Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 36 (2009) 1014–1019
Expert Systems with Applications
www.elsevier.com/locate/eswa

Price information evaluation and prediction for broiler using adapted case-based reasoning approach

B.W. Huang [a], M.L. Shih [b], Nan-Hsing Chiu [c], W.Y. Hu [a], C. Chiu [d,*]

[a] Department of Applied Economics, National Chung-Hsing University, Taichung, Taiwan
[b] Department of Social Studies Education, National Tai-Tung University, Tai-Tung, Taiwan
[c] Department of Information Management, Ching Yun University, Chungli, Taiwan
[d] Department of Information Management, Yuan Ze University, Chungli, Taiwan

Abstract

Predicting the upcoming broiler market price is important for the producers in developing their production plan. Effective price prediction model can aid producers to prevent over production or production shortage of broilers in advance. This research proposes an adapted CBR approach for predicting broiler price. The results indicate that the proposed adapted CBR approach demonstrates superior prediction performance than un-adapted CBR approach, CART, artificial neural nets and linear regression with at least 50% less of mean average error. This study finds that adjusting the price of the most similar case by considering the similarity distance to the case being predicted is a key to improve the prediction accuracy of the case-based broiler price estimation model.

Keywords: Broiler price; Case-based reasoning; Case adaptation; Genetic algorithms

1. Introduction

After Taiwan joining WTO in 2002, chicken has been gradually permitted to be imported. During the period 2002–2004 import quota was set; however quota-free import policy became effective when quota institution of imported chicken was completely deregulated after 2005. As a result, the broiler industry in Taiwan began to confront with more challenges among which the evaluation and prediction of broiler price become rather complicated. From the viewpoint of broiler growers, analyzing related price information could be the basis in deciding the best timing to buy in chick or to sale broiler. Under the liberalization circumstance, the producers need an effective price prediction model to grasp more accurate information based which a better production plan can be developed.

Broiler growers are accustomed to evaluate the upcoming market price according to their previous experiences and the related production information provided by Poultry Association. In the era of closed economic system that prohibits the chicken to import, the human way of price prediction might be simple and easy to proceed. However, as the chicken import regulation trending toward liberalization like an open economic system, the price information evaluation and prediction will become more difficult. The production planning and management for broiler growers tend to be rather challenging. For example, the heads slaughtered decreased from 207 million in 2004 to 170 million in 2005. Meanwhile, in 2005, the domestic dead pig events were outbreak, and the international chicken market was in higher price level. Therefore, the broiler price of per kilogram went upward 33.7 to 36.8 NT dollars. This implies that growers excessively dropped the production of broiler and induced a rise in price. It reveals that how to evaluate related information and predict the price will become an important issue for the sustainable development of broiler industry in Taiwan.

Confronting with the changeable international chicken market and unpredictable domestic livestock circumstances,

* Corresponding author.
E-mail address: imchiu@saturn.yzu.edu.tw (C. Chiu).

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.11.010

broiler growers greatly concern how to manage the fluctuant broiler price information and predict its trend. It is also one of interesting research topics about broiler industry. In the past those researches related to broiler price, the econometrics models are usually applied to conduct analysis. Nevertheless, these methodologies are constrained by their functional forms and restriction conditions. CBR is a technique that has been proposed for a long time as a valid alternative to expert judgment. The uses of case-based reasoning (CBR) approach has been evaluated and confirmed in many studies (Chiu and Chiu, 2003;Huang and Chiu, 2006). Most of the CBR estimation models retrieve the most similar case for generating estimates (Jorgenson, 2004;Shepperd and Schofield, 1997).

In general a historical case is seldom exact as the case being estimated. Essentially, the reused case needs to be adjusted when the most similar case has a similarity distance with the case being estimated. In this regard, one strategy for adjusting the reused case is to build an adjustment mechanism whose algorithm can derive the optimal adjustment on the reused case. Genetic algorithm (GA) is a stochastic algorithm that uses an adaptive search approach for solving problems (Impedovo, Lucchese, and Pirlo, 2006). Thus this study tries to apply the adapted CBR approach to analyze the broiler price information. Major study of the adapted CBR is to apply GA in the adjustment of the solution part of the reused case.

