Developing Agriculture Purchasing Managers' Index for Describing Taiwan's Agriculture Industry by Using Automatic Weighted *k*-means Algorithm

Tzong-Ru Lee[1] and Chien-Pang Lee[2,3]*

[1]Department of Marketing, National Chung Hsing University, Taichung City, Taiwan
[2]Department of Maritime Information and Technology, National Kaohsiung University of Science and Technology, Kaohsiung City, Taiwan
[3]Master's Program in Offshore Wind Energy Engineering, National Kaohsiung University of Science and Technology, Kaohsiung City, Taiwan

Received: 7 October 2020, Revised: 21 January 2021, Accepted: 18 February 2021

Abstract

Although the average agricultural total output is not greater than 2% of Taiwan's GDP, the Taiwan government still attaches great importance to the agriculture industry to ensure the food self-sufficiency rate. Taiwan still has no indicators to measure the status of the agriculture industry. This paper proposes an idea to develop Agriculture Purchasing Managers' Index (APMI) for Taiwan agriculture industries. To reduce the effect of some statistical assumptions and to provide more clarity and direct analysis of results, this paper proposes a novel automatic weighted *k*-means algorithm to develop the APMI. The results of this research suggest that four variables should be included in the APMI of the pig industry, namely "Trade amount", "Average weight per pig", "Price per pig", and "Slaughtered.". Among these, "Slaughtered" and "Trade amount" are the more important variables for developing the APMI of the pig industry. The proposed model offers three advantages: (a) it can be successfully used to construct APMI, (b) It can automatically search the weight of each variable without any human judgment in APMI, and (c) It avoids some statistical assumptions and explains the results more clearly and directly. Thus, the proposed model can be used to construct used APMI proposed in this work, and it describes the status of the agriculture industry.

Keywords: pig industry; trade of pigs; automatic weighted selection
DOI 10.14456/cast.2021.49

1. Introduction

Due to the geographical environment of Taiwan, agriculture is one of the first developed industries in Taiwan. Although the primary industry of Taiwan has changed as the technology industry in recent years, the agricultural industry is still important in Taiwan. According to the official document of the "National Accounts Yearbook" of Taiwan in Jan. 2021, the average agricultural

*Corresponding author: Tel.: (+886) 7-8100888 ext. 25322
E-mail: cplee@nkust.edu.tw

total output is about 520 billion New Taiwan dollars. That is not greater than 2% of Taiwan's GDP in recent years. Accordingly, the Taiwan government still attaches great importance to the agriculture industry to ensure the food self-sufficiency rate.

Because the agriculture industry plays a vital role in the stable operation of national economies [1, 2], many researchers focus on agriculture industry issues, especially in forecasting agricultural output [3-7]. Furthermore, there is an absence of an agricultural index to describe the macro-economic status of the market. Agricultural companies and government departments have no information except for agricultural gross domestic profit (GDP) to measure the current agriculture industry [8]. Therefore, this paper proposes the idea of developing an agriculture index to describe the agriculture industry's current status.

In the non-agriculture industry, Purchasing Managers' Index (PMI) and Non-Manufacturing (NMI) are two well-known indexes to describe industries' economic status. PMI and NMI are published by the two leading institutions, which are Markit and Institution for Supply Management (ISM). PMI is released ahead of other official indexes and is different from the measurement standards. It is used to evaluate the validity and reliability of the objective; for the reason that there are generally no "standards" of economic activity [8]. Although some countries have been releasing NMI, including the agriculture industry, NMI does not include agriculture industries in Taiwan. Even though the two indexes are not suitable to use directly for describing the status of the agriculture industry, some characteristics or features of them are suitable for us to search for the agriculture indexes.

