Dataset: Horse Movement Data and Analysis of its Potential for Activity Recognition

Jacob W. Kamminga
University of Twente
Enschede, The Netherlands
j.w.kamminga@utwente.nl

Nirvana Meratnia
University of Twente
Enschede, The Netherlands
n.meratnia@utwente.nl

Paul J.M. Havinga
University of Twente
Enschede, The Netherlands
p.j.m.havinga@utwente.nl

ABSTRACT

We describe and analyze a dataset that comprises horse movement. Data was collected during horse riding sessions and when the horses freely roamed the pasture over 7 days. The dataset comprises 1.8 million 2-second data samples from 18 individual horses, of which 93303 samples from 11 subjects were labeled. Sensor devices were attached to a collar around the neck of the horses while the orientation was not fixed. The devices contained a 3-axis accelerometer, gyroscope, and magnetometer that were sampled at 100 Hz. To demonstrate how this dataset can be used, we evaluated a Naive Bayes classifier with leave-one-out validation. Our results show that a performance of 90 % accuracy can be achieved using only the 3D acceleration vector as input. Furthermore, we demonstrate the effect of increased complexity, parameter tuning, and class balancing on classification performance and identify open research challenges. The complete dataset is available online with open access at the 4TU.Centre for Research Data [9].

CCS CONCEPTS

• **Information systems** → *Data mining*; • **Theory of computation** → *Machine learning theory*.

KEYWORDS

Animals, Horses, Activity Recognition, Accelerometer, Gyroscope, Compass, IMU, Orientation Independent, Neck

ACM Reference Format:

Jacob W. Kamminga, Nirvana Meratnia, and Paul J.M. Havinga. 2019. Dataset: Horse Movement Data and Analysis of its Potential for Activity Recognition. In *The 2nd Workshop on Data Acquisition To Analysis (DATA'19), November 10, 2019, New York, NY, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3359427.3361908

1 INTRODUCTION

The behavior of animals contains a tremendous wealth of information that not only provides insights into their life and well-being, but also their environment [2, 8, 12, 14, 16]. Animal activities can be classified from motion data [10, 12]. In this paper, we describe the collection process of a horse movement dataset, provide a brief evaluation, and discuss research challenges in the area of Animal Activity Recognition (AAR) that can be investigated using this dataset.

We chose to monitor horses that were ridden in an equestrian facility because they are exercising various activities over the course of a day. This could ease the task of collecting and labeling relatively large amounts of movement data from several activities and resulted in a more balanced dataset for different gaits. The dataset contains labeled data from activities that are very similar but slightly different, e.g. the difference in gait with- and without a rider on the horse. The largest part of our dataset is unlabeled data (denoted by `null`). This dataset is particularly suitable to benchmark unsupervised representation learning algorithms for AAR. Paper [13] on unsupervised representation learning for AAR uses part of this dataset. Other use cases for the dataset include: gait analysis and comparison, feature selection for AAR, and transfer learning. The data might be useful to validate AAR methods for other quadruped animals within the *Equidae* family, such as donkeys or zebras.

To demonstrate how this dataset can be used we trained and tested a Naive Bayes (NB) classifier. We demonstrate the effect of increased complexity in the classification problem on the performance of the classifier. Dealing with class imbalance is an ongoing and important area of research in machine learning [4, 5]. Therefore, we evaluate the classifier with and without balancing the dataset and discuss the research challenges.

2 DATA ACQUISITION AND LABELING

Movement data was collected from 18 individual horses that performed 17 different activities described in Table 1. All experiments with the animals complied with Dutch ethics law concerning working with animals. Ground truth was collected by cameras during riding sessions. More natural activities were observed while they were left to roam freely in an outdoor pasture as shown in Figure 1.