2. Literature review

The research works on the broiler price analysis usually applied time series models to analyze price fluctuation characteristics or predict the price trend. However, some of the research works such as Holt and Aradhyula (1998) and Kapombe and Colyer (1999) whose models always need to be verified for the stationary price variables, the test of the relationship of current and past prices, and the examining of the heteroskedasticity of residuals. Usually applying those models has to verify if the residuals comply with the basic assumptions of these model's theories. To resolve these concerns, further adjustment of the variables or the functional forms (e.g. first-order difference) is required. Moreover, these kinds of models are assumed as linear forms that become valid to be applied to conduct empirical analysis, but the fluctuation path of broiler price may not coincide with linear function that can affect the model effectiveness. Although some nonlinear models have been developed and applied to empirical analysis (Abdulai, 2002;Goodwin and Piggott, 2001;Sephton, 2003), those are still restricted by the functional forms. Nevertheless, the parameters estimation for the nonlinear models is another issue due to the initial values settings. Consequently, trials of different initial values are inevitable in order to achieve parameters convergence. In practice, the broiler price variation is not easy to be depicted by a specific functional form. Attempting to overcome the limitation of time series models, we propose an adapted CBR approach as an alternative tool for broiler price prediction.

CBR predictors compare the similarities between the case being predicted and all the historical cases that have been evaluated and confirmed in many studies (Beddoe and Petrovic, 2006;Chiu and Chiu, 2003;Huang and Chiu, 2006;Kim, 2003). Basically, a CBR process cycle is composed of four stages: retrieving, reusing, revising and retaining (Mair et al., 2000). The retrieval of appropriate cases relies on a similarity measure that takes into account the distance between pairs of cases. In most of the cases, the most similar case still has a similarity distance with the case being predicted in CBR predictors. For improving the prediction accuracy in the CBR predictor models, an appropriate adjustment approach based on the similarity distances between two cases is required. Establishing an adjustment mechanism whose algorithm can derive the optimal adjustment on the reused case is a potential solution for improving the prediction accuracy.

GA is an adaptive search technique which mimics the processes of evolution to solve optimization problems (Holland, 1975). It is an adaptive search approach for solving problems that is very useful in optimization problems requiring a complex search (Gong, Zulkernine, and Abolmaesumi, 2005;Huang and Chiu, 2006). Applying suitable adjustments is crucial to analogical reasoning for CBR predictors. Thus, the present paper aims to investigate the effect on the prediction accuracy when GA is adopted to determine a suitable adjustment on the reused case based on the similarity distances between pairs of cases in CBR predictors. There is no compelling reason in preferring one distance metric for CBR to another. Thus, the three different distance metrics are adopted for constructing their respective CBR predictor model in order to compare their effect on the accuracy improvement of broiler price predictions.

3. Methodology

3.1. Un-adapted CBR

A simple framework of un-adapted CBR predictors (traditional CBR approach) is shown in Fig. 1. The un-adapted CBR approach is to adopt the reused price without adjustment of reuse case. Among all of the historical cases, the closest analogue to a case being predicted is the case whose historical price is reused for deriving the price prediction. Three CBR predictors adopting widely used distance metrics of Euclidean distance (CBR–ED),

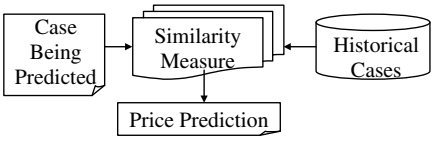

Fig. 1. A simple framework of un-Adapted CBR predictors.

Manhattan distance (CBR–MH), and Minkowski distance (CBR–MK) are introduced in Eqs. (1)–(3), respectively.

Among these equations, m is the number of price drivers; C_x is a case being predicted and C_y is one of the cases in the historical case base; C_{x_i} is the value of price driver i in case x and C_{y_i} is the value of price driver i in case y. In Eq. (2), it becomes the same as the CBR–ED approach when e is equal to 2. The reused price, R_x, is derived from the retrieved case with the closest analogue to the case x. Hence, the closest analogue to the case being predicted is the case with the minimum distance from among all the historical cases for CBR–ED, CBR–MH and CBR–MK approaches.