Since 2011, PMI and NMI are published by the Chung-Hua Institution for Economic Research in Taiwan [9]. Recently, Taiwan Manufacturing PMI has been released monthly. The indicators of PMI and NMI in Taiwan include "New orders", "Production", "Employment Level", "Supplier Deliveries", "Inventory of Purchase Materials", and so on. Many researchers have mentioned five important features in PMI [10-13]: (a) Validity and Reliability; PMI is used to evaluate validity and reliability. It is not used to measure the standard of economic activity, (b) Timing; all indexes are indicators of monthly variation and are calculated as diffusion indexes. For example, new orders of this month and the last month are compared whether the orders increase, decrease or no change, (c) Timeliness; if an index is irregularly releasing schedule, it does not help reference, (d) Stability; the index must be the limits of random fluctuations or a small random component compared to the trend and cyclical components of the activity measured, (e) No Revisions; if the index is released, it is not allowed to be revised.

Because PMI and NMI indicators have not been easy to collect in Taiwan's agricultural industries, a new index has to be developed for describing Taiwan's agricultural economic status. Accordingly, the purpose of this paper is to develop the Agriculture Purchasing Managers' Index (APMI) by selecting the main agriculture indexes. Although many statistical methods are used to solve agriculture-related issues, this paper uses data mining techniques to reduce the effects of certain statistical assumptions and to provide more clarity and direct analysis of results. Since the data is suitable for unsupervised learning algorithms, this paper proposes a novel model based on a k-means algorithm called the automatic weighted k-means algorithm, to search the weight of each agricultural variables without any human judgment for developing the APMI.

There are many kinds of agricultural industries in Taiwan. This paper chooses the pig industry as an example to develop the APMI of the pig industry because the annual export value of the pig industry was at 170 billion New Taiwan dollars before 1997. However, the industry has been until recently unable to export now due to the impact of foot-and-mouth disease. Fortunately, there is no a need to vaccinate at present, and exports are expected to resume in 2020. It is conceivable that the future pig industry will become one of the primary agricultural industries in Taiwan and this is the main reason for this paper to develop the APMI of the pig industry in Taiwan.

2. Materials and Methods

2.1 Materials

This paper uses data from the pig industry in Taiwan to develop the APMI for the industry. The data were collected from the Annual Report of the Council of Agriculture, Executive Yuan (Taiwan) and included six variables, such as, “Trade amount”, “Total weight”, “Average weight per pig”, “Price per pig”, “Slaughtered”, and “Total revenue”. The meaning of the variables are described as follows:

a. Trade amount: The number of pigs traded in a month; the unit is the number of “head”.

b. Total weight: The total weight of pigs traded in a month; the unit is “kg/head”.

c. Average weight per pig: The average weight of whole traded pigs in a month; the unit is “metric ton”.

d. Price per pig: The average price per traded pig in a month; the unit is “NT$/ton”.

e. Slaughtered: The number of pigs slaughtered in a month; the unit is the number of “head”.

e. Total revenue: The total revenue of traded pigs in a month; the unit is “NT 10 thousand dollars”.

2.2 The *k*-means algorithm

As described in the Introduction, the characteristic of the data used for developing the APMI is that there is no direct output variable. In other words, the data lends itself to an unsupervised learning algorithms. This paper uses the *k*-means algorithm as the main algorithm to search the weight of each agricultural variable and to develop the APMI. The proposed model is called the automatic weighted k-means algorithm. To understand the proposed model’s main algorithm, we now introduce the relevent theory and definitions.

The *k*-means algorithm, which was first proposed in 1967 [14], is a well-known cluster algorithm in data mining. The main idea of the *k*-means algorithm is to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean [15]. Euclidean distance is usually used to measure the distance *n* observations belonging to *k* clusters. We give a simple example to explain the algorithm of the *k*-means. Suppose a dataset $D=\{x_1, x_2,..., x_n\}$ is used to cluster to *k* groups. The procedure of the *k*-means algorithm includes four main steps.

Step 1. Randomly select *k* initial seeds. Subsequently, the *k* initial seeds are used as the centroids of the initial *k* clusters.

Step 2. Calculate the distance between each data point (observation) and *k* centroids. Subsequently, each data point (observation) is assigned to the cluster that is the smallest distance away.

Step 3. Create the new *k* centroids of the clusters by calculating the mean of the clusters.

