Enabling naturalistic neuroscience through behavior mining: Analysis of long-term human brain and video recordings

Satpreet H. Singh (satsingh@uw.edu)
University of Washington, Seattle, WA 98195 USA

Steven M. Peterson (stepeter@uw.edu)
University of Washington, Seattle, WA 98195 USA

Rajesh P. N. Rao (rao@cs.washington.edu)
University of Washington, Seattle, WA 98195 USA

Bingni W. Brunton (bbrunton@uw.edu)
University of Washington, Seattle, WA 98195 USA

Abstract

Much of our understanding in human neuroscience has been informed by data collected in pre-designed and well-controlled experimental tasks, where timings of cues, stimuli, and behavioral responses are known precisely. Recent technological advances have enabled us to study longer and increasingly naturalistic brain recordings, giving rise to a new paradigm named "naturalistic neuroscience" where neural computations associated with spontaneous behaviors are studied. Analyzing such unstructured, long-term, and multi-modal data with no *a priori* experimental design remains very challenging. Here we present an automated approach for analysing naturalistic datasets using behavior mining. Our analysis pipeline robustly uncovers and annotates instances of human upper-limb movements in long-term naturalistic behavior data ($\approx$18 million video frames per patient) using algorithms from computer-vision, time-series segmentation, and string pattern-matching. We analyze simultaneously recorded human electrocorticography (ECoG) brain recordings to uncover neural correlates associated with these naturalistic events, and show that they corroborate prior results from traditional controlled experiments. We also demonstrate the efficacy of our approach as a source of training data for brain-computer interface decoders.

Keywords: naturalistic neuroscience; big data; computer vision; data mining; brain-computer interfaces

Introduction

Human neuroscience experiments are traditionally performed in controlled laboratory settings using carefully designed experiments to isolate the effects of a treatment or perturbation from a baseline control condition. Typically, the timing of cues or stimuli presented to the human subject, as well as the timing and measurements of their behavioral responses, are known precisely. Enabled by advances in brain recording technology, "naturalistic neuroscience" (Huk, Bonnen, & He, 2018) is an emerging paradigm to study neural computations associated with spontaneous behaviors in freely behaving subjects.

Accompanying this trend, a rich set of computational tools are being developed for automated analysis of animal behavior, a field also known as "computational ethology" (Anderson & Perona, 2014; Todd, Kain, & de Bivort, 2017; Berman, 2018; Pereira et al., 2019; Markowitz et al., 2018; Wiltschko et al., 2015). These methods process the video recordings of one or more animals through an extensive analysis pipeline, often including steps such as: segmenting the subject(s) from the background, centering and rotating the subject to a standard pose, imputing missing frames, tracking specific joints across frames, and classifying actions.

There have been some recent attempts at applying such automated analysis in human neuroscience (N. X. R. Wang, Olson, Ojemann, Rao, & Brunton, 2016; N. X. Wang, Farhadi, Rao, & Brunton, 2018; Alasfour et al., 2019; Chen et al., 2018). Though there exists a large body of computer-vision research concerned with human action recognition (Ramasamy Ramamurthy & Roy, 2018), most methods are concerned with distinguishing coarse-grained activities such as sitting vs. walking, and consist of techniques requiring a huge amount of labeled training data. To answer the kinds of questions that neuroscientists and neuroengineers are focused on, we instead need to be able to produce a fine-grained segmentation with sub-second temporal resolution, ideally using none or few behavior labels ahead of time.

Here we take a behavior-mining approach that robustly uncovers instances of human upper-limb movements in long-term, naturalistic behavior data and connects them to simultaneously recorded brain recordings. In particular, we use pose estimation algorithms to extract human arm pose trajectories and then segment these trajectories in time using unsupervised latent-variable models. Interestingly, having a discrete sequential representation of pose allows us to simplify the problem of detecting behavioral events to that of pattern matching on strings. Our approach also extracts additional behavioral metadata associated with the events, such as movement angle, magnitude, and duration. Lastly, we uncover neural correlates associated with the events by analyz-

ing the simultaneously recorded human intracranial electro-corticography (ECoG) brain recordings; these naturalistic neural correlates further corroborate results from traditional, controlled experiments. Our preliminary investigations suggest that these naturalistic events may be used as training data for brain-computer interface (BCI) decoders.

Dataset

We use simultaneous brain and behavior recordings from consenting epilepsy patients undergoing long-term monitoring ($\approx$ $7 - 10$ days) at the University of Washington (UW) Harborview Medical Center in Seattle, WA, after obtaining approval from the UW Institutional Review Board's human subject division. Intracranial recordings are acquired at a sampling rate of 1000 Hz and are obtained through ≈ 80 electrodes on the cortical surface. Our data collection is opportunistic, with electrode locations and duration of recording being determined by clinical needs. Audio and video recording are obtained using a wall-mounted RGB/Infrared camera recording at ≈ 30 frames per second ($\approx$18 million video frames/patient over 10 days). Together, these amount to ≈ 250 GB of data per patient.

