, 2008.

[10] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, pp. 21-27, 1967.

[11] G. Y. Lu and D. W. Wong, "An adaptive inverse-distance weighting spatial interpolation technique," *Computers & Geosciences*, vol. 34, pp 1044-1055, 2008.

[12] M. A. Oliver and R. Webster, "Kriging: A method of interpolation for geographical information systems," *International Journal of Geographical Information Systems*, vol. 4, pp. 313-332, 1990.

[13] G. Laaha, J. O. Skøien, and G. Blöschl, "Spatial prediction on river networks: Comparison of top-kriging with regional regression," *Hydrological Processes*, vol. 28, pp. 315-324, 2014.

[14] M. T. Reza, W. Becker, K. Sachsenheimer, and J. Mumme, "Hydrothermal carbonization (HTC): Near infrared spectroscopy and partial least-squares regression for determination of selective components in HTC solid and liquid products derived from maize silage," *Bioresource Technology*, vol. 161, pp. 91-101, 2014.

[15] Y. H. Peng, C. S. Hsu, P. C. Huang, and Y. D. Wu, "An effective wavelength utilization for spectroscopic analysis on orchid chlorophyll measurement," in *Proc. the 2014 IEEE International Conference on Automation Science and Engineering*, Taipei, Taiwan, Aug. 18-21, 2014, pp. 716-721.

[16] I. H. Witten, E. Frank, L. Trigg, M. Hall, G. Holmes, and S. J. Cunningham, "Weka: Practical machine learning tools and

© 2015 J. Adv. Inf. Technol.

techniques with Java implementations," in *Proc. the ICONIP/ANZIIS/ANNES'99 Workshop on Emerging Knowledge Engineering and Connectionist-Based Information Systems*, 1999, pp. 192-196.

[17] T. Yoshida, S. Tsubaki, Y. Teramoto, and J. I. Azuma, "Optimization of microwave-assisted extraction of carbohydrates from industrial waste of corn starch production using response surface methodology," *Bioresource Technology*, vol. 101, pp. 7820-7826, 2010.

Chin-Shun Hsu received B.S. in electronic and computer engineering from National Taiwan University of Science and Technology and M.S. in computer science and engineering from National Sun Yat-sen University, Taiwan in 1999 and 2001, respectively. He is currently an R&D engineer in the Institute for Information Industry (III), and a Ph.D candidate of the institute of computer and communication engineering in National Cheng Kung University. His research interests include data mining, evolutionary algorithms, and network topology.

Yung-Hsing Peng received his B.S. and the M.S. degree in computer science and engineering from National Sun Yat-sen University, Kaohsiung, Taiwan, in 2003 and 2004, respectively. Then, he received the Ph.D. degree in computer science and engineering from National Sun Yat-sen University in 2010. He is currently an R&D engineer in the Institute for Information Industry (III). His research interests include data mining, evolutionary algorithms, sequence analysis, and pattern matching.

Po-Chuang Huang received his B.S. and the M.S. degree in Computer Science and Engineering from National Chen Kung University, Tainan, Taiwan, in 1999 and 2001, respectively. Then, he received the Ph.D. degree in computer science from National Cheng Kung University in 2010. He is currently an R&D engineer in the Institute for Information Industry (III). His research interests include machine learning, evolutionary algorithms, and pattern matching.

© 2015 J. Adv. Inf. Technol.