Non-Parametric Analysis of Inter-Individual Relations Using an Attention-Based Neural Network

Takashi Morita*[1], Aru Toyoda[2], Seitaro Aisu[1], Akihisa Kaneko[1], Naoko Suda-Hashimoto[1], Ikki Matsuda[2,3,4,5], and Hiroki Koda[1]

[1]Primate Research Institute, Kyoto University
[2]Chubu University Academy of Emerging Sciences
[3]Wildlife Research Center of Kyoto University
[4]Japan Monkey Centre
[5]Institute for Tropical Biology and Conservation, Universiti Malaysia Sabah

March 26, 2020

Abstract

In the past decade, social network analysis was widely adopted in animal studies, and it enabled the revelation of global characteristic patterns of animal social systems from pairwise inter-individual relations. Animal social networks are typically drawn based on geological proximity and/or frequency of social behaviors (e.g., grooming), but the appropriate choice of the distance/closeness metric is not clear especially when prior knowledge about the species/data is limited. In this study, the researchers explored a non-parametric analysis of inter-individual relations using a neural network with the attention mechanism, which plays a central role in natural language processing. The high interpretability of the attention mechanism and the flexibility of the entire neural network allow automatic detection of inter-individual relations included in raw data. As case studies, three-dimensional location data collected from simulated agents and real Japanese macaques were analyzed.

Keywords: interaction, social network analysis, animal society, Japanese macaque, deep learning, attention mechanism

Significance Statement

- Novel non-parametric analysis of inter-individual relations is proposed.
- The proposed method is based on an artificial neural network and automatically detects latent inter-individual dependencies behind raw data, without prior assumptions about the possible types of relations or random distributions involved.
- The performance of the method was tested on simulated and real three-dimensional location data of Japanese macaques.
- The analysis will be effective especially when little is yet known about the species under study.

*Corresponding author: tmorita@alum.mit.edu

1 Introduction

Understanding the characteristics of animal social groups is a major objective of animal ecology. Complex social systems of eusocial insects and non-human primates are often viewed as biological models or precursors of human societies, and as such, have also attracted the interest of social scientists. In the past decade, social network analysis was widely adopted in animal studies, and it enabled us to reveal global characteristic patterns of animal social systems from pairwise inter-individual relations (Krause et al., 2015).

The application of social network analysis to animal studies requires an appropriate metric for the relational distance/closeness between individuals. Popular options of the metric include geological proximity (often thresholded and converted to counts of co-appearance within the threshold; Silk et al., 2006b,a; Croft et al., 2008; Clark, 2011; Boyland et al., 2013; Castles et al., 2014; Krause et al., 2015; Schofield et al., 2019) and frequency of social behaviors (e.g., grooming of non-human primates; Chepko-Sade et al., 1989; Silk et al., 2006b,a; Croft et al., 2008; Clark, 2011; King et al., 2011b,a; Castles et al., 2014; Krause et al., 2015). However, the appropriate choice of the metric is not necessarily clear, especially when prior knowledge about the species/data is limited (cf. Castles et al., 2014; Farine and Whitehead, 2015). For example, location information for measuring the proximity is available from a wide variety of animals. This includes even those species whose social activities are not yet well-known. Location information also became more reliable and scalable thanks to the recent developments in bio-logging technology (Heupel et al., 2006; Rutz and Hays, 2009; Rutz et al., 2012; Fehlmann and King, 2016; King et al., 2018; Dore et al., 2020). However, naively measuring the Euclidean distance between individuals is not necessarily reasonable because the vertical and horizontal divergences would not have the same importance in social network analysis; Many terrestrial animals do not fly or dive, and cannot control the vertical distance from other individuals. Even the horizontal distance among animal agents could have different social meanings if other factors, such as trees and other objects, intervene. Hence, it is ideal to have a non-parametric, end-to-end analyzer of inter-individual relations that runs on raw data.

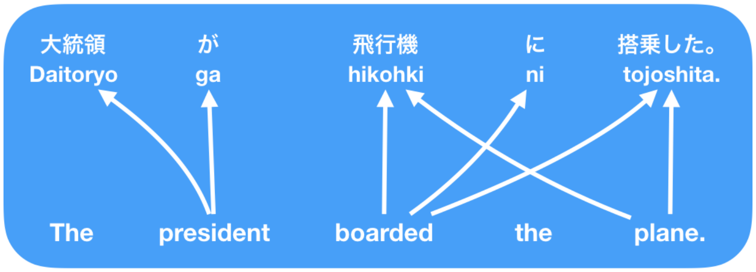
The present study explores non-parametric analysis of inter-individual relations using a neural network. Specifically, the researchers exploit an interpretable module called *attention mechanism*, as well as the flexibility of the entire neural network that fits to different types of data. The attention mechanism was originally invented for machine translation to represent relations between words in the source and target languages (Figure 1a; Bahdanau et al., 2015). More recently, studies begin to use attention mechanism to represent word relations within input sentences (self-attention, Figure 1b; Vaswani et al., 2017; Devlin et al., 2018). It is now considered central technology in today's natural language processing (NLP) in general. The present study applies this technology to three-dimensional (3D) location data collected from Japanese macaques (*Macaca fuscata*) to reveal their latent relations reflected in the data (Figure 1c), without prior assumptions about the possible types of relation structures or noise distributions. The said approach was validated by analyzing simulated data and recovering the latent relations among the virtual agents generating the data. Even though only the location data are discussed in this study, the possible applications of our relation analysis are not limited to the location data because of the flexibility of the neural network.