Data Analysis Pipeline

Video recordings are processed through a pipeline of pose-estimation, time-series segmentation, event detection, and metadata extraction.

Human Pose Estimation

We train a markerless pose-estimation tool (Mathis et al., 2018) on a small random sample (≈ 1000 frames) of the video recordings using manual annotations for multiple keypoints on the patient's body, namely, the nose, and both wrists, elbows, shoulders, and ears (Figure 1). The trained pose estimator is scaled using cloud computing to extract pose trajectories for the entire duration of the video recording.

Figure 1: Video frame showing typical patient pose when not asleep: Tracked body keypoints are shown by small colored circles. (Patient's face masked for anonymity.)
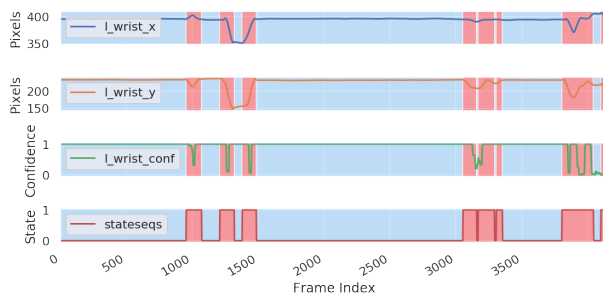

Figure 2: Autoregressive Hidden Markov Model with two discrete latent states robustly segments the left-wrist's pose dynamics into *rest* (State 0) and *move* (State 1) states with high temporal precision: `l_wrist_x[y]` is the x (or y) coordinates [in pixels] and `l_wrist_conf` is the tracking confidence of the left-wrist over time. `stateseqs` is the discrete latent-state sequence and has been used to shade times across all subplots.

Pose Time-series Segmentation

We seek a segmentation of the pose time series that is easily interpretable and temporally precise. With this in mind, we apply a first-order autoregressive hidden Markov model (Wiltschko et al., 2015) with two latent states to the pose time series of a single keypoint at a time (e.g. left wrist, see Figure 2). Segmentation results in a representation of the pose dynamics into discrete *rest* and *move* states with high (per-frame) temporal precision. It is also robust to variation in lighting, camera angle, and level of activity in the video.

Event Mining

Discretizing the continuous pose time-series reduces the problem of finding behavioral *events* to a string pattern matching problem, for which there exist several algorithmic tools. We use regular expressions to extract events from this discrete representation. For example, *movement-initiation* events are found by searching for a pattern of 15 consecutive *rest* states followed by 15 consecutive *move* states. Similarly, *no-movement* events are found by looking for a state sequence
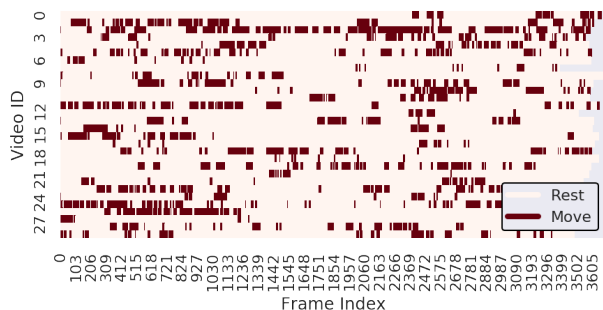

Figure 3: Raster plot representation of extracted discrete latent state trajectories for several video clips stacked vertically.
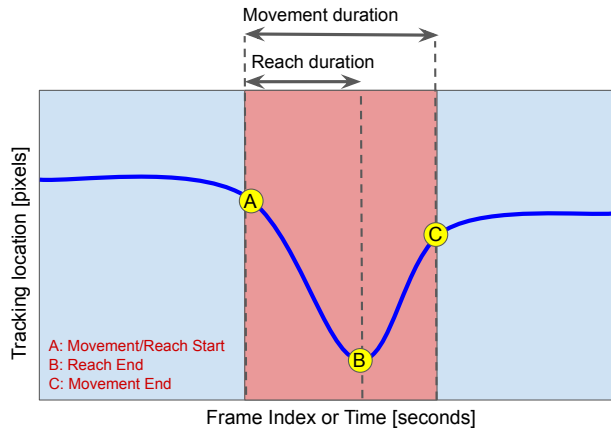
Figure 4: Event metadata extracted for a movement event: Movement duration is the duration of consecutive *move* states (i.e. from 'A' to 'C'). A *reach* is the subset of the movement event from the start of the movement (label 'A') to the point of maximum deviation (label 'B').

of prolonged *rest* states across both wrists and the nose.

