G Model
NSM-7732; No. of Pages 7
ARTICLE IN PRESS
Journal of Neuroscience Methods xxx (2017) xxx–xxx
Contents lists available at ScienceDirect
Journal of Neuroscience Methods
journal homepage: www.elsevier.com/locate/jneumeth

Learning to recognize rat social behavior: Novel dataset and cross-dataset application

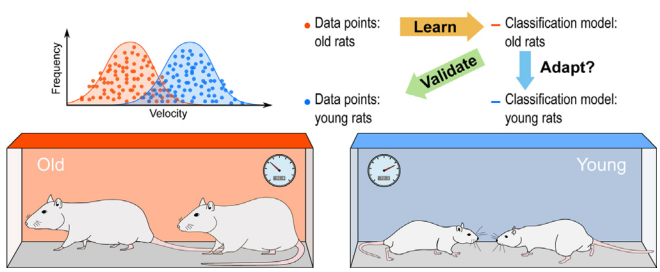
Malte Lorbach[a,c,*], Elisavet I. Kyriakou[b,c], Ronald Poppe[a], Elsbeth A. van Dam[c], Lucas P.J.J. Noldus[c], Remco C. Veltkamp[a]

[a] Department of Information and Computing Sciences, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands
[b] Department of Cognitive Neuroscience, Radboud University Medical Centre, Kapittelweg 29, 6525 EN Nijmegen, The Netherlands
[c] Noldus Information Technology BV, Nieuwe Kanaal 5, 6709 PA Wageningen, The Netherlands

HIGHLIGHTS

- Using video analysis to measure rodent social behavior receives growing attention.
- Developing and validating automated measuring methods requires annotated datasets.
- We introduce the first, publicly available rat social interaction dataset, RatSI.
- Cross-dataset validation of automated methods ensures validity in practice.
- Validity may be expanded by developing novel dataset adaptation techniques.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 31 January 2017
Received in revised form 4 May 2017
Accepted 5 May 2017
Available online xxx

Keywords:
Social behavior
Rodents
Automated behavior recognition
Dataset

ABSTRACT

Background: Social behavior is an important aspect of rodent models. Automated measuring tools that make use of video analysis and machine learning are an increasingly attractive alternative to manual annotation. Because machine learning-based methods need to be trained, it is important that they are validated using data from different experiment settings.
New method: To develop and validate automated measuring tools, there is a need for annotated rodent interaction datasets. Currently, the availability of such datasets is limited to two mouse datasets. We introduce the first, publicly available rat social interaction dataset, RatSI.
Results: We demonstrate the practical value of the novel dataset by using it as the training set for a rat interaction recognition method. We show that behavior variations induced by the experiment setting can lead to reduced performance, which illustrates the importance of cross-dataset validation. Consequently, we add a simple adaptation step to our method and improve the recognition performance.
Comparison with existing methods: Most existing methods are trained and evaluated in one experimental setting, which limits the predictive power of the evaluation to that particular setting. We demonstrate that cross-dataset experiments provide more insight in the performance of classifiers.
Conclusions: With our novel, public dataset we encourage the development and validation of automated recognition methods. We are convinced that cross-dataset validation enhances our understanding of

* Corresponding author.
E-mail address: m.t.lorbach@uu.nl (M. Lorbach).

http://dx.doi.org/10.1016/j.jneumeth.2017.05.006

rodent interactions and facilitates the development of more sophisticated recognition methods. Combining them with adaptation techniques may enable us to apply automated recognition methods to a variety of animals and experiment settings.

1. Introduction

Social interaction is an important component of psychiatric research as well as neurological testing of animal models in behavioral neuroscience (Urbach et al., 2010). As part of the emotional screening of a model it relates to aspects such as anxiety, stress, play and sexual behavior (File and Seth, 2003). Moreover, abnormal social behavior can be indicative of a psychopathology (Peters et al., 2015) and can therefore inform us of the onset or progression of conditions such as schizophrenia (Wilson and Koenig, 2014), Huntington's (Urbach et al., 2014) and Alzheimer's disease (Lewejohann et al., 2009) as well as Rett syndrome (Veeraragavan et al., 2016). Including social behavior in rodent models therefore increases their predictive power and value for the transition to clinical trials and treatments for humans (Peters et al., 2015; Richardson, 2015).

Whether we seek to enhance our understanding of social behavior or include it in a rodent model, we need to objectively measure and quantify it. Traditionally, this involves annotating the interactions among rodents in hours of either live observations or video recordings of social interaction tests. While this can be done manually, it is time-consuming and subjective. Subjectivity may be reduced by a meticulously defined ethogram and thorough training of the human annotators at the cost of additional work.