A sensor device from Gulf Coast Data Concepts [3] was attached to the neck by means of a collar fabricated from hook and loop fastener. The location was chosen so that the sensors could be worn without a saddle or halter. Additionally, this position is often used

Figure 1: Two subjects during the outdoor collection process

DATA'19, November 10, 2019, New York, NY, USA

ACM ISBN 978-1-4503-6993-0/19/11…$15.00
https://doi.org/10.1145/3359427.3361908

Table 1: Observed daytime activities exercised by horses

Activity	Description
Standing	Horse standing on 4 legs, no movement of head, standing still
Walking natural	No rider on horse, the horse puts each hoof down one at a time, creating a four beat rhythm
Walking rider	Rider on horse, the horse puts each hoof down one at a time, creating a four beat rhythm
Trotting natural	No rider on horse, 2 beat gait, one front hoof and its opposite hind hoof come down at the same time, making a two-beat rhythm, different speeds possible but always 2 beat gait
Trotting rider	Rider on horse, 2 beat gait, one front hoof and its opposite hind hoof come down at the same time, making a two-beat rhythm, different speeds possible but always 2 beat gait
Galloping natural	No rider on horse, one hind leg strikes the ground first, and then the other hind leg and one foreleg come down together, the the other foreleg strikes the ground. This movement creates a three-beat rhythm
Galloping rider	Rider on horse, can be right or left leaning, one hind leg strikes the ground first, and then the other hind leg and one foreleg come down together, the the other foreleg strikes the ground. This movement creates a three-beat rhythm
Jumping	All legs off the ground, going over an obstacle
Grazing	Head down in the grass, eating and slowly moving to get to new grass spots
Eating	Head is up, chewing and eating food, usually eating hay or long grass
Head shake	Shaking head alone, no body shake, either head up or down
Shaking	Shaking the whole body, including head
Scratch biting	Horse uses its head/mouth to scratch mostly front legs
Rubbing	Scratching body against an object, rubbing its body to scratch itself
Fighting	Horses try to bite and kick each other
Rolling	Horse laying down on ground, rolling on its back, from one side to another, not always full roll
Scared	Quick sudden movement, horse is startled

in studies that monitor wildlife such as zebra [6] which increases the usability of our dataset for research related to other animals. The orientation of the sensors was not fixed to be able to evaluate AAR approaches that are robust against the sensor orientation. Different colors were used for the collars to ease the identification of the horses in the videos. The sensor devices contained a 3-axis accelerometer, gyroscope and magnetometer with a sampling rate of 100 Hz.

The data was annotated with our labeling tool [11, 12] that is publicly available online [7]. Videos were synchronized with sensor data using metadata. Annotations were added by clicking on the visualized movement data. When a horse was performing multiple activities simultaneously, the activity that was mainly exercised was chosen as the label. For example, when a horse was eating while slowly walking, this activity was labeled as `grazing`, because the movement is part of the grazing behaviour. To minimize ambiguity in the labeling, all labeled data were visually inspected and corrected by a single person. The data from 6 subjects and 6 activities were labeled more extensively.

2.1 Dataset

The complete dataset is available online with open access at the 4TU.Centre for Research Data [9]. The data is organized in segments of continuous raw sensor data. Each segment has a unique identifier. Segments can have a varying length that depends on how long the subject exercised a given activity. The maximum segment length is 10 seconds because this improves the class balance when separating segments into train, tune, and test sets prior to windowing [12]. Each row in the dataset denotes one raw data sample. The columns of the dataset are described in Table 2 along with the sensor settings.

