Advanced Science and Technology Letters
Vol.95 (CIA 2015), pp.73-77
http://dx.doi.org/10.14257/astl.2015.95.14

Weighted sampling method for improving performance prediction: The case of pig management system

Ikhoon Jang,[1] Seo Youn Lee,[2] Young Chan Choe[3]

[1] First author, Seoul National University, Agricultural Economics and Rural Development, 200-8201, Gwanak-ro 1, Gwanak-gu, Seoul, 151-921, iks0404@snu.ac.kr
[2] Seoul National University, Agricultural Economics and Rural Development, 200-8201, Gwanak-ro 1, Gwanak-gu, Seoul, 151-921, nshaoba@hotmail.com
[3] Corresponding author, Seoul National University, Agricultural Economics and Rural Development, 200-8216, Gwanak-ro 1, Gwanak-gu, Seoul, 151-921, aggi@snu.ac.kr

Abstract. The main purpose of this paper is to introduce a weighted sampling method that can reduce bias in a swine production prediction model and to propose improved performance effect. The results of this study indicate that applying a weighted sampling method can increase prediction accuracy when predicting production by using swine management system users' data that have different distributions from a population. This can solve problems with decreasing prediction accuracy due to difficulty in the model's input data representing a population.

Keywords: Weighted sampling method, Population estimation, Swine production prediction, Pig management system, Prediction performance

1 Introduction

Prediction is studied in many different areas. Industry spends a large amount of money to predict the future as prediction can create profits and avoid risks. In agriculture, it is not easy to increase prediction accuracy because of various external factors such as climate, harmful insects, and price fluctuations. For example, Monsanto bought Climate Corporation for $1.1 billion, which indicates how important it is to predict the future in agriculture.

Many studies related to performance prediction seek to increase model accuracy by using the newest algorithm. This approach can be successfully implemented when input data represent a population. If sampled input data are biased, the prediction results can be incorrect. Thus, unbiasedness of input data is as important as accuracy in a prediction method. The main purpose of this study is to introduce a weighted sampling method, which can reduce bias in input data of a swine production prediction model and propose improved performance effect. First, this paper introduces previous studies regarding a sampling method and output prediction. Then, the paper explains swine production prediction and methods for improving performance when using data obtained from a swine information system. The conclusions propose implications based on the study results.

ISSN: 2287-1233 ASTL

2 Previous Research

2.1 Prediction for livestock production

The prediction of production has been actively studied in the area of stock price predictions. In the livestock sector, many studies have predicted livestock price [1-4]. In recent years, the increase performance of production prediction for pork, beef, and chicken is an important issue. In the past, [5] nonlinear optimization based on a heuristic search procedure was used to predict production. However, this particular procedure lacked accuracy. Some studies applied time series econometrics such as the random walk model or autoregressive model [6] as prediction techniques. However, current studies have focused on algorithms based on data-mining methods such as a neural network, which can better predict in comparison to econometric methods ([7],[8]). To improve prediction model performance based on a neural network, [9] used the Back Propagation algorithm with the Levenberg-Marquardt training method. Furthermore, the Support Vector Machine (SVM) algorithm is recognized for better prediction than a neural network, and [10] used SVM verified predictive performance of a livestock model.

2.2 Sampling method

To increase model performance, an accurately estimated population is important. However, measuring a population is almost impossible, and thus sampling without bias is also important. Weighted sampling is a classical sampling method [11-13]. [14] used the Bayesian approach to increase sampling accuracy. [15] proposed the weighted average importance sampling method and [16] presented the weighted random sampling method. In recent years, the respondent-driven method [17], capture-recapture method [17-19], random walk method [20-22], and biased sampling [23] were used to sample internet data, which do not have a normal distribution such as an online social network service. Considering various methods is important because population characteristics are an important factor to consider in determining the sampling method.