Our model organism, the Japanese macaque, exhibits rich affiliative and aggressive social behaviors and has been studied extensively for over more than 50 years for better understanding of the evolution of social systems in primate lineages, including humans (Nakagawa et al., 2010). This accumulated knowledge of the species is helpful to interpret/evaluate the results of our novel analysis even though the method is not limited to the species or other primates and could even be more valuable for further studies about animals that do not show intrepretable social activities.

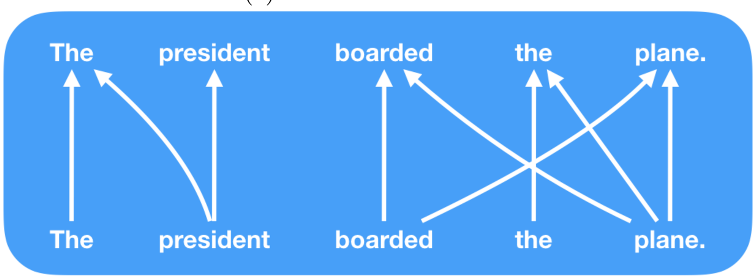
2 Materials & Methods

2.1 Overview of the Neural Network

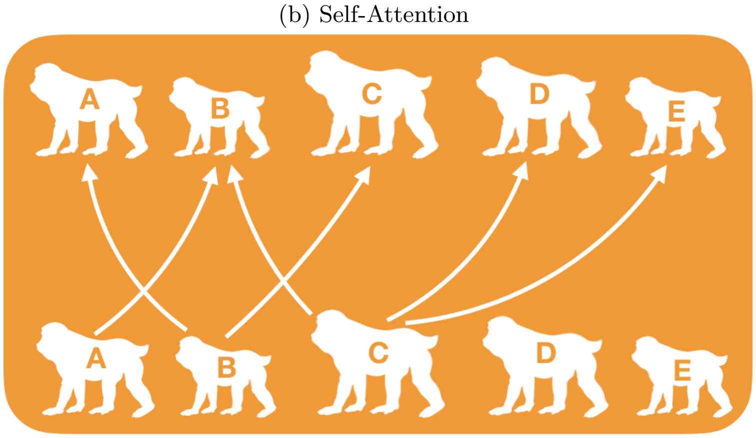
The framework of the relation analysis in this study is to (i) train a neural network that predicts the 3D location of each individual macaque—simulated or real—from that of the other individuals, and then (ii) check how much attention the network pays to each of the referent individuals when it makes the predictions. For simplicity, the network only receives the location data at a single timepoint and does not analyze temporal



(a) Machine Translation



(b) Self-Attention



(c) Inter-Individual Relations

Figure 1: History of the use of attention mechanism. (a) The original use of attention mechanism in machine translation, relating words in the source language (English) to ones in the target language (Japanese). (b) More recent application in NLP that represents relations between words within the input sentences. (c) Application to inter-individual relations explored in this study.

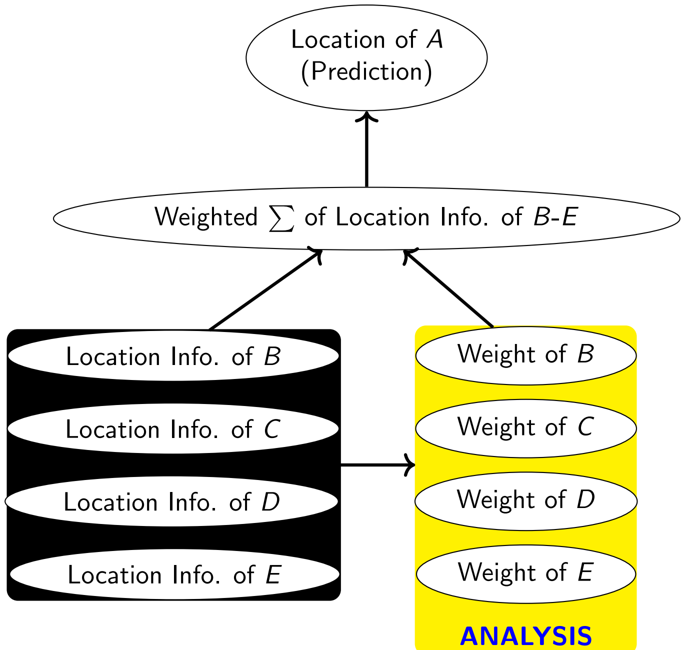


Figure 2: Schematic diagram of attention-based analysis.

changes in the location data, although the temporal information could include important traits of individuals and groups (cf. Morita et al., 2020) and thus should be discussed in future studies.