Step 4. Iterate Step 2 and Step 3 until the clusters stop changing or satisfy stop conditions.

Since the Euclidean distance is usually used in the *k*-means algorithm, the importance of each variable is the same. However, each variable should have a different influence on different issues in the real world. Accordingly, many different weighted *k*-means algorithms were proposed to counter this problem [16, 17]. This paper also follows the idea to propose a novel model, which is introduced in Section 2.3, to develop the APMI

2.3 Methods

As described in the Introduction, this paper is concerned with the development of the APMI from the agricultural variables that are to do with the prosperity of agriculture. However, the effects of agricultural variables in the APMI vary. Thus, this paper proposes a hybrid model to counter this problem. The proposed model's main idea is based on the use of the *k*-means algorithm to search each variable's weight without any human judgment in the APMI. The proposed model consists of two components, data preprocessing and the automatic weighted *k*-means algorithm. Because the unit of each agricultural variable is different, each variable has to perform data preprocessing to reduce the effect of the unit when building the proposed model. Subsequently, this paper uses the automatic weighted *k*-means algorithm to automatically determine the number of clusters (k) and each agricultural variable's weight. The proposed model's detailed procedures are shown in Figure 1 and are discussed in the following sections.

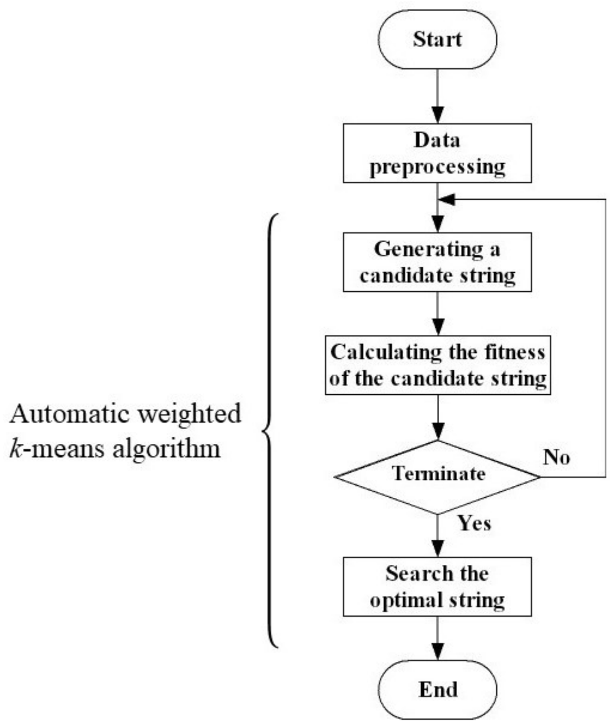

Figure 1. The flowchart of the proposed model

2.3.1 Data preprocessing

As described in the above sections, the unit of each agricultural variable is different. For example, the unit of trade amount is "head", and the unit of the price per pig is "NT$/ton". The different units would affect the proposed model's correctness, especially in calculating Euclidean distance, which is the primary measuring method in the *k*-means algorithm. Thus, each agricultural variable has to be preprocessed before building the proposed model. The equation of the data preprocessing is shown in Eq. (1).

$$PV_{in} = \frac{V_{in} - V_{i(n-1)}}{V_{i(n-1)}} \tag{1}$$

In Eq. (1), V_{in} and $V_{i(n-1)}$ denote the i^{th} variable of the n^{th} month's value and the i^{th} variable of the $(n-1)^{th}$ month's value, respectively; PV_{in} denotes the i^{th} variable of the n^{th} month's value after performing data preprocessing. Moreover, we would directly understand each agricultural variable's ratios (increase or decrease) according to the results of data preprocessing when PV_{in} is positive, which signifies that the value is greater than that of the last month.

2.3.2 Automatic weighted *k*-means algorithm

The automatic weighted *k*-means algorithm includes three steps to automatically determine the number of clusters and each variable's weight. We then use the determined weight of each variable to develop the APMI. The details are as described in Section 2.3.2.1 to Section 2.3.2.3.