Table 2: Column description

Column name	Description	Sampling rate (hz)	Range
Ax	Raw data from accelerometer x-axis	100	8g
Ay	Raw data from accelerometer y-axis	100	8g
Az	Raw data from accelerometer z-axis	100	8g
Gx	Raw data from gyroscope x-axis	100	2000 °/s
Gy	Raw data from gyroscope y-axis	100	2000 °/s
Gz	Raw data from gyroscope z-axis	100	2000 °/s
Mx	Raw data from compass (magnetometer) x-axis	12	
My	Raw data from compass (magnetometer) y-axis	12	
Mz	Raw data from compass (magnetometer) z-axis	12	
A3D	l2-norm (3D vector) of accelerometer axes	100	
G3D	l2-norm (3D vector) of gyroscope axes	100	
M3D	l2-norm (3D vector) of compass axes	12	
label	Label that belongs to each row's data		
segment	Each activity has been segmented with a maximum length of 10 seconds. Data within one segment is continuous. Segments have been numbered incrementally.		
subject	Subject identifier		

A summary of the data distribution is shown in Table 3. Each sample was obtained by windowing the activity segments with a window length of 2 seconds and 50 % overlap. We aggregated the data listed in Table 3 and grouped some of the activity classes. Figure 2 shows a visual representation of the data distribution over three statistical summary features. It can be seen that data clusters are overlapping and activities such as `galloping` and `head-shake` are more scattered.

Table 3: Amount of data samples per (grouped) activity. Each sample denotes a 2 second window of raw data.

Activity	null	standing	walking-rider	walking-natural	trotting	galloping	eating	other	total
nr samples	1191658	5297	35425	3609	25782	4036	18110	1044	1284961
fraction of labeled		6%	38%	4%	28%	4%	19%	1%	

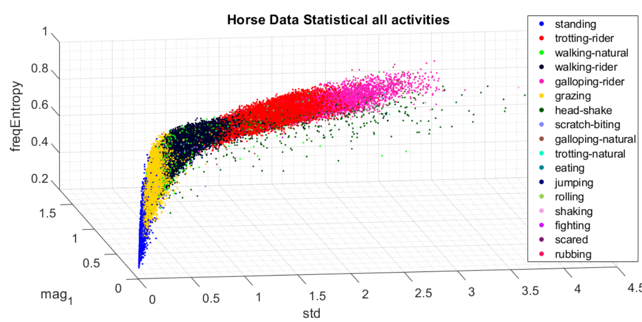

Figure 2: 3D data distribution over: frequency entropy, most dominant frequency component, and standard deviation.

3 EVALUATION

To demonstrate how this dataset can be used and to evaluate what AAR performance can be achieved with this dataset, we trained and tested a Naive Bayes (NB) classifier. NB was chosen because it has a good complexity to performance ratio for AAR [10, 12]. We used data from 6 subjects and 6 activities that contained sufficient labeled data so that leave-one-out cross-validation could be used. The activities shown in Table 3 were used during the evaluation, excluding `null` and `other`. Increased complexity in the classification problem was achieved by using 6 instead of 5 activities by dividing `walking` into `walking-natural` and `walking-rider`. We used 21 summary statistics that are commonly used for Activity Recognition (AR) [12] to describe the data: minimum, maximum, mean, standard deviation, median, 25th and 75th percentile, mean low pass signal, mean rectified high pass, skewness, kurtosis, zero-crossing rate, principal frequency, spectral energy, frequency entropy, and the six most dominant frequency component magnitudes. The evaluations were performed using Matlab [15]. We used only the magnitude of the 3D vector (ℓ^2-norm) of the accelerometer because it is orientation-independent and energy efficient [12]. The data was standardized through a Z-transformation while the test set was not used to calculate the mean and standard deviation. Data was balanced by simultaneously using random under-sampling [5] for majority classes and the Synthetic Minority Over-sampling Technique (SMOTE) for minority classes [1]. The classification performances (accuracy and F_1) for all scenarios are shown in Figure 3.