Event Metadata Extraction

For each detected event, we extract several metadata (Figure 4) that could be used for analytics (Figure 5) or for filtering events before further analysis. Metadata extracted include (x, y) pixel coordinates of a keypoint at the start and end of an event, duration of an event, rest duration before and after a movement event, and average pose-estimation confidence (for filtering out poorly tracked events). When processing wrist keypoints, we define a *reach* to be the wrist's maximum-displacement during the event compared to its location at the start of the event. In addition, we extract magnitude, angle, and duration of the reach associated with each movement event.

Towards Naturalistic Neuroscience

We evaluate the use of these naturalistic wrist movement-initiation events in two ways. First, we examine their neural correlates and, second, we use them to train neural decoders to detect future events.

Neural Correlates

We compute baseline-subtracted spectrograms for individual wrist movement-initiation events and average them to obtain event-averaged spectrograms (Figure 6). The observed activation pattern shows movement associated increases in a high-frequency band (HFB, $76 - 100$ Hz) and decreases in a low-frequency band (LFB, $8 - 32$ Hz) in motor-cortex electrodes, corroborating previously reported findings (Miller et al., 2007).
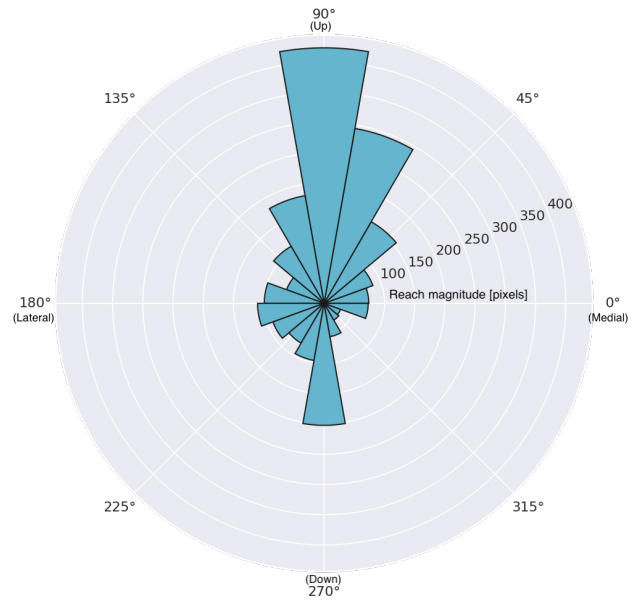

Figure 5: Polar histogram plot of reaches: Though dominated by movements along the upward-downward axis, reaches are found to be directed towards a broad distribution of angles.

Neural Decoding

We evaluate the use of our pipeline as a source of training data for a neural decoder classifier that detects wrist movement-initiation. In other words, it discriminates movement-initiation from no-movement events. We train a Random Forest classifier using spectral power features (HFB and LFB) extracted from the ECoG signal around event times
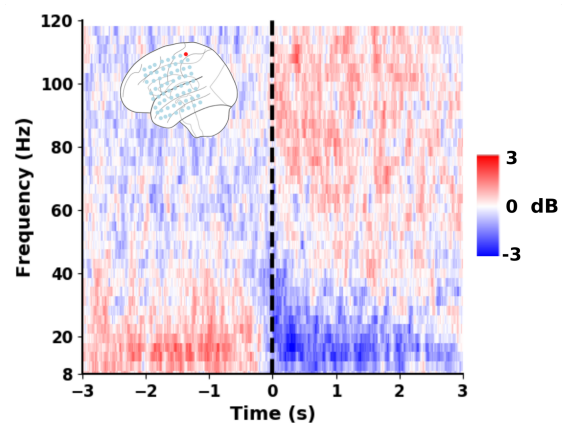

Figure 6: Spectrogram shows movement-associated activation reduction at low-frequencies and activation increase at high-frequencies, corroborating previously reported findings (Miller et al., 2007). [Inset] Electrode grid position (light blue) shown on left side view of patient's brain. The red electrode's spectrogram is shown here.

(400 ms before to 600 ms after each event). The classifier is trained on on events from days 3 & 4 of patient stay, and tested on events from day 5. Table 1 summarizes a preliminary assessment of performance on this task. The classifier performs well above chance, and its performance is comparable to previously reported decoders (N. X. Wang et al., 2018).