Figure 2 sketches the network structure, abstracting away from fine details (see the supporting information S1 for the full structure of the network). The location of an individual named A is predicted from that of four other individuals, B, C, D, and E. The prediction is made from the weighted sum of the (transformed) location information of the four referents, where the weight of each referent—called *attention weight*—is also derived from the location information. The attention weights govern the information flow from the referents; they intuitively represent *who is contributing how much to the prediction*. Our analysis of the relations between the target and the referents is performed on these attention weights.

The attention weights were normalized and always have a sum of 1. However, this formalization has one disadvantage. The model never detects the independence of the target individual, as at least one referent always has a non-zero weight. Accordingly, the predictive performance of the attention model was compared with a baseline model that predicts the target without reference to the other individuals. If the attention weights are redundant, the attention model will have the same level of predictive performance as the no-referent baseline. The baseline model was trained to maximize the likelihood.[1]

2.2 Data

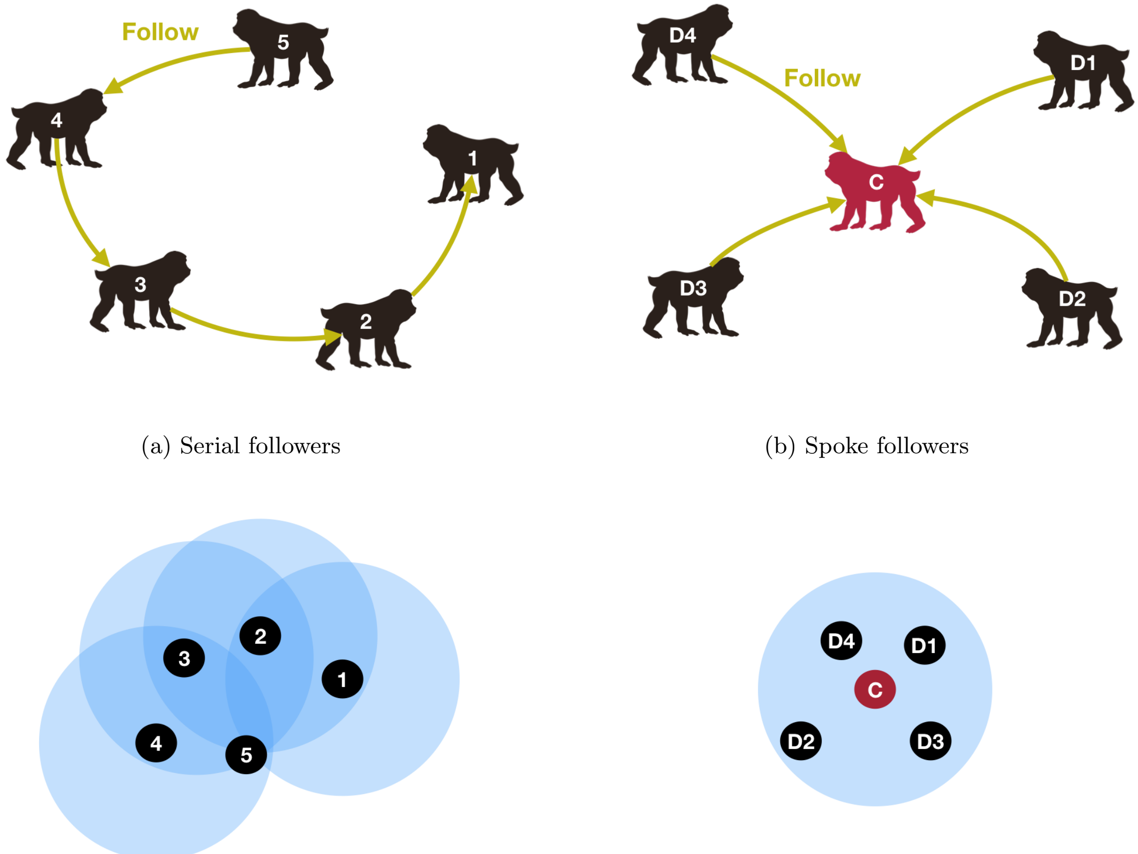
Attention-based relation analysis was performed on two simulated data (§2.2.1) and one real data (§2.2.2).