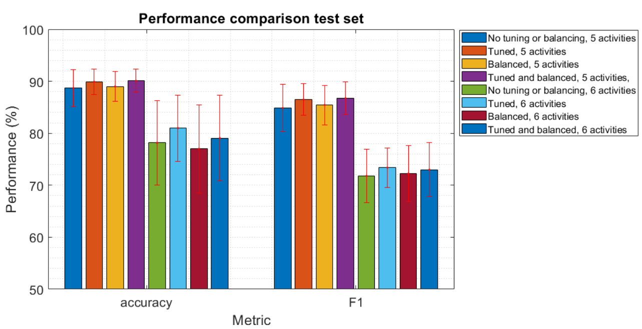

Figure 3: Comparison of the effect of complexity, tuning, and balancing on performance

The results in Figure 3 show that the F_1 performance for a simpler AAR task improved by tuning (1.6 %) and just slightly by balancing (0.6 %). In a more complex scenario with 6 activities, the improvements were 1.6 %, and 0.5 %, respectively. In this scenario, the balancing did slightly improve the F_1 performance but decreased accuracy. The decreased accuracy is probably due to the random under-sampling of the majority classes, which worsened their true positive rates. Thus, balancing through random under-sampling should only be done when the minority class is important. There was a 11.2 % drop in accuracy, and 13.7 % drop in F_1-performance when the complexity of the AAR task was increased from 5 to 6 activities.

Figure 4 shows the confusion matrix for AAR with 6 activities. The matrix shows the aggregated results of leave-one-out validation. The training data was balanced, and tuning was used. The last two columns denote the percentage of true and false positives per class. The eating activity is confused with standing and walking. This can be explained because during grazing and eating the horses are either standing still or slowly walking. Galloping and trotting are also often confused; this is not surprising because these activities largely overlap. A part of this confusion can also occur due to miss interpretation by the annotator during labeling as it is not always clear when the activity transitions occur. Walking-natural and walking-rider are mostly confused. The walking-rider class (38 %) is much larger than the walking-natural class (4 %) and the NB classifier is clearly biased towards the majority class, even when balancing is applied. We think that this has to do with limitations of the SMOTE [1] technique.

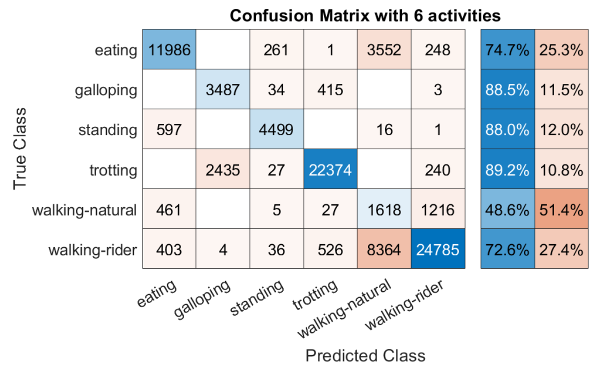

Figure 4: Predictions with 6 activities

Our results can probably be improved by investigating other features, balancing techniques, and classifiers. Moreover, this dataset can be used to exploit the potential of the vast amount of unlabeled data (`null`) to improve AAR performance.

4 CONCLUSION

We discussed the data collection process and composition of an extensive horse movement dataset. Moreover, we evaluated the labeled data through a NB classifier. In our evaluation, balancing the dataset was a trade-off between overall performance and the performance of the minority class. It is an open research challenge to improve or eradicate this trade-off. It seems somewhat wasteful to use random under-sampling since we are effectively discarding valuable labeled data. Therefore, we would like to invite other researchers to investigate better solutions to the balancing trade-off. Our results showed that parameter tuning for the NB improved the F_1 performance. The balancing of the data slightly improved the F_1 performance with less than a percent and even worsened the accuracy when using 6 activities. This dataset allows researchers to exploit the vast amount of null data to improve the performances we reported in this paper and address open research challenges.

ACKNOWLEDGMENTS

This research was supported by the Smart Parks Project, which involves the University of Twente, Wageningen University & Research, ASTRON Dwingeloo, and Leiden University. The Smart Parks Project is funded by the Netherlands Organisation for Scientific Research (NWO). We would like to thank the Horstlinde horse stable in Enschede, the Netherlands for their kind cooperation during the collection of this dataset. We would like to thank Lara Janßen, Lieke Hamelers, and Heleen Visserman for their help during the data collection and labeling process.