$$\mathrm{D}(C_x, C_y) = \sqrt{\sum_{i=1}^{m} \left(C_{x_i} - C_{y_i}\right)^2} \tag{1}$$

$$\mathrm{D}(C_x, C_y) = \sqrt[e]{\sum_{i=1}^{m} |C_{x_i} - C_{y_i}|^e} \tag{2}$$

$$\mathrm{D}(C_x, C_y) = \sum_{i=1}^{m} |C_{x_i} - C_{y_i}| \tag{3}$$

3.2. The Adapted CBR

The adapted CBR approach is to adjust the reused price based on the distances of two cases. GA approach is adopted to explore a suitably linear equation in order to adjust the reused price from CBR predictors. The adapted CBR predictor framework includes a feedback from the evaluation of price predictions to the linear adjustment mechanism throughout the development process. Three different adapted CBR predictor models based on the CBR–ED, CBR–MH and CBR–MK are proposed namely, the adapted CBR predictor using Euclidean distance (ACBR–ED), the adapted CBR predictor using Manhattan distance (ACBR–MH), and the adapted CBR predictor using Minkowski distance (ACBR–MK), respectively. Each adapted CBR adopts a linear equation to predict an adjustment value for the reused price. The construction of linear models ACBR–MH and ACBR–MK are the same as the processes in ACBR–ED model.

The framework of model construction for the adapted CBR predictor model is shown in Fig. 2. An adapted CBR predictor model applies a linear equation on the un-adapted CBR predictor model for adjusting the reused price. In constructing a linear equation, GA is adopted to search for a suitable linear equation in order to adjust the reused price derived from un-adapted CBR predictor model. Un-adapted CBR predictor model is the basic model for retrieving the most similar case for adapted CBR predictor model. GA explores a linear adjustment equation from the n cases being predicted via CBR–ED model. Among the n cases, the coefficients, ß$_i$, of this linear equation are treated as variables to be examined, and the mean absolute error (MAE) is considered as the value to be minimized.

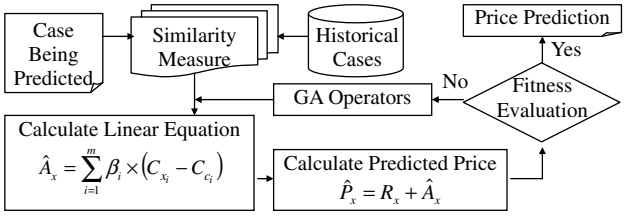

Fig. 2. The model construction for adapted CBR.

A linear equation determines the relationship between similarity distances and the predicted error. For a price driver i, C_{x_i} is the feature value for the case x being predicted and C_{c_i} is the feature value for the closest analogue c from among all of the historical cases. For n cases, $\widehat{A}_x$ is the adjustment value for the reused price in terms of the predicted error between the reused price and the actual price of the case being predicted, where R_x is the reused price of the closest analogue. The predicted price is therefore derived from two sources of the reused price and the adjustment value. The predicted price $\widehat{P}_x$ for the case being predicted is the sum of the reused price and the adjustment value.

GA operators include the processes of selection, crossover and mutation. It can be thought of as an evolutionary process wherein the fitness of each chromosome is calculated in each generation. The number of price drivers for a case is the number of coefficients to be examined for GA. The coefficients of this linear equation are encoded in a chromosome and groups of chromosomes make up a population. Chromosomes are selected for reproduction on the basis of their fitness. The reproduced chromosomes undergo a recombination that consists of crossovers and mutations. These iterative processes optimize the coefficients of this linear equation, and the combination of the best linear equation with the CBR–ED, CBR–MH and CBR–MK models are the price prediction model of ACBR–ED, ACBR–MH and ACBR–MK, respectively.

4. The experiments and results

4.1. The dataset description

From the viewpoint of industrial economics, the variables that affect the broiler price could be classified into industrial internal and external factors. In the industrial internal factors, those variables could include price of chick, price of feedstuff, quantity of broiler, and the related substitution products such as price of colorful broiler, price of pig, quantity and price of imported chicken. In the industrial external factors which are highly related with macroeconomic situations, includes variables such as business indicator, consumer price index, and wholesale price index.

