

Data mining as a hatchery process evaluation tool

Daniela Regina Klein^{1*}, Marcos Martinez do Vale², Mariana Fernandes Ribas da Silva³, Micheli Faccin Kuhn⁴, Tatiane Branco⁴, Mauricio Portella dos Santos¹

¹Universidade Federal de Santa Maria – Depto. de Zootecnia, Av. Roraima, 1000 – 97105-000 – Santa Maria, RS – Brasil.

²Universidade Federal do Paraná – Depto. de Zootecnia, R. dos Funcionários, 1540 – 80035-050 – Curitiba, PR – Brasil.

³Universidade Federal do Rio Grande do Sul – Depto. de Zootecnia, Av. Bento Gonçalves, 7712 – 91540-000 – Porto Alegre, RS – Brasil.

⁴Universidade Estadual de Campinas/FEAGRI – Depto. de Construções Rurais e Ambiente, Av. Cândido Rondon, 501 – 13083-875 – Campinas, SP – Brasil.

*Corresponding author <danniwk@yahoo.com.br>

Edited by: Thomas Kumke

Received March 14, 2018

Accepted February 11, 2019

ABSTRACT: The hatchery is one of the most important segments of the poultry chain, and generates an abundance of data, which, when analyzed, allow for identifying critical points of the process. The aim of this study was to evaluate the applicability of the data mining technique to databases of egg incubation of broiler breeders and laying hen breeders. The study uses a database recording egg incubation from broiler breeders housed in pens with shavings used for litters in natural mating, as well as laying hen breeders housed in cages using an artificial insemination mating system. The data mining technique (DM) was applied to analyses in a classification task, using the type of breeder and house system for delineating classes. The database was analyzed in three different ways: original database, attribute selection, and expert analysis. Models were selected on the basis of model precision and class accuracy. The data mining technique allowed for the classification of hatchery fertile eggs from different genetic groups, as well as hatching rates and the percentage of fertile eggs (the attributes with the greatest classification power). Broiler breeders showed higher fertility (> 95 %), but higher embryonic mortality between the third and seventh day post-hatching (> 0.5 %) when compared to laying hen breeders' eggs. In conclusion, applying data mining to the hatchery process, selection of attributes and strategies based on the experience of experts can improve model performance.

Keywords: attribute selection, classification tree, data management, data mining and information technology

Introduction

The expansion of poultry farming is founded on artificial incubation, one of the most important segments of the poultry chain, which ensures the performance of breeders and the quality of day-old chicks. With the evolution of available technology, the hatchery process now generates more data with potential for information extraction that can help increase process performance.

Classic data processing methods are at times not efficient in elaborating biological responses and factors (Mehri, 2013) such as genetic selection for meat or egg production, which influence the development and metabolism of the embryo during the hatching process (Buzala et al., 2015; Burggren et al., 2015).

A number of data and knowledge strategies to develop precision livestock farming are currently emerging such as the expert systems involving data mining (Vale et al., 2008), fuzzy logic (Sousa et al., 2016; Dominiak and Kristensen, 2017) and artificial neural networking (ANN; Kumar and Hancke, 2015). The performance of ANN when applied to egg hatchability prediction is satisfactory (Bolzan et al., 2008), but it does not explain how the features are used to solve the problem. On the other hand, data mining (DM) explains the acquisition of knowledge in terms of logical rules, such as the construction of classification trees. DM has been successfully applied in animal science to animal environment (Vale et al., 2010), laying breeders' production (Ferreira et al., 2013), breeding certification (Vieira et al., 2015), animal welfare (Moi et al., 2014), and egg quality prediction (Orhan et al., 2016).

Applying DM to organizing and analyzing a sizeable pool of data, the solutions are obtained faster than traditional methods with optimal predictive performance, which are needed in industrial applications (Koksal et al., 2011). Another advantage of the application of DM is that a mixture of numeric, categorical, and date data, can tolerate missing and noisy data, thereby increasing efficiency and detect potential quality problems in processes (Lei-da Chen and Frolick, 2000), including the poultry chain.

The aim of this study was to evaluate the applicability of the DM technique to building decision trees, data from hatchery results from broiler breeders and laying hen breeders housed under different systems.

