Automatic Pig’s Maternal Ability Evaluation System Based on Behavior Data of Sensor

Jie Liu, Chong Shi*, Guangmin Sun and Pan Ma
The Faculty of Information Technology, Beijing University of Technology, Beijing, China

*Corresponding author

Abstract. The high-risk behaviour of pigs from standing, sitting to lying and grovelling is the main reason that the piglets are always crushed to death. This paper conducts research based on the field of classification and recognition of pig behaviours by using triaxial acceleration sensor, which can evaluate the maternal ability of pigs and provide data basis for selecting high quality breeding pigs. Aiming at the problem of low data variability and small data range due to the small-scale activity of the pigs, which can cause poor classification accuracy. This paper first performs moving average filter processing on the x, y, and z-axis data collected by the triaxial acceleration sensor. After feature extraction and feature selection, an optimal feature subset is proposed. Experiments show that by adopting the optimal feature subset proposed, the random forest classifier adopted can classify and evaluate four basic behaviours of pigs in daily life better. The accuracy on the test set reaches 93.8%. Compared with decision tree and BP neural network, the AUC value of random forest reaches 0.957, which has obvious performance advantages. Finally, this paper also proposed an evaluation model of maternal ability for pigs and adopted maternal ability index to evaluate the maternal ability of the pigs according to the classification results.

1. Introduction

In 2000, relevant data from the US Department of Agriculture’s National Animal Health Monitoring System showed that the mortality rate of piglets before weaning was 11%, and the crush mortality rate accounted for 52.1% [1]; In 2013, the Danish Pig Research Center’s statistics from nine farms show that the average mortality rate of piglets before weaning was 13%, and the crush mortality rate accounted for 15% to 51% of them [2]. Studies have shown that the risk of crushed piglets always happens at the moment when certain physical behaviours occur of the pigs before delivery, which include some high-risk behaviours such as lying, grovelling, and rolling [3]. By Classifying and recognizing daily basic behaviours of pigs, and thus finding pigs with fewer occurrences of high-risk behaviours as breeding pigs, which can help improve the level of pigs’ maternal behaviour and reduce mortality rate of piglets before weaning.

The behaviour recognition technology based on the low-cost triaxial acceleration sensor can obtain collect of pig behaviours in real time. It is very simple to determine the behaviour of pigs quickly and correctly by combining with behaviour analysing methods. Thompson [4] and others performed feature extraction based on a 2s time sliding window on the data collected by the triaxial acceleration sensor in 2016. They examined the features like composite acceleration, maximum value, pitch angle, and roll angle. By using SVM classifier, they recognized the behaviours of pigs. However, this method has high complexity on calculation of the feature set, low implementation efficiency of the algorithm, and poor recognition effect on sitting behaviour.
This paper adopted low-cost triaxial acceleration sensor to collect information on the movement of pigs. After data analysing, data preprocessing, feature extraction, feature selection, an optimal feature subset that can represent various behavioural characteristics of pigs better was proposed. The training set trained the random forest classifier according to the selected features to obtain the reference model. The test set adopted the classifier trained for classification and evaluation based on the reference model. The research results can evaluate the maternal ability of pigs and provide data basis for selecting high quality breeding pigs. The structure of pig behaviour recognition is shown in Figure 1.

2. Materials and Methods

2.1. Data Collection
The experimental data comes from the measurement of a pig in a farm. The data was measured from August 17, 2019 to August 19, 2019. When the pig was moving, the sensor would occur triaxial acceleration: x-axis(horizontal axis), y-axis(vertical axis), and z-axis. In order to label the measured data, the real-time monitoring video was saved at the same time.

In order to prevent the interference data occurred by the stress response of the pig from affecting the experiment, this paper wore a triaxial acceleration sensor on the right ear of the pig.

The type of the sensor is MPU6050. The direction of the sensor is shown in Figure 2. The x-axis points to the right side of the pig's body, the y-axis points to the tail of the pig and the z-axis points directly above the pig's head. Meanwhile, the triaxial acceleration value can be converted into other information such as posture angle and composite acceleration by calculating.

Figure 2. Experimental pig and directions of triaxial acceleration sensor.

Figure 3. Triaxial acceleration data curve.
2.2. Data Analysis
The behaviour of pigs can be divided into four basic behaviours including standing, eating, lying and grovelling. Among them, eating is a dynamic behaviour while standing, lying and grovelling are relatively static behaviours. This paper only studies these four basic behaviours which doesn’t include the transition between behaviours. First, this paper drew a small amount of continuous samples of four behaviours for preliminary analysis. Figure 3 shows the data curves of the acceleration signal contains x-axis, y-axis, z-axis, and triaxial composite acceleration SMA which can reflect the overall situation of the pig’s behaviour.

\[
SMA = \left( a_x^2 + a_y^2 + a_z^2 \right)^{1/2}
\]  
(1)

It can be seen from Figure 3 that the dynamic behaviour such as eating behaviour fluctuates greatly while the static behaviour such as lying and grovelling behaviours fluctuates relatively smoothly. The mean and standard deviation can be used to distinguish these two curves, but a in-depth study about the behavioural distinction between dynamic or static behaviour needs to be done.