2.2.1 Simulated Data

The researchers tested the reliability of the attention-based relation analysis on two types of simulated data. Following the real data collected from Japanese macaques (see §2.2.2), both simulations had five individuals located in a bounded 3D space ($[0.0, 5.0] \times [0.0, 4.0] \times [0.0, 2.5]$), resembling the group cage where the macaques were kept.

The first simulation modeled the situation where the five agents moved in a sequence, each following the other (Figure 3a). This pattern was simulated by sampling the 3D location of each individual from

[1] As reported in §2.2, the location of the target individual was discretized, which often improves the performance of neural networks (van den Oord et al., 2016a,b). Hence, the maximum likelihood training of the baseline model was simple. The occurrences of each discretized location value were counted and then the gathered data were normalized.



(a) Serial followers

(b) Spoke followers

(c) Serial followers

(d) Spoke followers

Figure 3: Two patterns of simulated data. (a) and (c) model animal agents that move in a sequence. (b) and (d) model the situation where there is a single hub individual (red) that is followed by the other four spoke dependents (black). (a) and (b) depict the relations among the animal agents modeled in the simulations. (c) and (d) describe the distributions from which the location data are sampled: The blue circles represent the support of the uniform distribution centered at each following individual.

the neighborhood of its predecessor: (i) the first individual was located uniformly randomly in the domain, and (ii) the $(i + 1)$-th individual was located uniformly randomly in the intersection of the domain and the sphere with 1.5 radius centered at the i-th individual's location, which was implemented by rejection sampling (Figure 3c).

The second simulation modeled the hub and spoke relation in which a single hub individual was followed by the other four spoke dependents (Figure 3b). The researchers simulated this situation by (i) first sampling the location of the hub individual uniformly from the domain, and (ii) sampling the location of the spoke individuals uniformly from the intersection of the domain and the sphere with 1.5 radius centered at the hub individual's location.

For each of the two simulations, 100,000 samples were generated—each containing 3D coordinates of the five individuals.[2] All the samples were used for the training and the attention analysis afterward. The researchers did not hold out a test portion of the data, as the study's purpose was not to test the general applicability of the model predictions, but rather to reveal the dependency relations in the data (just as in

[2]The number of different target-referents pairs observed by the neural network was $5 \times 100,000 = 500,000$, since all the possible partitions of the five individuals were treated in the analysis as one target and four referents.

the linear regression tests).

Note that in both cases, the location samples were the only information available to the neural network analyzer. Furthermore, the settings of the simulations, including the type of the inter-individual relations and the form of the conditional distributions, were not hard-coded in the analyzer, unlike other classical analyses assuming a limited hypothesis space including the correct generative model (e.g., Bayesian inference). For example, when analyzing the data from the second simulation, the network had to (i) discover the hub structure among many other possible relation patterns behind the data, and (ii) identify the hub individual while all the five agents were always close to one another.

Although the simulated data took on continuous values (residing in the bounded subspace of the 3D Euclidean), it has been reported that neural networks are often better at predictions of discrete categories (due to the flexibility of categorical distributions; van den Oord et al., 2016a,b, similar discrete predictions of location data were also adopted by Takeda and Komatani, 2016). Accordingly, the domain was split into 400 cubes ($0.5 \times 0.5 \times 0.5$), and the network was trained to make a discrete prediction of the cube containing the target's location.

2.2.2 Real Data

In addition to the simulated data described in §2.2.1, the researchers analyzed real 3D location data of five captive Japanese macaques in the Primate Research Institute, Kyoto University (KUPRI), Japan. The subjects were two adult males and three adult females. The location recordings of the five macaques were conducted for two weeks in an outdoor group cage ($4 \times 5 \times 2.5$ m), where the subjects were allowed to freely move and interact with one another. Each subject carried two Bluetooth® Low-Energy beacons, and their 3D location was estimated based on the signals coming from these beacons (Quuppa Intelligent Locating System™). See Morita et al. (2020) for more details on the data collection procedure.

Although the system sampling frequency was set at 9 Hz, the actual rate was unstable because of uncontrollable signal interference and reflection perturbation. The sampling from the $2 \times 5 = 10$ beacons did not also synchronize their timing. Accordingly, the researchers took the median value of the samples from each individual (i.e., from the two beacons) for each dimension collected in every 3000 msec interval. The medians over the same interval constituted the input-output pair for the neural network.[3] This sacrifices time resolution of the collected data; however, this was not fatal to this study as the temporal information was ignored in the analysis. The preprocessing above yielded a total of 327,592 sets of data.[4] Just as in the simulations, all the data was used for both the model training and attention analysis without holding out a test portion, and the target locations were discretized into 400 cubes ($0.5 \times 0.5 \times 0.5$ m).

3 Results

Figure 4 reports the average attention weights among five individuals in two representations. The heatmaps on the left-hand side represent the attention weight assigned to each referent individual (column) upon the prediction of the target individual (row) by the color density of the corresponding cell. Moreover, the directed graphs on the right-hand side represent the same information by the color density of the arrows coming from the referents to the targets.