Table 1: Classification performance on test data

Event	Precision	Recall	F1 Score	Support
rest-to-move	0.71	0.60	0.65	205
no-movement	0.65	0.75	0.70	205

Discussion

In summary, we describe a workflow for analyzing simultaneously recorded long-term naturalistic human brain and behavioral video data that does not involve any *a priori* experimental design. The events and their metadata derived using this pipeline are amenable to a variety of scientific analyses; here we have examined their neural correlates, and have used them to train neural decoders.

There are several extensions of our work that we are currently exploring. We can investigate multi-keypoint interactions at a fine-grained temporal resolution by combining single-keypoint state sequences. In particular, we are interested in the various possible temporal sequence combinations of wrist movement. (e.g. raising the left and right wrists simultaneously vs. one after the other.)

As a natural extension to predicting movement initiation, we plan to predict entire movement trajectories. Additionally, our pipeline has the capability of generating an order-of-magnitude more events compared to the traditional controlled laboratory experiment paradigm (Schalk & Leuthardt, 2011). This substantial expansion in training data could improve the accuracy of brain-computer interface decoders and make them more robust to variability when deployed in real-world scenarios.

Acknowledgements

References

Alasfour, A., Gabriel, P., Jiang, X., Shamie, I., Melloni, L., Thesen, T., ... others (2019). Coarse behavioral context decoding. *Journal of Neural Engineering*, *16*(1), 016021.

Anderson, D. J., & Perona, P. (2014, October). Toward a Science of Computational Ethology. *Neuron*, *84*(1), 18-31. doi: 10.1016/j.neuron.2014.09.005

Berman, G. J. (2018, December). Measuring behavior across scales. *BMC Biology*, *16*(1). doi: 10.1186/s12915-018-0494-7

Chen, K., Gabriel, P., Alasfour, A., Gong, C., Doyle, W. K., Devinsky, O., ... Gilja, V. (2018). Patient-Specific Pose Estimation in Clinical Environments. *IEEE Journal of Translational Engineering in Health and Medicine*, *6*, 1-11. doi: 10.1109/JTEHM.2018.2875464

Huk, A., Bonnen, K., & He, B. J. (2018, July). Beyond Trial-Based Paradigms: Continuous Behavior, Ongoing Neural Activity, and Natural Stimuli. *The Journal of Neuroscience*, 1920-17. doi: 10.1523/JNEUROSCI.1920-17.2018

Markowitz, J. E., Gillis, W. F., Beron, C. C., Neufeld, S. Q., Robertson, K., Bhagat, N. D., ... Datta, S. R. (2018, June). The Striatum Organizes 3D Behavior via Moment-to-Moment Action Selection. *Cell*, *174*(1), 44-58.e17. doi: 10.1016/j.cell.2018.04.019

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). *DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning* (Tech. Rep.). Nature Publishing Group.

Miller, K. J., Leuthardt, E. C., Schalk, G., Rao, R. P. N., Anderson, N. R., Moran, D. W., ... Ojemann, J. G. (2007, February). Spectral Changes in Cortical Surface Potentials during Motor Movement. *Journal of Neuroscience*, *27*(9), 2424-2432. doi: 10.1523/JNEUROSCI.3886-06.2007

Pereira, T. D., Aldarondo, D. E., Willmore, L., Kislin, M., Wang, S. S.-H., Murthy, M., & Shaevitz, J. W. (2019). Fast animal pose estimation using deep neural networks. *Nature methods*, *16*(1), 117.

Ramasamy Ramamurthy, S., & Roy, N. (2018). Recent trends in machine learning for human activity recognition - A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1254.

Schalk, G., & Leuthardt, E. C. (2011). Brain-computer interfaces using electrocorticographic signals. *IEEE Reviews in Biomedical Engineering*, *4*, 140–154.

Todd, J. G., Kain, J. S., & de Bivort, B. L. (2017, February). Systematic Exploration of Unsupervised Methods for Mapping Behavior. *Physical Biology*, *14*(1), 015002. doi: 10.1088/1478-3975/14/1/015002

Wang, N. X., Farhadi, A., Rao, R. P., & Brunton, B. W. (2018). AJILE movement prediction: Multimodal deep learning for natural human neural recordings and video. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Wang, N. X. R., Olson, J. D., Ojemann, J. G., Rao, R. P. N., & Brunton, B. W. (2016, April). Unsupervised Decoding of Long-Term, Naturalistic Human Neural Recordings with Automated Video and Audio Annotations. *Frontiers in Human Neuroscience*, *10*. doi: 10.3389/fnhum.2016.00165

Wiltschko, A. B., Johnson, M. J., Iurilli, G., Peterson, R. E., Katon, J. M., Pashkovski, S. L., ... Datta, S. R. (2015, December). Mapping Sub-Second Structure in Mouse Behavior. *Neuron*, *88*(6), 1121-1135. doi: 10.1016/j.neuron.2015.11.031