The prices of chick and feedstuff are the major cost components of broiler production. If those prices change, the cost of broiler would also change and the broiler price

would be subsequently affected. When the supply quantity of broiler increases, the broiler price would drop, and the balance of demand and supply can be reached. The colorful broiler and pig are important substitution products for broiler. Therefore if their prices increase, consumers might alternatively buy chicken and broiler price would induce to increase. The quantity of imported chicken may affect the total chicken supply quantity in domestic market. For example, if the quantity of imported chicken increases, this would impose the pressure to decrease the broiler price. The price level of imported chicken could influence the price competitiveness of domestic broiler. If the price of imported chicken decreases, then the broiler price would be forced to reduce.

Regarding macroeconomic situations, business condition could reflect the social consumption ability, which would affect the broiler price. Suppose the business condition is fine, the social consumption would be encouraged. Therefore, the demand quantity of chicken might increase. Then the broiler price would have a potential to increase. The consumer price index is related to the purchase ability of consumers. It implies that consumer purchase power would decrease when the consumer price index increases. In this situation, consumers might reduce the demand quantity of chicken, which might have an indirect impact on the broiler price. The wholesale price index would be, in some sense, expressing the changes of farm prices. That might connect with broiler price more or less.

What are the practical influences of all mentioned variables on the broiler price? It still needs to be investigated. In other words, these factors might have positive or negative influences in theoretical inference. But their relationship might be or not significant in practice.

This study has collected 72 records from January 2000 to December 2005 in Taiwan. As shown in Table 1, 10 independent variables are adopted for model construction in this research. The dependent variable is the broiler price. The mean value of the dependent variable is 21.4 NT dollars with a minimum price of 18. 2 and maximum price of 26.4. All of these variables are continuous type. The whole data set is divided into pair of training and test sets. 48 records from January 2000 to December 2003 are for training. The other 24 records from January 2004 to November 2005 are for test.

Table 1
Variables of the dataset

Description	Maximum	Minimum	Mean	Stand deviation
Quantity of broiler	20358947.0	11823066.0	15763360.7	2194230.0
Price of colorful broiler	49.4	16.8	25.5	7.2
Price of chick	17.9	6.0	10.9	2.7
Price of feedstuff	12.0	10.3	11.1	0.5
Quantity of imported chicken	10618000.0	0.0	2680847.2	2504793. 2
Price of imported chicken	39.3	NA	23.7	5.9
Price of pig	66.9	37.0	49.1	7.6
Business indicator	115.1	96.1	107.3	4.8
Consumer price index	105.5	98.5	100.7	1.6
Wholesale price index	113.7	97.3	104.0	4.7
Broiler price (dependent variable)	26.4	18.2	21.4	1.6

4.2. The Experiments

Three CBR predictors using Euclidean distance (CBR–ED), Manhattan distance (CBR-MH), and Minkowski distance (CBR–MK) were adopted in order to construct their respective prediction models. The adapted CBR predictors using Euclidean distance (ACBR–ED), Manhattan distance (ACBR–MH), and Minkowski distance (ACBR–MK) were applied to the above-mentioned data sets and compared with the CBR predictors without any adjustment. The other widely used approaches of the classification and regression trees (CART) (Jeffery, Ruhe, and Wieczorek, 2001), the artificial neural networks (ANN) (Heiat, 2002) and the linear regression (LR) (Jeffery, Ruhe, and Wieczorek, 2000) were also used for constructing their respective prediction models in order to compare their prediction abilities.

For the GA parameters of ACBR–ED, ACBR–MH and ACBR–MK, they are decided based on the previous work (Huang and Chiu, 2006) and a certain amount of our experimentations. A roulette wheel selection method is adopted in conjunction with the uniform crossover approach. Mutation and crossover rates are 0.06 and 0.6, respectively. $10*N$ organisms are in the population, where N is the number of variables to be explored for GA. The entire learning process is stopped after $1000*N$ trials or when the best result did not change in the last $100*N$ trials.