2.3. Data Preprocessing
The data collected contains noise due to the shaking of the pig's ears. In addition, the range and difference of the collected data is small due to the small-scale activity of the pig. It is impossible to use the raw data directly. Therefore, this paper adopted moving average filter to smooth [5] the raw data first. For each spatial coordinate axis \( i \), given raw signal samples \( a(t) \) in 1s, \( a(t)' \) can be obtained by moving average filter :

\[
a(t) = (a_x, a_y, a_z)^T
\]

\[
a_i(t)' = \beta \cdot a_i(t - 1) + (1 - \beta) a_i(t) \quad i \in \{x, y, z\}
\]

\( \beta \) is used as a weight for adjusting the acceleration change rate. The experimental sample contained 122177 sets of raw data. Because of the temporality of the experimental sample, this paper adopted 1s-time sliding window as a standard in order to combine several groups of data to obtain 19779 sets of behaviour data vectors, which can be divided into four classes: standing, eating, lying, grovelling. The amount of the raw data from each class is shown in the Sample Size in Table 1. In order to solve the problem of the imbalanced samples, the class with less samples were over-sampled by randomly copying small samples from itself for data expansion. Thus making the ratio between four classes samples reach 1: 1: 1: 1. Then, feature extraction was performed on each class of samples. The data set was divided into train set and test set according to the ratio of 7: 3, which is shown in the Train Size and Test Size in Table 1.

| Behavior | Sample Size | Oversampling | Train Size | Test Size |
|----------|-------------|--------------|------------|-----------|
| Stand    | 4621        | 7843         | 5490       | 2353      |
| Eat      | 3469        | 7843         | 5490       | 2353      |
| Grovel   | 7843        | 7843         | 5490       | 2353      |
| Lie      | 3846        | 7843         | 5490       | 2353      |

2.4. Feature Extraction
Feature parameters include the mean, standard deviation, upper quartile, lower quartile, range, maximum, minimum, zero-crossing, Euclidean distance similarity, mean values of pitch and roll of the x, y, and z-axis and the composite axis within 1 s, which constitutes 37-dimensional feature vectors, like\([M, S, U, D, J, B, X, Z, O, R, P]\). The Description of parameters is shown as follows.

\( a_i \) stands for the acceleration of each axis, \( i \in \{x, y, z\} \).

- \( M = [M_{ax}, M_{ay}, M_{az}, M_{a_{xyz}}] \) is the mean vector of x, y, z-axis and the composite axis in 1s, which can reflect different trends of the data and stand for feature 1-4.
\[ M_{ai} = \frac{1}{n} \sum_{k=0}^{n} a_i \]  

- \( S = [S_{ax}, S_{ay}, S_{az}, S_{a_{xyz}}] \) is the standard deviation vector of x, y, and z-axis in 1s, and the composite axis, which can reflect the dispersion degree of the data and stand for feature 5-8.

\[ S_{ai} = \left( \frac{1}{n-1} \sum_{k=0}^{n} (a_i - mean_i)^2 \right)^{1/2} \]  

- \( U = [U_{ax}, U_{ay}, U_{az}, U_{a_{xyz}}] \) \( D = [D_{ax}, D_{ay}, D_{az}, D_{a_{xyz}}] \) are the vectors of the upper and lower quartiles of x, y, and z-axis and the composite axis in 1s, which can stand for feature 9-16.

\[ U_{ai} = sort(a_i) \left[ \frac{n+1}{4} \right] \]  

\[ D_{ai} = sort(a_i) \left[ \frac{3(n+1)}{4} \right] \]  

- \( J = [J_{ax}, J_{ay}, J_{az}, J_{a_{xyz}}] \) is the range vector of x, y, and z-axis and the combined axis in 1s, which can reflect the fluctuation range of the data and stand for feature 17-20.

\[ J_{ai} = |\text{max}(a_i) - \text{min}(a_i)| \]  

- \( B = [B_{ax}, B_{ay}, B_{az}, B_{a_{xyz}}] \) \( X = [X_{ax}, X_{ay}, X_{az}, X_{a_{xyz}}] \) are the maximum and minimum vectors of x, y, and z-axis and the composite axis in 1s, which can reflect the instantaneous acceleration change of the pig's violent movement and stand for feature 21-28.

\[ B_{ai} = \text{max}(a_i) \]  

\[ X_{ai} = \text{min}(a_i) \]  

