Turkey Gait Analysis: Predicting Expert Score With Machine Learning Based on IMU Data.

J.E. Doornweerd[1], A. Savchuk[2,3], B. Visser[2], A.C. Bouwman[1]

1 Animal Breeding and Genomics, Wageningen University & Research, Wageningen, the Netherlands.
2 Hendrix Genetics, Boxmeer, the Netherlands.
3 Jheronimus Academy of Data Science (JADS), 's-Hertogenbosch, the Netherlands.
janerik.doornweerd@wur.nl

Abstract

In livestock production, locomotion is an important health & welfare trait. Breed4Food aims to improve locomotion through breeding and herd management using sensor technology for precision phenotyping. Usually, locomotion is scored as a one-off subjective snapshot by a human expert. Motion sensor application in poultry is largely underdeveloped. Therefore, this study aimed to predict expert locomotion scores based on inertial measurement unit (IMU) data in turkeys. In turkey breeding, breeding candidates are subject to selection on locomotion scores. Each bird is individually scored from poor (1) to good (6) by an expert whilst walking through a corridor. During this routine procedure, three IMUs were attached to each bird (N = 83) with Velcro straps, one on each leg and one on the neck. The IMU provides 3D accelerometer, gyroscope & magnetometer data. Gradient boosting was used for step recognition based on leg IMU data (F-score: 0.82, on an allowed distance of 0.2s). The steps (N=1736 after quality control) served as input for feature extraction, which were subsequently used for prediction of locomotion scores of individual steps with gradient boosting. The model had a mean per class error of 0.37 on the test set. The current approach shows promise in providing objective locomotion scoring, possibly leading to more frequent scoring or continuous scoring of locomotion. Knowledge gained could also enhance the application of motion sensor technology in other livestock species.

Introduction

In livestock production, locomotion is an important health & welfare trait. Impaired locomotion compromises welfare and production. Generally, locomotion is scored as a one-off subjective snapshot by a human expert. In turkey, locomotion scores are heritable [1], repeatable, and valuable for selection but the process to acquire them is laborious, invasive, and subjective. Technological methods (*e.g.* force platforms [2], cameras [3], accelerometers [4]) could provide effortless, non-invasive and objective measurements. Accelerometers have found the most widespread use in livestock production, especially in cows and pigs, for the detection of behavioural changes as indicators of estrus, health & welfare (*e.g.* [5] & [6]). However, motion sensor application in poultry is largely underdeveloped, though, with technological advancements sensors are becoming smaller, cheaper and more accurate making them more viable for application in the poultry sector.

Accelerometers, cameras, and force platforms have been used to assess differences between animals of different locomotion scores [2,7], but not for the direct scoring of locomotion. Therefore, this study aimed to predict expert locomotion scores based on inertial measurement unit (IMU) data in turkeys. Inertial measurement units (IMUs) are like accelerometers but more extensive, providing 3D accelerometer, gyroscope & magnetometer data. Where previous work (Bouwman et al., under review) was concerned with step segmentation, this study focusses on feature extraction from those segmented steps and prediction of locomotion scores.

Volume 2 of the Proceedings of the joint 12th International Conference on Methods and Techniques in Behavioral Research and 6th Seminar on Behavioral Methods held online, May 18-20, 2022. www.measuringbehavior.org. Editors: Andrew Spink, Jarosław Barski, Anne-Marie Brouwer, Gernot Riedel, & Annesha Sil (2022). DOI 10.6084/m9.figshare.20066849. ISBN: 978-90-74821-94-0

Materials & Methods

Data collection

Data were collected on 85 animals during a standard walkway test applied in the turkey breeding program of Hybrid Turkeys (Hendrix Genetics, Kitchener, Canada). Each bird was individually scored from behind on a scale from poor (1) to good (6) by a human expert whilst walking through a corridor within the barn. During the test, the animals were equipped with IMUs (MTw Awinda, XSens Technologies B.V., Enschede, the Netherlands) on each upper leg, and stimulated to walk in one direction for approximately 5 meters. The animals often needed stimulation to start or keep walking. Since the data was collected during a routine process there was little time for habituation to the sensor presence.