3 Prediction for production based on a pig management system

This study used the number of weaned piglets to predict production by using Pig management system data. Fig. 1 is a pig production life cycle. Farms that use this particular pig management system have the most abundant data as they input weaned piglet numbers. They can predict production after 23 weeks (about 160 days).

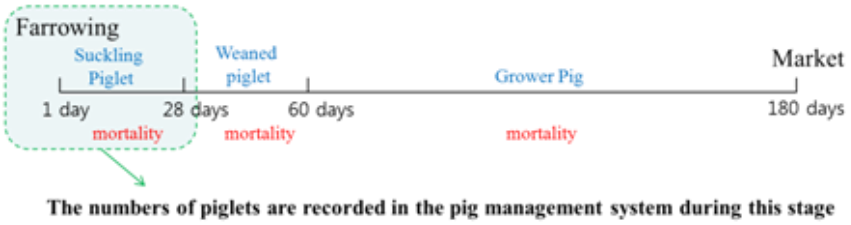

Fig 1. Life cycle of pig production

The number of weaned piglet data matches 100% with the shipments after 23 weeks, unless there is no mortality from illness or variation in shipment timing. However, these variables have an effect such that it is required to design a prediction model. To verify the effect of population estimation based on weighted samples, this study assumed that 100% of the weaned piglets are marketed.

4 Weighted sampling and test for improvement of prediction

If the weaned piglets recorded in the pig management system are 100% marketed, the number of pigs takes up 25% of total number of pigs slaughtered in Korea. The hog farms that use a management system are one-quarter of the entire swine farms. However, these farms have not been chosen from random sampling, and thus can have distinct characteristics. Table 1 presents the production ratio of different scales of farms that are selected from random sampling from a population and farms that use a management system. Table 1 illustrates that the number of pigs per farm that use a management system are larger than that of farms randomly selected from a population. A difference in productivity exists in the swine industry because large-size pig farms usually have higher productivity. Thus, the sampling method after applying a weighted value by farm size can revise the difference in distribution and productivity caused by different farm sizes.

Table 1. Ratio of normal hog farms and system users according to farm size in 2013

(unit: percent)	Under 1,000	1,000-2,000	2,000-3,000	3,000-5,000	Over 5,000
Sample from population	23	37.1	19.7	11.6	8.6
System users	4.7	26.4	20.4	22.1	26.4
Sample / system users	4.89	1.41	0.97	0.52	0.33

Table 2 shows the application of the machine learning technique, which manipulated and put farms over 10,000 pigs to 1. Weighted values for every case from 0.5 to 1.5 in the unit of 0.1 are multiplied by the number of weaned piglets in the other four size farms. We calculated the mean square error (MSE) and sum of the data added in the weekly basis and the actual number of slaughtered pigs after 23 weeks in the wholesale market. When we compared the default weighted value is 1, we found that

the minimum error case MSE value is lower. Farms that are under 1,000 pigs have a lower component ratio and number of shipments. Thus, these small-size farms have less influence on shipments, which results in less effect on the MSE value in every range. In the case 1,000 to 3,000 pig farms, the weighted value that is larger than 1 is found because the ratio of system users was lower than the normal size farms. In the case of 3,000-10,000 pig farms, the ratio of system users is larger than the normal size farms. Thus, the weighted value less than 1 is found. These results coincide with the results in Table 1.

5 Conclusions

This study proves that applying the weighted sampling method can increase prediction accuracy when predicting productivity by using data of a swine management system that have different distributions from a population. This result can solve the problem of decreasing prediction accuracy since prediction model input data cannot represent the entire population. However, the results have a limitation in that they can only be applied to the swine industry. Thus, future research should apply production prediction accuracy to other types of livestock.

Acknowledgements. This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the C-ITRC (Convergence Information Technology Research Center) (IITP-2015-H8601-15-1007) supervised by the IITP(Institute for Information &communications Technology Promotion).