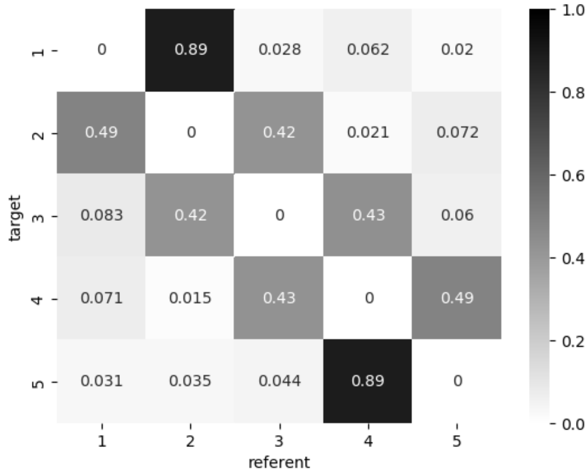
The attention-based analysis correctly distinguished the two simulation patterns. Given the data from the serial follower simulation, the analyzer correctly assigned heavy weights (> 0.1) to the relations between the individuals with successive indices (e.g., "2" and "3"; Figure 4a,b). On another note, the hub-spoke simulation concentrated the attention weights to the hub individual when predicting the spoke dependents, and evenly distributed the weights to the spokes when predicting the hub (Figure 4c,d).

The analysis of the real data revealed that the female macaques were more informative referents than the males, especially when predicting the other females. The attention weight to Male 1 as a referent was smaller compared with all the females, except when Male 2 was predicted (where Female 1 collected zero

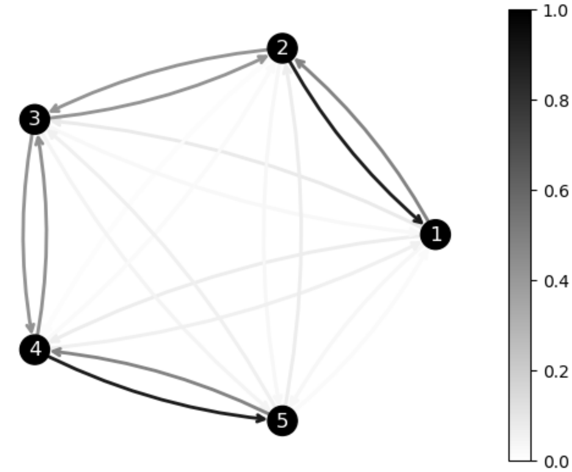
[3] The network simultaneously analyzed all the possible choices of one individual as the prediction target and four as the referents.

[4] Again, the researchers analyzed all the possible partitions of the five individuals into one target and four referents, and thus, the number of different target-referents pairs observed by the neural network was $5 \times 327,592 = 1,637,960$.

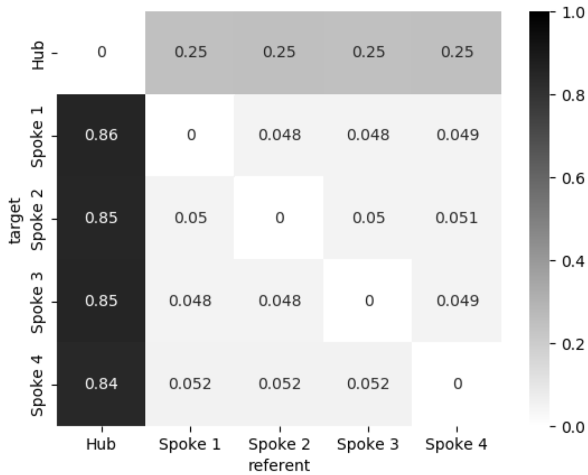
bioRxiv preprint doi: https://doi.org/10.1101/2020.03.25.994764; this version posted March 26, 2020.