The IMUs (16 g, 47x30x13 mm) recorded at 100Hz and recording was manually turned on and off, averaging 20s of material per animal. IMU output consisted of calibrated time series data for 3D acceleration (m s^{-2}), 3D angular velocity (° s^{-1}), and 3D magnetic field (arbitrary unit A.U., normalized to 1 during factory calibration). Orientation data was provided in Euler representation (Pitch, Roll, Yaw) and unit quaternions (q = [W X Y Z]) [8].

Feature extraction and training

Previous work (Bouwman et al., under review) was concerned with automated step segmentation from the IMU profiles. Several methods (change point detection, local extrema approach and Gradient Boosting Machine) were applied, of which the gradient boosting machine (GBM) had the best performance (F-score: 0.82, on an allowed distance of 0.2s) [*F-score is the harmonic mean of precision and recall, in which 1 is perfect*]. Feature extraction was based on the step segmentation of this method, see Figure 1.

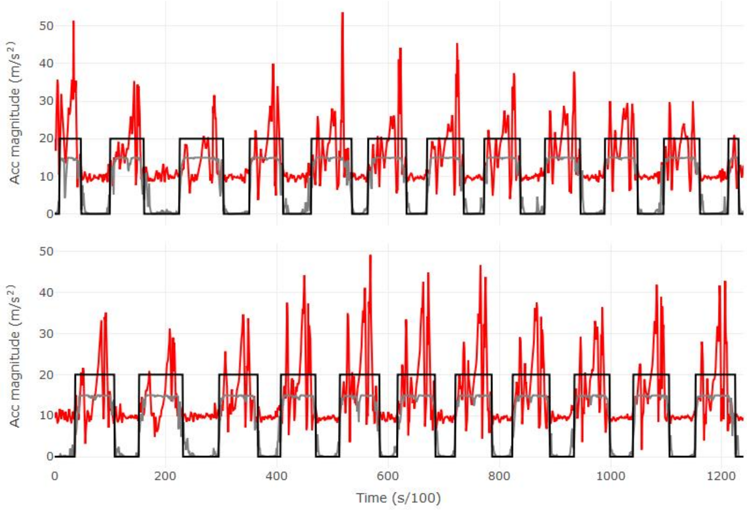

Figure 1. Results of GBM step segmentation for both legs of one turkey (085). Acceleration magnitude ($\sqrt{Acc_{x^2} + Acc_{y^2} + Acc_{z^2}}$) is plotted in red, chance of being a step according to GBM model is plotted in gray (x17 for visualization)

The steps (N=1736 after quality control) served as input for feature extraction. Per animal step contribution to the total number of steps differed (min: 8, mean: 20.92, max: 90). Quality control was based on the within leg median absolute deviation of step duration with a cut-off of 3.65. Extracted features included the minimum value, the 1st quartile, the median, the 3rth quartile, the maximum value, the geometric mean, skewness, the mean, kurtosis, the trimmed mean, the standard deviation/variance, the mode, the interquartile range, the coefficient of variation, the range, the median absolute deviation, the sum of each variable and the step duration.

Volume 2 of the Proceedings of the joint 12th International Conference on Methods and Techniques in Behavioral Research and 6th Seminar on Behavioral Methods held online, May 18-20, 2022. www.measuringbehavior.org. Editors: Andrew Spink, Jarosław Barski, Anne-Marie Brouwer, Gernot Riedel, & Annesha Sil (2022). DOI 10.6084/m9.figshare.20066849. ISBN: 978-90-74821-94-0

Feature selection was based upon absolute pair-wise correlations with a cut-off of 0.9, the feature with the highest mean absolute correlation within the pair was removed. Mean absolute correlations were re-evaluated after each removal, resulting in 201 remaining features.