References

1. Zhang, W. S., Chen, H. F., Wang, M. S. A forecast model of agricultural and livestock products price. Paper presented at the Applied Mechanics and Materials. (2010)
2. Ikerd, J. E.: Forecasting spreads between retail beef and live cattle prices: a model for routine use by market analysts. Paper presented at the NCR-134 Conference on Applied Commodity Price Analysis and Forecasting. Iowa State Univ. (1984)
3. Brandt, J. A., Bessler, D. A.: Forecasting with vector autoregressions versus a univariate ARIMA process: An empirical example with US hog prices. North Central Journal of Agricultural Economics, 29-36. (1984)
4. Sun, J.: Pork price forecast based on breeding sow stocks and hog-grain price ratio. Transactions of the Chinese Society of Agricultural Engineering, 29(13), 1-6. (2013)
5. Trapp, J. N.: Forecasting Short-Run Fed Beef Supplies with Estimated Data. American Journal of Agricultural Economics, 63(3), 457-465. (1981)
6. Lohano, H. D., Soomro, F. M.: Unit root test and forecast of milk production in Pakistan. International Journal of Dairy Science, 1(1). (2006)
7. Sanzogni, L., Kerr, D.: Milk production estimates using feed forward artificial neural networks. Computers and Electronics in Agriculture, 32(1), 21-30. (2001)

8. Salle, C., Guahyba, A., Wald, V., Silva, A., Salle, F., Nascimento, V.: Use of artificial neural networks to estimate production variables of broilers breeders in the production phase. British poultry science, 44(2), 211-217. (2003)
9. Wei, X., Qi, G., Shen, W., Jian, S.: Using LM BP Algorithm Forecase the 305 Days Production of First-Breed Dairy *Computer and Computing Technologies in Agriculture III* (pp. 359-363): Springer. (2010)
10. Tang, Y., Li, C.: The Study on Livestock Production Prediction in Heilongjiang Province Based on Support Vector Machine. Paper presented at the Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering. (2013)
11. Patil, G. P., Rao, C. R.: Weighted distributions and size-biased sampling with applications to wildlife populations and human families. Biometrics, 179-189. (1978)
12. Wong, C.-K., Easton, M. C.: An efficient method for weighted sampling without replacement. SIAM Journal on Computing, 9(1), 111-113. (1980)
13. Chen, X.-H., Dempster, A. P., Liu, J. S.: Weighted finite population sampling to maximize entropy. Biometrika, 81(3), 457-469. (1994)
14. Lo, A. Y.: A Bayesian method for weighted sampling. The Annals of Statistics, 2138-2148. (1993)
15. Hesterberg, T.: Weighted average importance sampling and defensive mixture distributions. Technometrics, 37(2), 185-194. (1995)
16. Efraimidis, P. S., Spirakis, P. G.: Weighted random sampling with a reservoir. Information Processing Letters, 97(5), 181-185. (2006)
17. Salganik, M. J., Heckathorn, D. D.: Sampling and estimation in hidden populations using respondent-driven sampling. Sociological methodology, 34(1), 193-240. (2004)
18. Chao, A., Lee, S., Jeng, S.: Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal. Biometrics, 201-216. (1992)
19. Lu, J., Li, D.: Estimating deep web data source size by capture–recapture method. Information retrieval, 13(1), 70-95. (2010)
20. Lu, J.: Ranking bias in deep web size estimation using capture recapture method. Data & Knowledge Engineering, 69(8), 866-879. (2010)
21. Lu, J., Li, D.: Sampling online social networks by random walk. Paper presented at the Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research. (2012)
22. Dasgupta, A., Jin, X., Jewell, B., Zhang, N., Das, G.. Unbiased estimation of size and other aggregates over hidden web databases. Paper presented at the Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. (2010)
23. Dasgupta, A., Das, G., Mannila, H.. A random walk approach to sampling hidden databases. Paper presented at the Proceedings of the 2007 ACM SIGMOD international conference on Management of data. (2007)