Materials and Methods

The study used a database of hatcheries of brown egg laying hen breeders and broiler breeders from Apr to Sept 2011. Broiler breeders were housed in pens with shavings used for litters and reproductive management with natural mating. Laying hen breeders were housed in cages using artificial insemination. It was expected that the different systems would generate differences that the DM technique would probably explain.

The analysis used the DM technique for a classification task, having as classifier two genetic classes: broiler breeders (BB) and laying hen breeders (LH). The database had 49 egg hatching batches, 33 from LH and 16 from BB, and the attributes and distributions are in Table 1.

Table 1 – Database attribute distribution for laying hen breeders (LH) and broiler breeders (BB) hatchery process.

| Attributes | Mean | Minimum | Maximum | Std. deviation |
|-----------------------|-------|---------|---------|----------------|
| Dbtemp incubator (°C) | 36.24 | 35.07 | 37.09 | 0.58 |
| Wbtemp incubator (°C) | 31.62 | 30.44 | 32.43 | 0.75 |
| Dbtemp hatcher (°C) | 36.58 | 35.17 | 37.33 | 0.53 |
| Wbtemp hatcher (°C) | 31.14 | 29.17 | 34.25 | 1.20 |
| Hatching eggs | 1056 | 140 | 5678 | 923.57 |
| Hatch rate (%) | 75.40 | 44.47 | 92.61 | 10.89 |
| Hatchability (%) | 82.73 | 51.53 | 95.07 | 10.56 |
| Fertile eggs (%) | 91.21 | 74.19 | 99.21 | 6.62 |
| Infertile eggs (%) | 8.79 | 0.79 | 25.81 | 6.62 |
| M1 (%) | 2.34 | 0 | 5.78 | 1.29 |
| M2 (%) | 1.81 | 0 | 13.20 | 2.87 |
| M3 (%) | 0.92 | 0 | 10.84 | 1.68 |
| M4 (%) | 2.50 | 0 | 10.69 | 2.31 |
| Pecked eggs (%) | 2.32 | 0 | 13.31 | 3.00 |
| Contaminated egg (%) | 0.34 | 0 | 4.51 | 0.81 |

Wbtemp = Wet bulb temperature; Dbtemp = Dry bulb temperature; Embryonic mortality determined with embryo diagnosis (M1: in the first 48 h, M2: 3rd to 7th day, M3: 8th to 14th day, M4: 15th to 20th day).

The database was analyzed using the Weka® software tool version 3.7.8, and the classification algorithm J48 for classification tasks. A classification tree, a graphical view of inverted tree form, was generated in which the root node is the first variable with the highest classification power presenting below the branches, formed by the other attributes that allow for classification.

The data mining technique was applied according to Vale et al. (2008) following the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology. Three approaches to database analysis were used as follows: 1. analysis of database without attribute selection; 2. using attribute selection algorithms available in Weka®; and 3. attribute selection based on the help of experts with a minimum of two years of experience in hatchery.

The selection of attributes by Weka® algorithms seeks to find in databases those attributes which are determinants in the process of knowledge extraction. The algorithms used for the attribute selection were: InfoGain, which evaluates the value of an attribute by measuring the gain of information with respect to the class; CFS (Correlation-based Feature Selection), which evaluates the value of a subset of attributes by considering the individual predictive capacity of each resource along with an inherent degree of redundancy; and Gain-Ratio, which evaluates the value of an attribute by measuring the rate of gain with respect to the class (Witten et al., 2016).

The selection of the best model was based on the general precision of the models and the accuracy of the classes according to Vale et al. (2008) and Vale et al. (2010), using a contingency matrix and interpretation of the classification rules by experts.

The database was submitted to an analysis of variance using the R® software package, version 2016.

Results

DM resulted in three classification trees of great relevance. The first classification modeled from the database without attribute selection (Figure 1, model precision of 76 %), classified the performance of hatching eggs from LH with better performance than those of BB, and class accuracy of 0.82 and 0.63 respectively.

The second classification approach used only relative values as directed by an expert. The percentage of fertile eggs was the main attribute able to classify the strains (Figure 2, model precision of 90 %, class accuracy of 0.94 and 0.82 respectively for LH and BB). The model rules determined that the percentage of fertile eggs less or equal or are less than 95 % was the variable with the highest classification power to define LH. This result confirms the results of Figure 1.