Ninety-five percent of the total available data (N=1650) were used for model building to ensure class balance, the remaining 5 percent will be used to evaluate the finalised model. From the model-building data, 80% was used for training, 12.7% for validation, and 7.3% for testing. Current data only consists of animals from score 1 (N=30), score 2 (N=26), score 3 (N=23), and score 4 (N=4). Score 3 and score 4 animals were combined to form score 3 (N=27).

The R version of H2O (Open-source software from H2O.ai, version 3.28.0.1) was used to train the gradient boosting model to predict the scores of individual steps. Gradient boosting was chosen based on preliminary results with the AutoML function of H2O. Training was done with 5-fold cross-validation, a learn rate of 0.06, learn rate annealing of 0.995, a column sample rate of 0.1, a row sample rate of 0.9, a max depth of 11, a minimum number of rows at 5, a stopping tolerance of 0.0275 with 5 stopping rounds and a minimal split improvement of 0.0001 with the total number of trees of being 113.

Results

In Table 1, the performance of the model on the validation and test set is shown. The three most important variables were the average median absolute deviation of free acceleration on the Y-axis and the X-axis, and the minimum pitch value. In Table 2, the confusion matrix of the test set is shown. The overall mean per-class error is 0.37 with and associated logloss of 0.78.

Table 1. Performance results on validation and test set

1. Metric	Validation	Test
MSE	0.24	0.28
RMSE	0.49	0.53
Logloss	0.70	0.78
Mean per-class error	0.28	0.37
R^2	0.63	0.58

Table 2. Confusion matrix of the test set

	Predicted				
Actual	1	2	3	Error	Rate
1	26	8	6	0.35	14/40
2	5	25	10	0.375	15/40
3	9	6	25	0.375	15/40
Total	40	39	41	0.37	44/120

Discussion

The aim of this study is to predict expert locomotion scores based on inertial measurement unit (IMU) data in turkeys. To this end, a machine learning technique called gradient boosting was applied to IMU data of individual steps to predict the locomotion score based on features of each step.

The current model is trained on predicting the score of each individual step, whereas the expert scores the turkey on a series of steps. Predicting the score of each individual step instead of per animal was a necessity due to the limited number of animals. Despite the scoring of individual steps instead of animals, the results (mean per-class error of 0.37) indicate the possibility of using single steps to predict animal score. However, this is under the assumption that each step within an animal's step profile is unique yet indicative of the score of the animal. Additionally, given the split over steps, the possibility exists that the animal is detected instead of the locomotion score. It is unclear if this phenomenon occurs, and if so, to what effect this phenomenon affects the predictions. However, if this phenomenon occurs to its full extent, it would be expected that the mean per-class error would have been lower. Furthermore, the expert only considers 'good' steps, however, what constitutes a good step? Should step filtering within an animal's step profile consist of more than a conservative step duration filtering?

Initial results show a logloss of 0.78 on the test set, with a logloss of 1.10 associated with random guessing. Hence, the model shows that it has found links between the IMU features and the expert locomotion score. However, it should be noted that in the current model pitch, roll, and yaw were included as variables for feature extraction

Volume 2 of the Proceedings of the joint 12th International Conference on Methods and Techniques in Behavioral Research and 6th Seminar on Behavioral Methods held online, May 18-20, 2022. www.measuringbehavior.org. Editors: Andrew Spink, Jarosław Barski, Anne-Marie Brouwer, Gernot Riedel, & Annesha Sil (2022). DOI 10.6084/m9.figshare.20066849. ISBN: 978-90-74821-94-0

despite problems with gimbal lock. Gimbal lock is the occurrence of axis alignment which causes singularities. The occurrence of gimbal lock is dependent on the initial sensor placement and the movement of the animal. The algorithm could have picked up on the phenomenon and partially base the predictions on it.