An attractive alternative to manual scoring are automated measuring tools (Schaefer and Claridge-Chang, 2012; Steele et al., 2007; Egnor and Branson, 2016; Noldus et al., 2001). Such tools track the locations of the rodents in video recordings and provide quantitative measures such as the distance traveled and the time spent in proximity (Spruijt et al., 1992; Sams-Dodd, 1995; Dell et al., 2014). Recent advances in video analysis have made the tracking of rodents more robust and accurate (Hong et al., 2015; Pérez-Escudero et al., 2014). This allows us to take the next step and consider the automated recognition of specific *interactions* such as approaching and following. Although the interaction categories that can currently be handled automatically are not as fine-grained and large in quantity as the categories that humans are able to annotate, automated methods can still support manual annotation and reduce labor. For example, by providing a first segmentation into these broader categories with high accuracy, the human effort can be reduced to annotating fine-grained behaviors only in the relevant video segments instead of the full length of the video.

The automated recognition of interactions typically involves applying classification algorithms to a quantified representation (features) of the visual information in the video (Hong et al., 2015; Kabra et al., 2012; Burgos-Artizzu et al., 2012; Giancardo et al., 2013). The features are derived from the tracked animals and may include velocity and distance. In order to distinguish between the different interactions, the parameters in the classification algorithms are determined using labeled feature examples. In this training phase, the classifier learns the similarities among the examples and thereby creates a *model* of each interaction. For instance, it may learn that whenever a rat approaches another, it moves at a certain velocity while the distance between the two decreases. It is important how the classifier learns such models. A classifier that simply "remembers" the feature values will not perform well on unseen examples which have slightly different values. Instead, it must generalize from the empirical examples to the inherent variations of the interaction classes.

Generally, there are two types of variation in the examples of an interaction. First, two animals will perform the same interaction slightly differently every time, for instance, at a slightly different velocity or from a different starting point. We consider this the natural variation of an interaction. Second, there is a systematic bias in the natural variation that depends on the tested population and the environment in which the interactions are observed. Rats from the tested population, which is characterized by the genetic background, the age and possibly the progress of a condition or its treatment, could for example move slower than rats from another population. The environment, which is often created by the researcher to study specific behaviors, comprises factors such as the available space and the presence of hiding places or novel objects that may allow or prevent interactions to be performed in certain ways.

As a consequence, the models learned by the classifier depend on the distribution of training examples with respect to the systematic bias. If the bias changes due to modifications to the animal population or the environment (Schneider and Levine, 2014), the models could lose their effectiveness.

Therefore, when we evaluate the performance of a trained classifier, we typically use test examples that follow the same distribution as the training examples. Both training and test examples are usually taken from a dataset of video recordings of one specific experiment (Hong et al., 2015; Kabra et al., 2012; Burgos-Artizzu et al., 2012; Giancardo et al., 2013; Eyjolfsdottir et al., 2014; Kuehne et al., 2016). That ensures that the bias is kept constant during evaluation and that we obtain a plausible measure of the performance.

This evaluation scheme becomes critical when we apply the trained classifier in practice. Beyond the tested experiment setting, the evaluation is of limited value as it cannot predict the classifier's performance in another setting. Given the difficulty of precisely replicating experiment settings (Crabbe et al., 1999) as well as appeals to increase experiment heterogeneity (Richter et al., 2009), we argue for an evaluation of interaction classifiers across settings and therefore across datasets. Only with cross-dataset evaluation can we be confident about the performance of the classifier in practice (van Dam et al., 2013) and judge to which settings we can apply it without retraining.

We argue that there is a need for datasets for at least two purposes: to train classifiers and to evaluate them across experiment settings. Currently, there are only two rodent social behavior datasets publicly available for researchers and both focus on mice: the Caltech Resident-Intruder Mouse dataset (CRIM13) (Burgos-Artizzu et al., 2012) and the Mice Behavior Analysis dataset (MBADA) (Giancardo et al., 2013).

Given the increasing interest in rats for studying social behavior (Veeraragavan et al., 2016; Homberg et al., 2016), we introduce the first rat social interaction dataset (RatSI).[1] It contains 2.25 h of annotated video recordings of two interacting rats in an open-field arena, including accurate 3-point tracking of the animals. The

[1] http://www.noldus.com/innovationworks/phenorat-dataset.