2.3.2.1 Generating a candidate set of the number of clusters and the importance of variables

Generally, a set of the number of clusters and the weight of variables would affect the weighted *k*-means algorithms' correctness. However, determining a suitable set of the number of clusters and the weight of variables for the weighted *k*-means algorithms is difficult. Therefore, this paper uses a strategy that imitates chromosomes of the genetic algorithm, a popular optimization algorithm [1, 4, 18], to search for the optimal set of the number of clusters and the weight of variables.

In the strategy, we imitate chromosomes of the genetic algorithm to generate many strings as the candidate sets of the number of clusters and the weight of variables for the weighted *k*-means algorithms. However, the proposed model uses an integer encode method to generate a string (chromosome) for a set of the number of clusters and the weight of variables. To reduce the computing complexity, this step utilizes the importance of variables to temporarily replace the weight of variables.

A candidate string is generated in the following procedure. Firstly, a candidate string is randomly generated, and the string's structure is determined, as shown in Figure 2. Figure 2 shows that the last bit (g_k) denotes the number of clusters and ranges from 2 to 5. The remainder bits (g_1 to g_n) denote the importance of variables and range from 0 to 10. Subsequently, a candidate string is generated based on the above strategy.

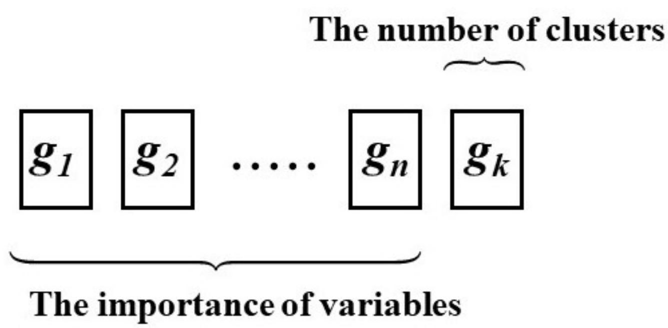

Figure 2. The structure of a candidate string

2.3.2.2 Generating a candidate set of the number of clusters and the importance of variables

After constructing a set of the number of clusters and the importance of variables, the importance of variables has to transform as each variable's weight format for performing Euclidean distance in the weighted *k*-means algorithm. Eq. (2) is used to transform each variable's importance into each variable's weight format. In Eq. (2), W_i and g_i denote the weight and the importance of the i^{th} variable, respectively. Subsequently, the algorithm of *k*-means according to Eq. (3) for minimizing the within-cluster sum of squares is performed. In Eq. (3), C_i denotes the i^{th} cluster; X_j denotes the j^{th} observation; $\overline{C_i}$ denotes the mean of C_i; W denotes the weight vector of the variables.

We next define a fitness function to measure the performance according to the characteristic of the *k*-means algorithm. Accordingly, Eq. (4) is defined as the proposed model's fitness function and is used to measure the performance of the proposed model. In Eq. (4), "SST" denotes the total sum of squares, and "SSB" denotes the sum of squares between clusters. According to the Eq. (4), a candidate string with higher fitness means that its performance is better than the others.

$$W_i = \frac{g_i}{\sum_{j=1}^{n} g_j} \tag{2}$$

$$\arg\min_{C} \sum_{i=1}^{k} \sum_{X_j \in C_i} W \left\| X_j - \overline{C_i} \right\|^2 \tag{3}$$

$$fitness = \frac{SSB}{\mathrm{SST}} \tag{4}$$

2.3.2.3 Search the optimal set the number of clusters and the weight of variables

Section 2.3.2.1 and Section 2.3.2.2 are analogous to the generation of genetic algorithms. In the proposed model, Section 2.3.2.1 and Section 2.3.2.2 have to be performed many times to generate many candidate sets of the number of clusters and the weight of variables. After generating many candidate strings, we then compare the fitness value of each candidate string. The string with the best fitness value is the optimal string. Finally, the optimal string is used to determine the weights of the agricultural variables for developing the APMI.