Table 2 shows the prediction results of the different price prediction models by considering three CBR predictors without adjustment (CBR–ED, CBR–MH and CBR–MK), three CBR predictors with adjustments (ACBR–ED, ACBR–MH and ACBR–MK), and three widely used prediction approaches (ANN, CART and LR) using the evaluation criteria MAE between the training and the test

Table 2
The prediction results of different approaches

Approach		Training	Test
CBR without adjustment	CBR–ED	0.868	2.589
	CBR–MH	0.898	2.785
	CBR–MK	0.923	2.675
CBR with adjustment	ACBR–ED	0.728	0.801
	ACBR–MH	0.767	0.773
	ACBR–MK	0.793	0.872
	ANN	0.585	1.670
	CART	0.503	1.280
	LR	0.589	1.176

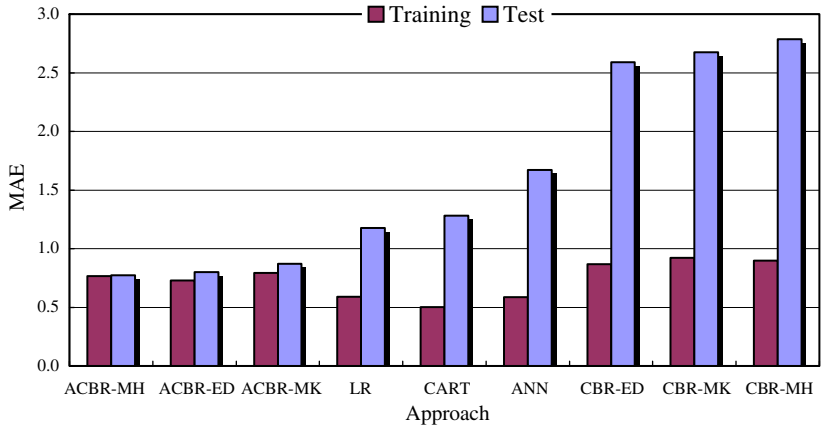

Fig. 3. MAE results of different approaches.

stages. By comparing these different price prediction models, a prediction model with a lower value of MAE presents an accurate price prediction for the broiler.

Fig. 3 shows the prediction results of all considered models at the training and test stages for the MAE evaluation criterion. Applying GA to explore the suitable adjustment of ACBR–ED, ACBR–MH and ACBR–MK models produces better predictions than the CBR predictors without adjustment of CBR–ED, CBR–MH and CBR–MK. For these three adapted CBR predictors, the CBR predictors using the different distance metrics of Euclidean, Manhattan and Minkowski present insignificant differences at both the training and test phases. Although the CART, ANN and LR approaches outperformed the models of the adapted CBR at the training stage, the test results of adapted CBR are superior to these models. In general, applying GA to adjust the CBR predictors enhanced the prediction performance of CBR predictors based on the MAE evaluation criterion. It is also comparable to the other price prediction models of ANN, CART and LR.

For the proposed adapted CBR predictors, Table 3 shows the improvement of MAE based on the CBR predictors without adjustment between the training and the test stages. For the CBR models with different distances metrics, the highest improvement is 16.1% at the training stage from the CBR–ED to ACBR–ED model, and 72.2% at the test stage for the CBR–MH to ACBR–MH model. The mean improvement for these models is 14.9% and 69.6% for training and test stages, respectively. Generally, the adapted CBR predictors of ACBR–MH, ACBR–MK and ACBR–ED improved the prediction abilities with the MAE performance indicators as compared to the CBR predictors without adjustment of CBR–ED, CBR–MH and CBR–MK.

Table 3
Improvements of different CBR approaches

Approach		Training	Test
Euclidean distance	CBR–ED	0.868	2.589
	ACBR–ED	0.728	0.801
	Improvement (%)	16.1	69.1
Manhattan distance	CBR–MH	0.898	2.785
	ACBR–MH	0.767	0.773
	Improvement (%)	14.6	72.2
Minikowski distance	CBR–MK	0.923	2.675
	ACBR–MK	0.793	0.872
	Improvement (%)	14.1	67.4
	Mean improvement (%)	14.9	69.6

5. Discussion and conclusion

The research results indicate that the proposed adapted CBR price prediction approach can support the producers to more effectively foresee the broiler prices. The experiment results show that our proposed adapted CBR is significantly superior to un-adapted CBR approach. Further, the adapted CBR approach presents better prediction performance than CART, ANN and LR with at least 50% less of MAE. Three different distance metrics of Euclidean, Manhattan and Minkowski seems making not much performance differences at both the training and test phases. Predicting the broiler price from the adjustment of reused broiler case based on the similarity distance is less explored in the existing broiler price prediction literature. Moreover, adjusting the price of the most similar case by considering the similarity distance to the case being predicted is a key to improve the prediction accuracy of the case-based broiler price estimation model.