- \( Z = [Z_{ax}, Z_{ay}, Z_{az}, Z_{a_{xyz}}] \) is the vector of zero-crossing points of x, y, and z-axis and the composite axis in 1s, which can reflect the direction of the pig's movement and stand for feature 29-32.

\[ Z_{ai} = \text{amount}(a_i > 0) \]  

- \( O = [O_{ax}, O_{ay}, O_{az}, O_{a_{xyz}}] \) is the Euclidean distance similarity of the accelerometer in axes like x and y, y and z, x and z in 1s, which can reflect the degree of interaction between the two axis data and stand for feature 33-35.

\[ O_{aij} = (\sum_{k=0}^{n} (a_i - a_j)^2)^{1/2} \]  

- \( R = [r], P = [p] \) are the mean vector of the pitch and roll angles of x, y, and z-axis in 1s, which can reflect the horizontal and vertical movement of the pig and stand for feature 36 and feature 37.

\[ p = \frac{1}{n} \sum_{k=0}^{n} \arctan \left( \frac{a_y}{(a_x^2 + a_y^2)^{1/2}} \right) \quad (pe[-90°, 90°]) \]  

\[ r = \frac{1}{n} \sum_{k=0}^{n} \arctan \left( \frac{-a_z}{a_x} \right) \quad (re[-180°, 180°]) \]  

2.5. Feature Selection and Behaviour Classification

This paper adopted random forest classifier to classify and recognize the behaviour of the pig. The steps of random forest are shown in Figure 4. As for parameter selection, according to the results about the classifier’s performance of multiple experiments, the n-estimators (the number of decision subtrees) was set to 100, the min-samples-leaf (the minimum number of samples about leaf nodes) was set to 1, and the max-features (The maximum number of features used in a single decision tree) was set to the square root of the total number of features. Finally a better classification result can be achieved according to
the above parameter settings. In order to improve the accuracy of classification and recognition, it is also necessary to perform feature selection on the above-mentioned 37-dimensional feature vectors to set up the optimal feature vector subset. Feature selection is a kind of process selecting better features from the original feature subset according to the evaluation criteria clearly defined in advance to achieve the effect of eliminating irrelevant and redundant features. Compared with dimensionality reduction techniques such as linear discriminant analysis or principal component analysis (PCA), feature selection does not change the original representation of the features set [6]. In the process of constructing a random forest, each decision tree uses different data, which makes the feature importance of each tree behave differently. Therefore, it is necessary to regard the importance of each feature on each tree as a basis for feature selection.

This paper adopted the feature importances index for feature selection. Figure 5 shows the importance score of each feature under the random forest classifier. It can be seen from Figure 5 that the two kinds of angle features make a significant contribution to the classifier while the feature importances among features like Euclidean distance and range are almost 0, the feature importances of each axis standard deviation are also relatively low. Therefore, this paper deleted the above 12-dimensional features with low feature importances from the original 37-dimensional feature vector and sets up the optimal 25-dimensional feature vector subset.

3. Result

3.1. Validation and Evaluation

In order to evaluate the performance of the random forest classifier, True Positive (TP), False Positive (FP), and False Negative (FN) can be calculated according to the labeled Ground Truth and the prediction samples from test set. According to the prediction results of the samples in the test set, several evaluation standards like precision, recall, F1 score, and confusion matrix can be calculated to evaluate the performance of the classifier.

Precision indicates how many samples predicted to be positive are truly positive samples.

\[
Precision = \frac{TP}{TP + FP}
\]  

Recall indicates how many positive samples among all samples were predicted correctly.

\[
Recall = \frac{TP}{TP + FN}
\]

F1 score is based on the harmonic mean of precision and recall.

\[
F_1 = \frac{2 \times (Precision \times Recall)}{Precision + Recall}
\]

In the experiment, the test set contained 9412 samples, each behaviour class had 2353 samples. As a result, 8832 samples were classified correctly while 580 samples were misclassified. The accuracy was 93.84% on the test set. The specific test results are shown in Table 2 and Figure 6.
Table 2. Result.

| Behavior | Correct samples | Precision | Recall | F1 scores |
|----------|-----------------|-----------|--------|-----------|
| Stand    | 2342            | 0.980     | 0.995  | 0.988     |
| Eat      | 2011            | 0.926     | 0.854  | 0.889     |
| Grovel   | 2230            | 0.898     | 0.948  | 0.923     |
| lie      | 2249            | 0.949     | 0.958  | 0.952     |
| Mean     | 2208            | 0.939     | 0.938  | 0.938     |

Based on the above evaluation results of the test set, it can be seen from Table 2 that the average accuracy rate of each behaviour class reached 93.9%, the average recall of each behaviour class reached 93.8% and the F1 score reached 0.938, indicating that less than 1/10 of the positive samples were misclassified. On the other hand, the F1 score of the other behaviours exceeded 0.9 except the eating behaviour. This shows that the classifier can easily confuse the eating behaviour with other behaviours. The confusion matrix in Figure 6 shows that the classifier almost misclassified eating behaviours into lying and grovelling.