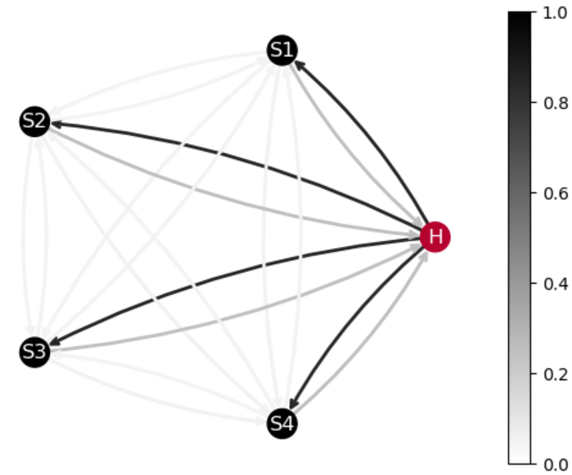
(a) Serial followers



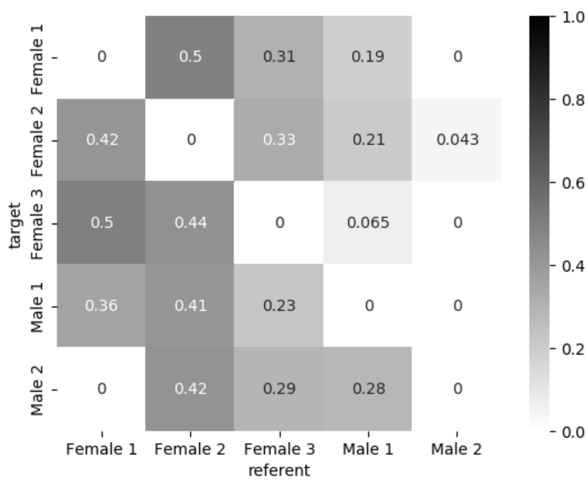
(b) Serial followers



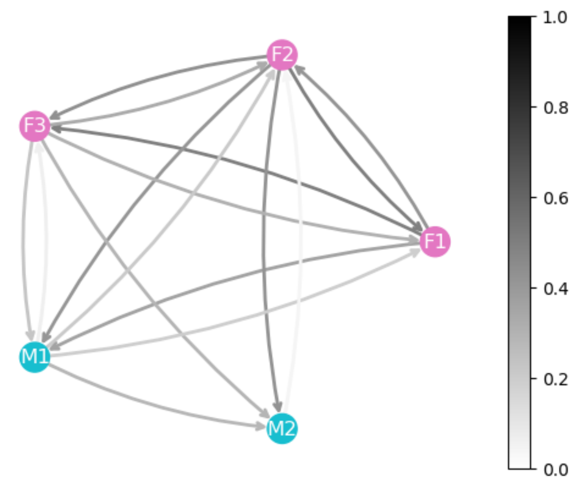
(c) Spoke followers



(d) Spoke followers



(e) Real data



(f) Real data

Figure 4: Attention weights. The same data are reported in heatmaps on the left-hand side and in directed graphs on the right-hand side. Top (a,b): results from the simulation with serial followers (Figure 3a,c). Middle (c,d): results from the simulation with a single hub individual and four spoke dependents (Figure 3b,d). Bottom (e,f): results from the real location data collected from two male and three female Japanese macaques (§2.2.2).

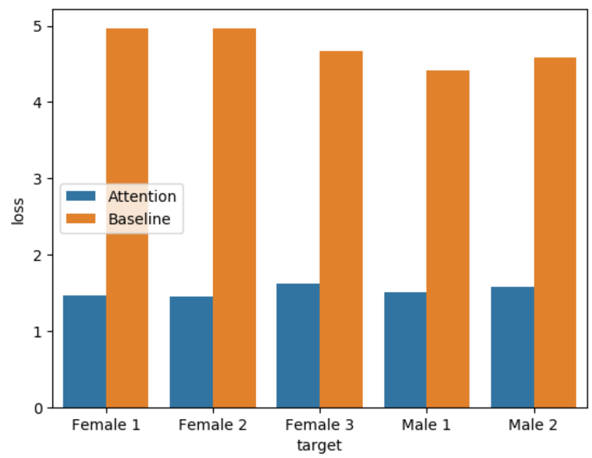


Figure 5: Average prediction loss (negative log probability) of the attention-based neural network and the no-referent baseline.

attention),[5] and Male 2 was almost never used as a referent (Figure 4e,f). It is of note, however, that the males did not move independently of the other individuals. The predictions of the males' location were based on three other individuals (Male 1 was dependent on the females, and Male 2 was dependent on Female 2, 3, and Male 1). While the neural network was unable to flag independence as the attention weights were never all zeros, these detected dependencies were not redundant because the network outperformed the no-referent baseline. As shown in Figure 5, the average prediction loss (i.e., negative log probability) of the attention model was smaller compared with the baseline across all the individuals.

4 Discussion

In the study, the attention analysis correctly identified the latent relations behind the two types of simulated data, one having serial followers and the other having the spoke followers. While these simulations were simple, it is of note that the analyses were performed without prior assumptions about the possible types of inter-individual relations or random distributions involved. Compared with the traditional distance/closeness-based analyses (Silk et al., 2006b,a; Croft et al., 2008; Clark, 2011; Boyland et al., 2013; Castles et al., 2014; Krause et al., 2015; Schofield et al., 2019), this method would be more effective in visualizing latent dependency patterns, especially when little is yet known about the properties of data.