The classifications shown in Table 2 show that steps of animals with an actual score of 1 could be predicted as a score 3 step and vice versa. A certain overlap between scores is expected, however, one would expect that the largest overlap for steps of score 3 animals would occur with steps of score 2 animals, not steps of score 1 animals. However, certain steps within an animal's step profile could be considered anomalies given the rest of the steps within that animal's step profile. Currently, the reason(s) as to why the misclassifications happen is being investigated.

Conclusion

Although the research is still in progress, the preliminary results show promise in predicting expert locomotion scores based on IMU data. The IMU data seems to contain the information which the expert considers in scoring the turkeys. Further refinement of the features and model hyperparameters could improve results.

Ethical statement

Ethical review and approval was not required for the animal study because The Animal Welfare Body (AWB) of Wageningen Research decided ethical review was not necessary because the applied units were low in weight (<1% of body weight), the units were attached for less than one hour, the animal is not isolated in the corridor and more or less familiar with the corridor.

References

1. Quinton, C. D., Wood, B. J., & Miller, S. P. (2011). Genetic analysis of survival and fitness in Turkeys with multiple-trait animal models. *Poultry Science* **90**, 2479–2486. https://doi.org/10.3382/ps.2011-01604

2. De Alencar Naas, I., De Lima Almeida Paz, I. C., Baracho, M. dos S., De Menezes, A. G., De Lima, K. A. O., De Freitas Bueno, L. G., … De Souza, A. L. (2010). Assessing locomotion deficiency in broiler chicken. *Scientia Agricol*, **67**: 129–135. Retrieved from www.scielo.org

3. Kashiha, M. A., Bahr, C., Ott, S., Moons, C. P. H., Niewold, T. A., Tuyttens, F., & Berckmans, D. (2014). Automatic monitoring of pig locomotion using image analysis. *Livestock Science* **159**: 141–148. https://doi.org/10.1016/j.livsci.2013.11.007

4. Stevenson, R., Dalton, H. A., & Erasmus, M. (2019). Validity of micro-data loggers to determine walking activity of Turkeys and effects on Turkey gait. *Frontiers in Veterinary Science* **5**: 1–12. https://doi.org/10.3389/fvets.2018.00319

5. Martiskainen, P., Järvinen, M., Skön, J. P., Tiirikainen, J., Kolehmainen, M., & Mononen, J. (2009). Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. *Applied Animal Behaviour Science* **119**: 32–38. https://doi.org/10.1016/j.applanim.2009.03.005

6. Thompson, R., Matheson, S. M., Plötz, T., Edwards, S. A., & Kyriazakis, I. (2016). Porcine lie detectors: Automatic quantification of posture state and transitions in sows using inertial sensors. *Computers and Electronics in Agriculture*, **127**: 521–530. https://doi.org/10.1016/j.compag.2016.07.017

7. Aydin, A., Cangar, O., Ozcan, S. E., Bahr, C., & Berckmans, D. (2010). Application of a fully automatic analysis tool to assess the activity of broiler chickens with different gait scores. *Computers and Electronics in Agriculture* **73**: 194–199. https://doi.org/10.1016/j.compag.2010.05.004

8. Paulich, M., Schepers, M., Rudigkeit, N., & Bellusci, G. (2013). Xsens MTw : Miniature Wireless Inertial Motion Tracker for Highly Accurate 3D Kinematic Applications. *Xsens Technologies* (April), 1–9.

Volume 2 of the Proceedings of the joint 12th International Conference on Methods and Techniques in Behavioral Research and 6th Seminar on Behavioral Methods held online, May 18-20, 2022. www.measuringbehavior.org. Editors: Andrew Spink, Jarosław Barski, Anne-Marie Brouwer, Gernot Riedel, & Annesha Sil (2022). DOI 10.6084/m9.figshare.20066849. ISBN: 978-90-74821-94-0