The third tree using the Weka® attribute selection approach is shown in Figure 3 (94 % model precision, and class accuracy 1.00 and 0.84 respectively for LH and BB). The percentage of fertile eggs was the only variable capable of classifying LH and BB egg performance.

The attributes of root nodes from the three decision trees were submitted to analysis of variance, and the means between the classes were different ($p < 0.001$).

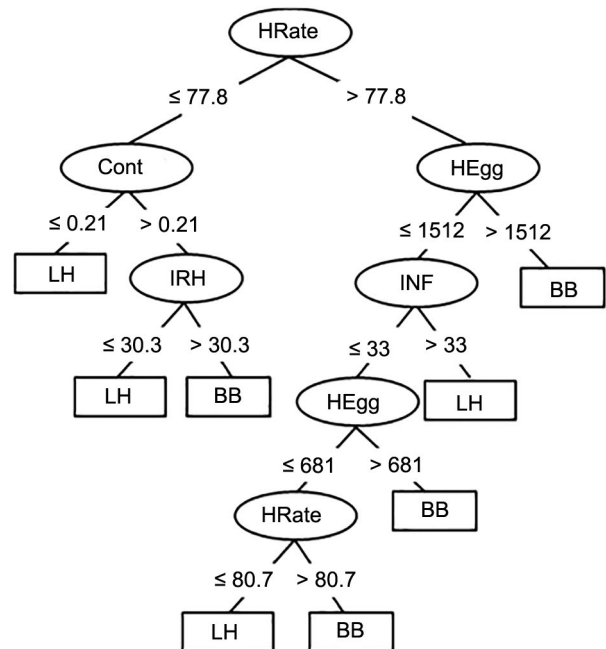


Figure 1 – Hatchery process classification from laying hen breeders (LH) and broiler breeders (BB) eggs. HRate = Hatching rate; HEgg = Hatched eggs; IRH = Incubator wet bulb temperature; Cont = Percentage of contaminated eggs; INF = Infertile eggs; FER = Fertile eggs.

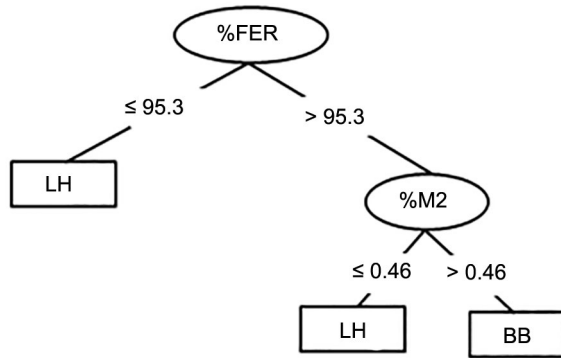


Figure 2 – Classification tree for hatchery process from laying hen breeders (LH) and broiler breeders (BB) eggs, evaluated by experts. %FER = Percentage of fertile eggs; %M2 = Percentage of embryo mortality in M2 period.

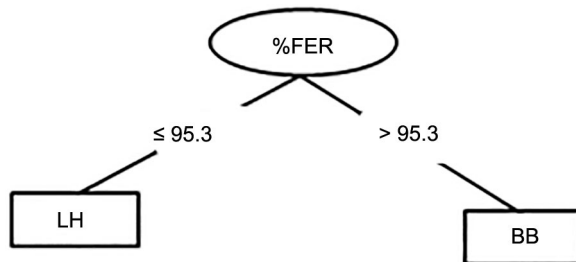


Figure 3 – Classification tree of laying hen breeders (LH) and broiler breeders (BB) hatchery process, using attribute selection. %FER = Percentage of hatched fertile eggs.

Discussion

The LH housed in cages results in a lower percentage of contaminated eggs because the eggs do not have contact with litter nor feces. The bacterial count in eggs of birds housed in litter systems is higher when compared to eggs from caged systems (Roberts and Chousalkar, 2014; Parisi et al., 2015).

A higher number of infertile eggs for LH is associated with artificial insemination, using diluted semen once a week (Getachew, 2016; Hughes, 1978). The ability of DM to find variations in hatching rate and infertility can identify eggs from different farms with different sanitary status and male quality.