The study also demonstrated that the attention analysis is applicable to raw real data, reporting a case study of Japanese macaques. The researchers found that the location of females is more informative than that of males in locating the other female individuals. This result is consistent with the female-oriented social system of the macaques: Japanese macaques are a female-philopatrics species (represented by parturition in the natal group and male emigration), and females usually build firmer and more stable social relationships with one another than males (e.g., matrilineal social rank succession; Yamada, 1963; Furuichi, 1984; Mitani, 1986; Nakagawa et al., 2010). The females of the species are also reported to exhibit more tight and frequent spatial cohesion (Otani et al., 2014). Therefore, knowing the location of some females is expected to increase the chance of locating other females in the same group. In addition, the researchers' earlier study on the same data showed that the males' movements (i.e. trajectories of the 3D location) reflect a greater degree of individuality than that of the females (Morita et al., 2020). Because of such idiosyncratic movements, the males' location may not be informative predictors of other individuals' location. The analysis of the real data also revealed that one individual (Male 2) was almost never attended by the others. The discovery of such distinguished individuals is useful to build new hypotheses about the animal society and plan behavioral experiments targeting them.

[5] The average attention to Male 1, upon the predictions of Male 2, looked equal to the attention to Female 3, but it actually was statistically significantly smaller, considering the 10,000 samples of bootstrapping.

Attention analysis does not exploit any behavioral patterns that are particular to Japanese macaques, and researchers can use it for location data of other animals, including ones that do not exhibit prototypical social behaviors such as grooming. It can also be used if the conventional metric of distance/closeness is unavailable. The researchers further emphasize that attention analysis is not limited to location data because the flexibility of the neural network would allow its application to a wide variety of data with interacting components. For example, vocal communication has been of major interest to behavioral biologists as well as evolutionary linguists (Levinson, 2016). Attention analysis could potentially reveal latent communication flow when applied to the vocalization recordings of multiple individuals. It is to be noted, however, that the proposed method needs to be extended and refined for such broader applications and for more detailed analyses of location/movement. For example, the current version does not process temporal changes in location, whereas the implementation of temporal analysis is crucial in analyzing dynamic behaviors and vocal communication. Since the attention mechanism was developed for NLP, however, processing time series data should not be so hard (see Vaswani et al., 2017; Shaw et al., 2018; Dai et al., 2019, for popular ways to encode time information).

Acknowledgements

We appreciate the animal care by our technicians and research assistants as well as their suggestions and support for this project. In particular, we wish to express special thanks to Norihiko Maeda, Mayumi Morimoto, and Takayoshi Natsume for animal arrangements, and Panasonic Solution Technologies Co., Ltd. for the apparatus technical supports. We also acknowledge the profound comments and suggestions from Zin Arai and Hiroshi Takeuchi regarding the analysis. This study was performed under the Cooperative Research Program at KUPRI (2018-C-27, 2019-B-27).

Funding

This study was mainly funded by the Japan Science and Technology Agency, Core Research for Evolutional Science and Technology 17941861 (#JPMJCR17A4) and partly by MEXT Grant-in-aid for Scientific Research on Innovative Areas #4903 (Evolinguistic), 17H06380.

Ethical Statement

All the procedures were reviewed and approved by the Animal Welfare and Care Committee of KUPRI (Permission # 2018-203) and complied with the institutional guidelines (Primate Research Institute, 2010).

Author contributions

Project organization: IM HK; Animal arrangements: HK SA NSH; Apparatus building: AT IM HK; Data acquisition: AT IM HK; Animal cares: AK AT IM HK; Data management; HK AT TM; Computational modeling: TM; Manuscript writing: IM HK TM.

References

Bahdanau, D., Cho, K., and Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings.*

Boyland, N. K., James, R., Mlynski, D. T., Madden, J. R., and Croft, D. P. (2013). Spatial proximity loggers for recording animal social networks: consequences of inter-logger variation in performance. *Behavioral Ecology and Sociobiology*, 67(11):1877–1890.

Castles, M., Heinsohn, R., Marshall, H. H., Lee, A. E., Cowlishaw, G., and Carter, A. J. (2014). Social networks created with different techniques are not comparable. *Animal Behaviour*, 96:59–67.

Chepko-Sade, B., Reitz, K. P., and Sade, D. S. (1989). Sociometrics of macaca mulatta iv: Network analysis of social structure of a pre-fission group. *Social Networks*, 11(3):293–314. Special Issue on Non-Human Primate Networks.

Clark, F. E. (2011). Space to choose: network analysis of social preferences in a captive chimpanzee community, and implications for management. *American Journal of Primatology*, 73(8):748–757.

Croft, D. P., James, R., and Krause, J. (2008). *Exploring Animal Social Networks*. Princeton University Press.

Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q. V., and Salakhutdinov, R. (2019). Transformer-XL: Attentive language models beyond a fixed-length context.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805.