At the same time, this paper also uses decision tree model and BP neural network model to conduct comparative experiments with the random forest classifier based on the same test conditions. In order to compare the performance of these three classifiers, the parameters of these three classifiers were adjusted to be optimal at first. The min-samples-leaf (The minimum number of samples about leaf nodes) of the decision tree is set to 5 while the BP neural network adopts a structure of an input layer, 3 hidden layers (300 nodes per layer), and an output layer (4 classification nodes). Figure 7 shows the ROC curves predicted by three different methods under the same test set.

It can be seen from the Figure 7 that under the same TPR condition, the FPR of the random forest is the smallest. The ROC curve of random forest is closer to the upper left of the coordinate axis and its AUC reaches the largest of three classifiers at 0.957. The experimental result shows that the effect of random forest on the classification and recognition of pig behaviour is better than that of decision trees and BP neural networks under the same test set sample conditions.

3.2. Maternal Ability Index

This paper evaluated the pig’s maternal ability index based on the caution degree of various high-risk[7]. Many researches shows that lying and grovelling behaviours are important standard of the index. The risk scores of each behaviour were defined as shown in Table 3.
Table 3. Definition of High-risk Behavior.

| Behavior           | Sample Size | Score |
|--------------------|-------------|-------|
| Stand to grovel    | 64          | 3     |
| Grovel to stand    | 58          | 1     |
| Stand to lie       | 4           | 4     |
| Lie to stand       | 10          | 2     |
| Grovel to lie      | 68          | 5     |
| Lie to grovel      | 61          | 5     |

Then the risk index can be calculated as follows by combining the frequency of each behaviour:

\[ S_{\text{danger}} = \sum_{i=1}^{k} s_i p_i \] (18)

\( s_i \) is the frequency of behaviour and \( p_i \) is the corresponding score.

Converting the risk index to maternal ability index:

\[ \text{Maternal\_ability} = 10(1 - S_{\text{danger}}) \] (19)

The total score was set to 10. It means that the higher the score of the maternal ability index is, the more careful pigs will be when moving. In other words, the mortality rate of piglets crushed before weaning will be lower. This paper took the pig tested as an example to count the frequency of each behaviour. The risk index was 0.047 and the maternal ability index was 9.529 by calculating based on the behaviour recognition results. This indicated that the pig tested had a higher level of maternal ability.

4. Conclusion

This paper collected pig behaviour data by wearing a triaxial acceleration sensor at the right ear of a pig. After rough data analysis, this paper adopted a moving average filter to smooth the raw data. The optimal feature subset was constructed by combining the zero-crossing point, the upper and lower quartile points, the maximum value, the pitch angle, and the roll angle with the traditional features like mean and standard deviation characteristics after feature selection. By constructing the optimal feature subset proposed in this paper, the random forest classifier adopted is more suitable for situations where the behaviour changes frequently. The overall recognition rate of the classifier is 93.8%. The performance of random forest was better than that of decision trees and BP neural networks under the same test condition. While the recognition effect of eating behaviour is not ideal. Finally, this paper proposed an evaluation model calculated the maternal ability index according to the classification results, which can provide a good data basis for evaluating pigs’ maternal ability and selecting high-quality breeding pigs.

References

[1] Lay Jr D C. Management tips to reduce pre-weaning mortality[J]~/'Forty-sixth Annual North Carolina Pork Conference, 2002.
[2] Chidgney K L, Morel P C H, Stafford K J, et al. Sow and piglet productivity and sow reproductive performance in farrowing pens with temporary crating or farrowing crates on a commercial New Zealand pig farm[J]. Livestock Science, 2015, 173 (3): 87 -94.
[3] Moustsen V A, Hales J, Lahrmann H P, et al. Confinement of lactating sows in crates for 4 days after farrowing reduces piglet mortality [J]. Animal, 2013, 7 (4) :648 - 654.
[4] Thompson R, Matheson S M, Plötz T, et al. Porcine lie detectors: Automatic quantification of posture state and transitions in sows using inertial sensors[J]. Computers and electronics in agriculture, 2016, 127: 521-530.
[5] Swiatek J, Grzech A, Swiatek P, et al. Advances in Systems Science: Proceedings of the International Conference on Systems Science 2013 (ICSS 2013)[J]. Springer Ebooks, 2013, 240.
[6] San S R, Montero J M, Barrac R, Fernandof, Manuel P. Feature extraction from smartphone inertial signals for human activity segmentation [J]. Signal Processing, 2016, 120: 359-372.
[7] Marek, Špinka, and, et al. Dimensions of maternal behaviour characteristics in domestic and wild×domestic crossbred sows[J]. Applied Animal Behaviour Science, 2000.