3. Results and Discussion

3.1 Data

This paper uses the pig trade in Taiwan as an example. The monthly data, including six variables from 12/2011 to 12/2018 obtained from the Annual Report of the Council of Agriculture, Executive Yuan (Taiwan), are then collected to develop the APMI by the proposed model. Table 1 shows a part of the collected data. After collecting the data, the data have to be preprocessed using Eq. (1) before the performance of the automatic weighted *k*-means algorithm. The transformed results are shown in Table 2.

Table 1. A part of the collected data of the traded pigs in Taiwan

Year	Month	Trade amount	Total Weight	Average Weight per pig	Price per pig	Slaughtered	Total revenue
2011	Dec.	665,289	79,981.04	120.22	7,021	402,329	56,155
2012	Jan.	631,948	75,738.97	119.85	6,856	401,495	51,927
2012	Feb.	599,717	72,901.60	121.56	6,080	350,498	44,324
2012	Mar.	700,451	85,630.13	122.25	5,349	403,953	45,804
…	…	…	…	…	…	…	…
2018	Nov.	575,657	71,577.19	124.34	7,156	374,301	51,221
2018	Dec.	581,915	72,622.99	124.80	7,189	384,052	52,209

Table 2. A part of the collected data of the traded pigs in Taiwan after data preprocessing

Year	Month	Trade amount	Total weight	Average weight per pig	Price per pig	Slaughtered	Total revenue
2012	Jan.	-0.050	-0.053	-0.003	-0.024	-0.002	-0.075
2012	Feb.	-0.051	-0.037	0.014	-0.113	-0.127	-0.146
2012	Mar.	0.168	0.175	0.006	-0.120	0.153	0.033
…	…	…	…	…	…	…	…
2018	Nov.	-0.028	-0.018	0.010	0.006	-0.029	-0.012
2018	Dec.	0.011	0.015	0.004	0.005	0.026	0.019

3.2 Descriptive statistics

Table 3 shows the descriptive statistics of variables from 12/2011 to 12/2018 after data preprocessing. Although each variable has increased slightly (mean is greater than 0), each variable's variation (SD) has significantly fluctuated apart from the average weight per pig. Subsequently, we plot line charts of each variable by month for each year to explore the trend or the variation of each variable, as shown in Figure 3. Obviously, each variable's trend in different years is almost the same except for some particular timepoints. For example, the "Trade amount" in March is greater than the "Trade amount" in February for each year; on the contrary, the fluctuation of "Slaughtered" in July 2018 is different from the other years.

To explore the trend or variation among variables, we also calculate the Pearson's correlation coefficient matrix to compare the correlation coefficient between variables, as shown in Table 4. According to Table 4, apart from the price per pig, the correlation coefficients among variables are positive. Furthermore, the correlation coefficients among some variables are close to 1. Accordingly, we only keep four variables ("Trade amount", "Average weight per pig", "Price per pig", and "Slaughtered") close the proposed model for developing the APMI.

Table 3. The descriptive statistics of variables after data preprocessing

	Trade amount	**Total weight**	**Average weight per pig**	**Price per pig**	**Slaughtered**	**Total revenue**
Min	-0.453	-0.450	-0.022	-0.179	-0.468	-0.404
Mean	0.007	0.008	0.001	0.002	0.010	0.009
SD	0.127	0.132	0.011	0.057	0.144	0.143
Median	0.0125	0.014	0.001	-0.004	0.013	0.008
Max	0.383	0.419	0.026	0.169	0.636	0.534