The proposed approach is also very similar to the way of broiler price adjustment in practice. In the past the broiler price was estimated by producers based their accumulated experience and subjective judgment. However the human estimated price was always biased from the market price and thereby having great impacts on the production revenues. It's believed that the use of the prediction model can be of help to the producers to obtain more objective price information and based on which their production plans and management can be more effective. That is, with this prediction model, over production or production shortage of broilers can be prevented in advance.

On the other hand, the agriculture authority can construct the price prediction system as a pre-warning mechanism to direct the nation-wide poultry resources allocation and provide decent production suggestion for the producers. For the chicken importers, the price prediction information may also provide valuable clues for their logistic planning.

The use of different configurations or other different design of adapted case-based predictions (e.g. nonlinear adjustment or Kalman filter) may bring in further enhancement of the proposed method in the future. Other, the appropriate features combination is also possible for improving the model prediction performance. Of course, the collection of more market and production data is helpful to obtain an effective prediction mode as well.

References

Abdulai, A. (2002). Using threshold cointegration to estimate asymmetric price transmission in the Swiss pork market. *Applied Economics, 34*, 679–687.

Beddoe, G. R., & Petrovic, S. (2006). Selecting and Weighting Features using a Genetic Algorithm in a Case-based Reasoning Approach to Personnel Rostering. *European Journal of Operational Research, 175*, 649–671.

Chiu, C., & Chiu, N. H. (2003). The prediction of airplane landing gravity using case-based reasoning. *International Journal of Industrial Engineering: Theory, Applications, and Practice, 10*, 82–89.

Gong, R. H., Zulkernine, M., & Abolmaesumi, P. (2005). A software implementation of a genetic algorithm based approach to network intrusion detection. *Proceedings of the Sixth International Conference on Software Engineering*, 246–253.

Goodwin, B. K., & Piggott, N. E. (2001). Spatial market integration in the presence of threshold effects. *American Journal of Agricultural Economics, 83*(2), 302–317.

Heiat, A. (2002). Comparison of artificial neural network and regression models for estimating software development effort. *Information and Software Technology, 44*, 911–922.

Holland, J. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, Mich., USA: University of Michigan Press.

Holt, M. T., & Aradhyula, S. V. (1998). Endogenous risk in rational-expectations commodity models: A multivariate generalized ARCH-M approach. *Journal of Empirical Finance, 5*, 99–129.

Huang, S. J., & Chiu, N. H. (2006). Optimization of analogy weights by genetic algorithm for software effort estimation. *Information and Software Technology, 48*, 1034–1045.

Impedovo, S., Lucchese, M. G., & Pirlo, G. (2006). Optimal zoning design by genetic algorithms. *IEEE Transactions on Systems, Man and Cybernetics, 36*, 833–846.

Jeffery, R., Ruhe, M., & Wieczorek, I. (2000). A comparative study of two software development cost modeling techniques using multi-organizational and company-specific data. *Information and Software Technology, 42*, 1009–1016.

Jeffery, R., Ruhe, M., & Wieczorek, I. (2001). Using public domain metrics to estimate software development effort. *Seventh IEEE International Software Metrics Symposium*, 16–27.

Jorgenson, M. (2004). A review of studies on expert estimation of software development effort. *The Journal of Systems and Software, 70*, 37–60.

Kapombe, C. M., & Colyer, D. (1999). A structural time series analysis of us broiler exports. *American Journal Agricultural Economics, 21*, 295–307.

Kim, K. (2003). Financial time series forecasting using support vector machines. *Neurocomputing, 55*(1-2), 307–319.

Mair, C., Kadoda, G., Lefley, M., Phalp, K., Schofield, C., Shepperd, M., et al. (2000). An Investigation of Machine Learning Based Prediction Systems. *The Journal of Systems and Software, 53*, 23–29.

Sephton, P. S. (2003). Spatial market arbitrage and threshold cointegration. *American Journal of Agricultural Economics, 85*(4), 1041–1046.

Shepperd, M., & Schofield, C. (1997). Estimating software project effort using analogies. *IEEE Transactions on Software Engineering, 23*(12), 736–743.