Dore, K. M., Hansen, M. F., Klegarth, A. R., Fichtel, C., Koch, F., Springer, A., Kappeler, P., Parga, J. A., Humle, T., Colin, C., Raballand, E., Huang, Z.-P., Qi, X.-G., Di Fiore, A., Link, A., Stevenson, P. R., Stark, D. J., Tan, N., Gallagher, C. A., Anderson, C. J., Campbell, C. J., Kenyon, M., Pebsworth, P., Sprague, D., Jones-Engel, L., and Fuentes, A. (2020). Review of GPS collar deployments and performance on nonhuman primates. *Primates*.

Farine, D. R. and Whitehead, H. (2015). Constructing, conducting and interpreting animal social network analysis. *Journal of Animal Ecology*, 84(5):1144–1163.

Fehlmann, G. and King, A. J. (2016). Bio-logging. *Current Biology*, 26(18):R830–R831.

Furuichi, T. (1984). Symmetrical patterns in non-agonistic social interactions found in unprovisioned Japanese macaques. *Journal of Ethology*, 2(2):109–119.

Heupel, M. R., Semmens, J. M., and Hobday, A. J. (2006). Automated acoustic tracking of aquatic animals: scales, design and deployment of listening station arrays. *Marine and Freshwater Research*, 57(1):1–13.

King, A. J., Clark, F. E., and Cowlishaw, G. (2011a). The dining etiquette of desert baboons: the roles of social bonds, kinship, and dominance in co-feeding networks. *American Journal of Primatology*, 73(8):768–774.

King, A. J., Fehlmann, G., Biro, D., Ward, A. J., and Fürtbauer, I. (2018). Re-wilding collective behaviour: An ecological perspective. *Trends in Ecology & Evolution*, 33(5):347–357.

King, A. J., Sueur, C., Huchard, E., and Cowlishaw, G. (2011b). A rule-of-thumb based on social affiliation explains collective movements in desert baboons. *Animal Behaviour*, 82(6):1337–1345.

Krause, J., James, R., Franks, D. W., and Croft, D. P., editors (2015). *Animal Social Networks*. Oxford University Press.

Levinson, S. C. (2016). Turn-taking in human communication – origins and implications for language processing. *Trends in Cognitive Sciences*, 20(1):6–14.

Mitani, M. (1986). Voiceprint identification and its application to sociological studies of wild Japanese monkeys (*Macaca fuscata yakui*). *Primates*, 27(4):397–412.

Morita, T., Toyoda, A., Aisu, S., Kaneko, A., Suda-Hashimoto, N., Matsuda, I., and Koda, H. (2020). Animals exhibit consistent individual differences in their movement: A case study on location trajectories of Japanese macaques. *Ecological Informatics*, 56:101057.

Nakagawa, N., Nakamichi, M., and Sugiura, H. (2010). *The Japanese Macaques*. Primatology Monographs. Springer Japan.

Otani, Y., Sawada, A., and Hanya, G. (2014). Short-term separation from groups by male Japanese macaques: Costs and benefits in feeding behavior and social interaction. *American Journal of Primatology*, 76(4):374–384.

Primate Research Institute (2010). *The Care and Use of Laboratory Primates*. The Primate Research Institute, Kyoto University, third edition.

Rutz, C., Burns, Z. T., James, R., Ismar, S. M. H., Burt, J., Otis, B., Bowen, J., and St Clair, J. J. H. (2012). Automated mapping of social networks in wild birds. *Current Biology*, 22(17):R669–R671.

Rutz, C. and Hays, G. C. (2009). New frontiers in biologging science. *Biology Letters*, 5(3):289–292.

Schofield, D., Nagrani, A., Zisserman, A., Hayashi, M., Matsuzawa, T., Biro, D., and Carvalho, S. (2019). Chimpanzee face recognition from videos in the wild using deep learning. *Science Advances*, 5(9).

Shaw, P., Uszkoreit, J., and Vaswani, A. (2018). Self-attention with relative position representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 464–468, New Orleans, Louisiana. Association for Computational Linguistics.

Silk, J., Alberts, S., and Altmann, J. (2006a). Social relationships among adult female baboons (*Papio cynocephalus*) II. variation in the quality and stability of social bonds. *Behavioral Ecology and Sociobiology*, 61:197–204.

Silk, J., Altmann, J., and Alberts, S. (2006b). Social relationships among adult female baboons (*Papio cynocephalus*) I. variation in the strength of social bonds. *Behavioral Ecology and Sociobiology*, 61:183–195.

Takeda, R. and Komatani, K. (2016). Sound source localization based on deep neural networks with directional activate function exploiting phase information. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 405–409.

van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., and Kavukcuoglu, K. (2016a). Wavenet: A generative model for raw audio.

van den Oord, A., Kalchbrenner, N., Espeholt, L., kavukcuoglu, k., Vinyals, O., and Graves, A. (2016b). Conditional image generation with PixelCNN decoders. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems 29*, pages 4790–4798. Curran Associates, Inc.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc.

Yamada, M. (1963). A study of blood-relationship in the natural society of the Japanese macaque. *Primates*, 4(3):43–65.