The higher relative air humidity in the incubator for BB is related to the larger egg size and greater eggshell pore number, losing more water during embryo development changing the incubation environment (Boleli et al., 2016; Araújo et al., 2017). The presence of this variable in the construction of rules is consistent since many incubation problems arise from changes in the variables of the physical incubation environment.

The model considered the absolute number of incubated eggs when the hatchery batch size is constant. In conditions where the number of eggs per batch is not

constant, absolute egg number information may not be a good indicator of the incubation process, requiring relative values that can be preprocessed.

In the second tree BB has higher egg fertility and higher embryo mortality in the M2 phase, between the third and day of incubation (Buzala et al., 2015; Van Emous et al., 2015). This initial higher embryonic mortality is due to eggshell quality, egg size and, higher egg contamination from litter and nests (Ishaq et al., 2014; Parisi et al., 2015). The result reinforces the efficiency of the technique in extracting patterns for the differentiation of groups with different patterns.

The third classification tree focuses on the minimum information necessary to describe knowledge (Buczak and Guven, 2016). The rates of fertile eggs had great classification power and premise to identify classification rules linked with the production process (Vale et al., 2008).

The three approaches used in this DM confirm the importance of preprocessing data. Feature selection by experts or using algorithms is necessary in order to improve model precision and accuracy, removing irrelevant and redundant data (Kashef and Nezamabadi-Pour, 2015; Ramya et al., 2017).

The differences between the means of root node attributes of LH and BB prove that there are differences between the classes analyzed thereby improving the applicability of the DM technique and use of decision trees for extracting useful patterns from hatchery databases. Thus, taking note of other applications in animal science, the decision trees can be applied across the industry to quickly and efficiently find the patterns and solutions for quality problems.

Conclusion

The data mining technique allows for the classification of hatchery fertile eggs from different genetic groups. Hatching rates and the percentage of fertile eggs are the attributes with the highest classification power. Feature selection and the implementation of strategies based on the experience of experts improve model performance.

Authors' Contributions

Conceptualization: Klein, D.R.; Vale, M.M.; Silva, M.F.R. Data acquisition: Klein, D.R.; Kuhn, M.F. Data analysis: Klein, D.R.; Silva, M.F.R.; Vale, M.M. Design of methodology: Klein, D.R.; Vale, M.M.; Silva, M.F.R. Writing and editing: Klein, D.R.; Vale, M.M.; Silva, M.F.R.; Branco, T.; Santos, M.P.

Acknowledgments

To the FAPERGS (Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul) for providing a Master's degree scholarship.