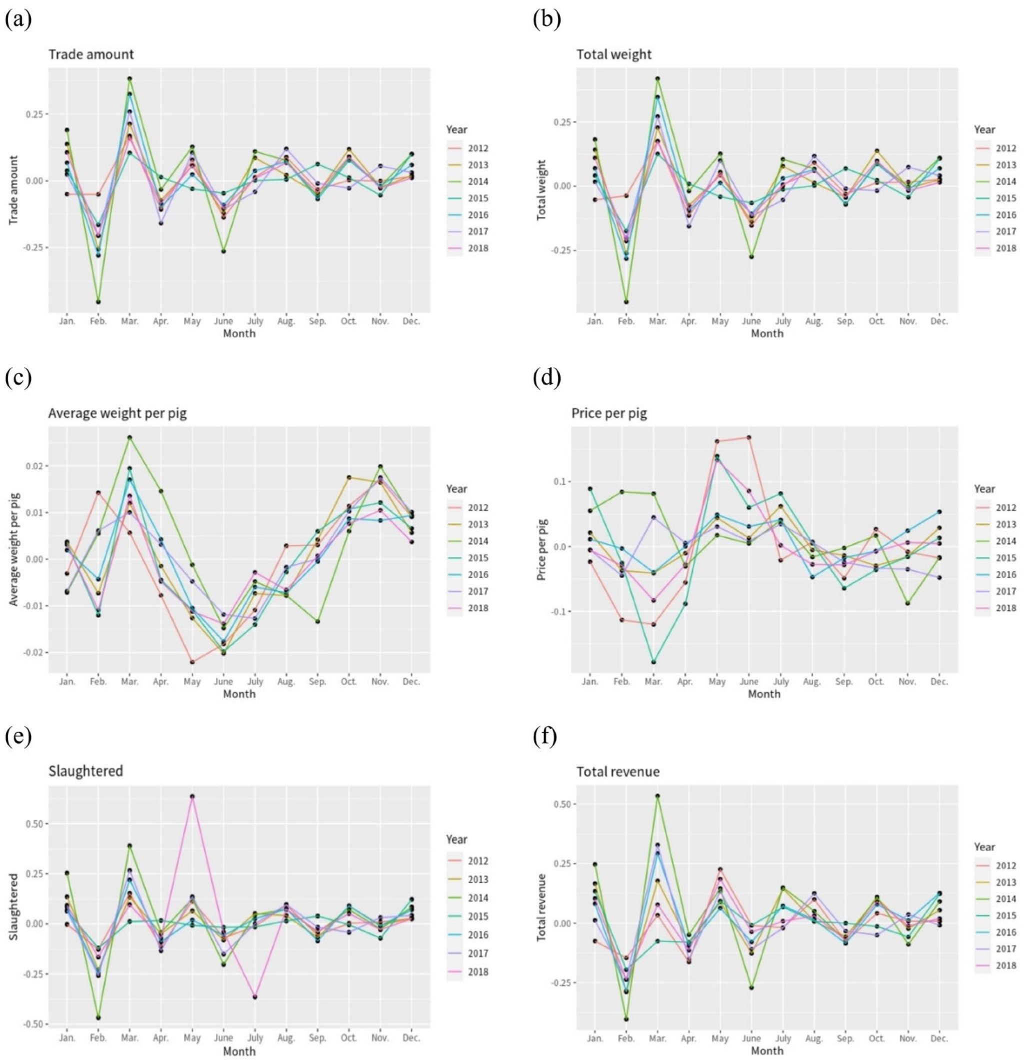

Figure 3. The variation of each variable in different years. (a) Trade amount; (b) Total weight; (c) Average weight per pig; (d) Price per pig; (e) Slaughtered; (f) Total revenue

Table 4. The Pearson's correlation coefficient matrix of the variables

	Trade amount	**Total weight**	**Average weight per pig**	**Price per pig**	**Slaught ered**	**Total revenue**
Trade amount	1.000	0.997	0.365	-0.032	0.821	0.923
Total weight	0.997	1.000	0.436	-0.069	0.807	0.911
Average weight per pig	0.365	0.436	1.000	-0.461	0.161	0.218
Price per pig	-0.032	-0.069	-0.461	1.000	0.193	0.345
Slaughtered	0.821	0.807	0.161	0.193	1.000	0.841
Total revenue	0.923	0.911	0.218	0.345	0.841	1.000

3.3 Parameter setting

Three parameters have to set before performing the proposed model to construct the APMI. Firstly, the number of clusters ranges from 2 to 5 in a candidate string. Secondly, the importance of each variable ranges from 0 to 10. Finally, the number of candidate strings was set at 20,000 candidate strings. Subsequently, we use the parameter setting to perform the proposed model to generate an optimal string.