REFERENCES

[1] Nitesh V Chawla, Kevin W Bowyer, and Lawrence O Hall. 2002. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research 16 (2002), 321–357. https://doi.org/10.1613/jair.953

[2] Jamali Firmat Banzi. 2014. A Sensor Based Anti-Poaching System in Tanzania National Parks. International Journal of Scientific and Research Publications 4, 4 (2014), 1–7.

[3] LLC Gulf Coast Data Concepts. 2019. Human Activity Monitor: HAM. online. (2019). http://www.gcdataconcepts.com/ham.html

[4] Haibo He and Edwardo A. Garcia. 2009. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering 21, 9 (2009), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

[5] Justin M. Johnson and Taghi M. Khoshgoftaar. 2019. Survey on deep learning with class imbalance. Journal of Big Data 6, 1 (2019), 27. https://doi.org/10.1186/s40537-019-0192-5

[6] Philo Juang, Hidekazu Oki, Yong Wang, Margaret Martonosi, Li Shiuan Peh, and Daniel Rubenstein. 2002. Energy-efficient computing for wildlife tracking. ACM SIGOPS Operating Systems Review 36, 5 (2002), 96. https://doi.org/10.1145/635508.605408

[7] Jacob W. Kamminga. 2019. Matlab Movement Data Labeling Tool. (8 2019). https://doi.org/10.5281/zenodo.3364004

[8] J. Kamminga, E. Ayele, N. Meratnia, and P. Havinga. 2018. Poaching detection technologies-A survey. Sensors (Switzerland) 18, 5 (2018), 1474. https://doi.org/10.3390/s18051474

[9] Jacob W. Kamminga. 2019. Horsing Around – A Dataset Comprising Horse Movement. (7 2019). https://doi.org/10.4121/uuid:2e08745c-4178-4183-8551-f248c992cb14

[10] Jacob W. Kamminga, Helena C. Bisby, Duc V. Le, Nirvana Meratnia, and Paul J. M. Havinga. 2017. Generic online animal activity recognition on collar tags. In Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers on - UbiComp '17. Association for Computing Machinery, Maui, HI, 597–606. https://doi.org/10.1145/3123024.3124407

[11] Jacob W Kamminga, Michael Jones, Kevin Seppi, Nirvana Meratnia, and Paul J.M. Havinga. 2019. Synchronization between Sensors and Cameras in Movement Data Labeling Frameworks. In The 2nd Workshop on Data Acquisition To Analysis (DATA'19), November 10, 2019, New York, NY, USA. Association for Computing Machinery, New York, NY. https://doi.org/10.1145/3359427.3361920

[12] Jacob Wilhelm Kamminga, Duc V. Le, Jan Pieter Meijers, Helena Bisby, Nirvana Meratnia, and Paul J.M. Havinga. 2018. Robust Sensor-Orientation-Independent Feature Selection for Animal Activity Recognition on Collar Tags. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies IMWUT 2, 1 (2018), 1–27. https://doi.org/10.1145/3191747

[13] Jacob W Kamminga, Duc V. Le, Nirvana Meratnia, and Paul J.M. Havinga. 2019. Deep Unsupervised Representation Learning for Online Animal Activity Recognition. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. (2019), submitted.

[14] Paula Martiskainen, Mikko Järvinen, Jukka-Pekka Skön, Jarkko Tiirikainen, Mikko Kolehmainen, and Jaakko Mononen. 2009. Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. Applied Animal Behaviour Science 119, 1-2 (2009), 32–38. https://doi.org/10.1016/j.applanim.2009.03.005

[15] MATLAB. 2018. version 9.5.0.944444 (R2018b). The MathWorks Inc., Natick, Massachusetts.

[16] Ran Nathan. 2008. An emerging movement ecology paradigm. Proceedings of the National Academy of Sciences of the United States of America 105, 49 (2008), 19050–19051. https://doi.org/10.1073/pnas.0808918105