References

- Araújo, I.C.S.D.; Leandro, N.S.M.; Mesquita, M.A.; Café, M.B.; Mello, H.H.C.; Gonzales, E. 2017. Water vapor conductance: a technique using eggshell fragments and relations with other parameters of eggshell. *Revista Brasileira de Zootecnia* 46: 896-902.
- Boleli, I.C.; Morita, V.S.; Matos Jr, J.B.; Thimotheo, M.; Almeida, V.R. 2016. Poultry egg incubation: integrating and optimizing production efficiency. *Brazilian Journal of Poultry Science* 18: 1-16.
- Bolzan, A.C.; Machado, R.A.F.; Piaia, J.C.Z. 2008. Egg hatchability prediction by multiple linear regression and artificial neural networks. *Brazilian Journal of Poultry Science* 10: 97-102.
- Buczak, A.L.; Guven, E. 2016. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials* 18: 1153-1176.
- Burggren, W.W.; Mueller, C.A.; Tazawa, H. 2015. Hypercapnic thresholds for embryonic acid-base metabolic compensation and hematological regulation during CO₂ challenges in layer and broiler chicken strains. *Respiratory Physiology & Neurobiology* 215: 1-12.
- Buzala, M.; Janicki, B.; Czarnecki, R. 2015. Consequences of different growth rates in broiler breeder and layer hens on embryogenesis, metabolism and metabolic rate: a review. *Poultry Science* 94: 728-733.
- Dominiak, K.N.; Kristensen, A.R. 2017. Prioritizing alarms from sensor-based detection models in livestock production: a review on model performance and alarm reducing methods. *Computers and Electronics in Agriculture* 133: 46-67.
- Ferreira, P.B.; Vale, M.M.; Macedo, A.; Boemo, L.S.; Rorato, P.R.N.; Beck, T.B. 2013. Phenotypic production characteristics from different breeds of laying hens through data mining. *Ciência Rural* 43: 164-171 (in Portuguese, with abstract in English).
- Getachew, T. 2016. A review article of artificial insemination in poultry. *World's Veterinary Journal* 6: 26-35.
- Hughes, B.L. 1978. Efficiency of producing hatching eggs via artificial insemination and natural mating of broiler pullets. *Poultry Science* 57: 534-537.
- Ishaq, H.M.; Akram, M.; Baber, M.E.; Jatoi, A.S.; Sahota, A.W.; Javed, K.; Husnain, F. 2014. Embryonic mortality in Cobb broiler breeder strain with three egg weight and storage periods at four production phases. *Journal of Animal and Plant Sciences* 24: 1623-1628.
- Kashef, S.; Nezamabadi-Pour, H. 2015. An advanced ACO algorithm for feature subset selection. *Neurocomputing* 147: 271-279.
- Koksal, G.; Batmaz, İ.; Testik, M.C. 2011. A review of data mining applications for quality improvement in manufacturing industry. *Expert Systems with Applications* 38: 13448-13467.
- Kumar, A.; Hancke, G.P. 2015. A zigbee-based animal health monitoring system. *IEEE Sensors Journal* 15: 610-617.
- Lei-da Chen, T.S.; Frolick, M.N. 2000. Data mining methods, applications, and tools. *Information Systems Management* 17: 67-68.
- Mehri, M. 2013. A comparison of neural network models, fuzzy logic, and multiple linear regression for prediction of hatchability. *Poultry Science* 92: 1138-1142.
- Moi, M.; Nääs, I.A.; Caldara, F.R.; Paz, I.C.L.A.; Garcia, R.G.; Cordeiro, A.F.S. 2014. Vocalization data mining for estimating swine stress conditions. *Engenharia Agrícola* 34: 445-450.
- Orhan, H.; Eyduran, E.; Tatliyer, A.; Saygici, H. 2016. Prediction of egg weight from egg quality characteristics via ridge regression and regression tree methods. *Revista Brasileira de Zootecnia* 45: 380-385.
- Parisi, M.A.; Northcutt, J.K.; Smith, D.P.; Steinberg, E.L.; Dawson, P.L. 2015. Microbiological contamination of shell eggs produced in conventional and free-range housing systems. *Food Control* 47: 161-165.
- Ramya, R.S.; Venugopal, K.R.; Iyengar, S.S.; Patnaik, L.M. 2017. Feature extraction and duplicate detection for text mining: a survey. *Global Journal of Computer Science and Technology* 16: 1-21.
- Roberts, J.R.; Chousalkar, K.K. 2014. Effect of production system and flock age on egg quality and total bacterial load in commercial laying hens. *Journal of Applied Poultry Research* 23: 59-70.
- Sousa, R.V.; Canata, T.F.; Leme, P.R.; Martello, L.S. 2016. Development and evaluation of a fuzzy logic classifier for assessing beef cattle thermal stress using weather and physiological variables. *Computers and Electronics in Agriculture* 127: 176-183.
- Vale, M.M.; Moura, D.J.; Nääs, I.A.; Oliveira, S.R.M.; Rodrigues, L.H.A. 2008. Data mining to estimate broiler mortality when exposed to heat wave. *Scientia Agricola* 65: 223-229.
- Vale, M.M.; Moura, D.J.I.; Nääs, I.A.; Pereira, D.F. 2010. Characterization of heat waves affecting mortality rates of broilers between 29 days and market age. *Brazilian Journal of Poultry Science* 12: 279-285.
- Van Emous, R.A.; Kwakkel, R.P.; Van Krimpen, M.M.; Van den Brand, H.; Hendriks, W.H. 2015. Effects of growth patterns and dietary protein levels during rearing of broiler breeders on fertility, hatchability, embryonic mortality, and offspring performance. *Poultry Science* 94: 681-691.
- Vieira, F.D.; Oliveira, S.R.D.M.; Paiva, S.R. 2015. Data mining-based technique on sheep breed certification. *Engenharia Agrícola* 35: 1172-1186.
- Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. 2016. *Data Mining: Practical Machine Learning Tools and Techniques*. 4ed. Morgan Kaufmann, Cambridge, MA, USA.