3.4 Stability analysis

Although the proposed model could search for an optimal string for constructing the APMI, the proposed model still had a disadvantage. Because the candidate strings of the proposed model are generated randomly, the optimal string is different when performing the proposed model at different times. To ensure the proposed model can generate a stable result, the proposed model needs to be performed 2,000 times to generate 2,000 optimal strings, which can and then be used to verify the optimal strings from the proposed model's stability. Table 5 shows the results of the 2,000 optimal weights of variables. Table 5 shows that the mean and median of weights of variables are very close, and their SD are also less than 0.01. That is, the wights of each optimal string should be very close. According to the result, we believe that the result of the proposed model is stable. We could use the proposed model to search the optimal weights of variables for developing the APMI.

Table 5. The result of the optimal weights

	Weights			
	Trade amount	**Average weight per pig**	**Price per pig**	**Slaughtered**
Mean	0.334	0.201	0.086	0.379
Median	0.311	0.202	0.088	0.399
SD	0.093	0.098	0.035	0.091

3.5 Discussion

The average weights of variables are calculated in Section 3.3. Thus, we use the result to develop the APMI, as shown in Eq. (5):

$$\text{APMI}=(0.334X_1+0.201X_2+0.086X_3+0.379X_4)\times 100\% \quad (5)$$

where X_1 denotes the "Trade amount"; X_2 denotes the "Average weight per pig"; X_3 denotes the "Price per pig", and X_4 denotes the "Slaughtered".

According to Eq. (5), the "Slaughtered" and the "Trade amount" are the more important variables in the APMI of the pig industry because their weights are significantly greater than the other two variables. Among them, "Slaughtered" is the most important variable for developing the APMI of the pig industry because it has the largest weight in Eq. (5). The result is reasonable and interpretable because the "Slaughtered" and the "Trade amount" denote the pork market's demand in Taiwan. When pigs are slaughtered immediately after the trade, it represents that the pork market's demand is high in the month. In short, when the "Slaughtered" and the "Trade amount" of pig's trade are increasing, the trade of pig in Taiwan is busier. Accordingly, a higher APMI denotes a higher agricultural economic environment.

4. Conclusions

4.1 Research contributions

Although the average agricultural total output is not greater than 2% of Taiwan's GDP in recent years, the Taiwan government still attaches great importance to the agriculture industry to ensure the food self-sufficiency rate. Accordingly, many researchers are invested in agricultural research issues. Since Taiwan still has no indicators to measure the agriculture industry's status, this paper proposes an idea to develop the APMI for the Taiwan agriculture industry.

According to the results, the APMI of the pig industry has been developed based on four variables: "Trade amount", "Average weight per pig", "Price per pig", and "Slaughtered". Among them, "Slaughtered" and "Trade amount" are the more important variables for developing the APMI of the pig industry, especially "Slaughtered". Accordingly, we could observe the fluctuations of "Slaughtered" and "Trade amount" to judge the pig industry's economic status.

The proposed model offers three advantages: (a) it can be successfully used to develop the APMI, (b) it can automatically search the weight of each variable without any human judgment in APMI, and (c) it is based on data mining techniques that reduce the effects of some statistical assumptions and provide more clarity and direct analysis of the results. Accordingly, the proposed model can be suggested for the development of the APMI and describes well the agriculture industry's status.

4.2 Research limitations and future research

This paper's dataset is the monthly trade of pigs in Taiwan, expressed as six variables, from 12/2011 to 12/2018, obtained from the Annual Report of the Council of Agriculture, Executive Yuan (Taiwan). As described in the Introduction, the Taiwan pig industry announced in 2019 that it was free of foot and mouth disease. Therefore, we give the following suggestions for future research: (a) Future studies should pay attention to the impact of years, especially across 2019, (b) Future studies could include more factors (variables) in the APMI and its development because the agriculture industry may be affected by more external factors, (c) Future studies should include a situation or a time point to update or reconstruct the weightings in the APMI, and (d) Future studies should use the proposed model to develop the APMI for different agriculture industries.

References

[1] Ou, S.-L., 2012. Forecasting agricultural output with an improved grey forecasting model based on the genetic algorithm. *Computers and Electronics in Agriculture*, 85, 33-39.
[2] Xiong, T., Li, C., Bao, Y., Hu, Z. and Zhang, L., 2015. A combination method for interval forecasting of agricultural commodity futures prices. *Knowledge-Based Systems*, 77, 92-102.
[3] Lee, C.-P., 2017. Reduced the risk in agriculture management for government using support vector regression with dynamic optimization parameters. *LEX LOCALIS - Journal of Local Self-Government*, *15*, 243-261.
[4] Jheng, T., Li, T. and Lee, C.-P., 2018. Using hybrid support vector regression to predict agricultural output. *Proceedings of 2018 27th Wireless and Optical Communication Conference (WOCC)*, 30 April-1 May 2018, 1-3.
[5] Nieh, S.-C., Lin, W.-S., Hsu, Y.-H., Shieh, G.-J. and Kuo, B.-J., 2014. The effect of flowering time and distance between pollen source and recipient on maize. *GM Crops & Food*, 5, 287-295.
[6] Higgins, A., Prestwidge, D., Stirling, D. and Yost, J., 2010. Forecasting maturity of green peas: An application of neural networks. *Computers and Electronics in Agriculture*, 70, 151-156.
[7] Lee, C.-P., Shieh, G.-J., Yiu, T.-J. and Kuo, B.-J., 2013. The strategy to simulate the cross-pollination rate for the co-existence of genetically modified (GM) and non-GM crops by using FPSOSVR. *Chemometrics and Intelligent Laboratory Systems*, 122, 50-57.
[8] Tien, T.-L., 2018. *The Trial Compilation of Agriculture Purchasing Managers' Index for Taiwan: The Case of Taiwanese Hog Industry*. Taichung: National Chung-Hsing University.
[9] Chen, S.-H., 2011. *The Compilation of Purchasing Managers' Index for Taiwan*. Taipei: Chung-Hua Institution for Economic Research.
[10] Kauffman, R.G., 1999. Indicator qualities of the NAPM report on business®. *Journal of Supply Chain Management*, 35, 29-37.
[11] Koenig, E.F., 2002. Using the purchasing managers' index to assess the economy's strength and the likely direction of monetary policy. *Economic and Financial Policy Review*, 1(6), 1-14.
[12] Harris, E.S., 1991. Tracking the economy with the purchasing managers' index. *Federal Reserve Bank of New York Quarterly Review*, 16(3), 61-69.
[13] Lindsey, M.D. and Pavur, R.J., 2005. As the PMI turns: A tool for supply chain managers. *Journal of Supply Chain Management*, 41, 30-39.
[14] MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. *Proceedings of Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, California, 281-297.
[15] Li, Y.-g. and Huang, Y.-s., 2013. Product requirements cluster analysis based on K-means. *Proceedings of 20th International Conference on Industrial Engineering and Engineering Management*, Berlin, Heidelberg, 455-461.
[16] Thiyagarajan, R., Thangavel, K. and Rathipriya, R., 2013. Recommendation of web pages using weighted K-means clustering. *International Journal of Computer Applications*, 86, 44-48.
[17] Thangadurai, D.K., Uma, M. and Punithavalli, D.M., 2010. A study on rough clustering. *Global Journal of Computer Science and Technology*, 10(5), 55-58.
[18] Lee, C.-P. and Lin, W.-S., 2016. Using the two-population genetic algorithm with distance-based k-nearest neighbour voting classifier for high-dimensional data. *International Journal of Data Mining and Bioinformatics*, 14, 315-331.