

Matrix factorization approach to behavioral mode analysis from acceleration data

Yehezkel S. Resheff^{*†‡}, Shay Rotics[†], Ran Nathan[†], Daphna Weinshall[‡]

^{*}Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, 91904, Israel.

Email: Yehezkel.Resheff@mail.huji.ac.il

[†]Movement Ecology Lab, Department of Ecology, Evolution and Behavior, The Hebrew University of Jerusalem

[‡]School of Computer Science and Engineering, Hebrew University of Jerusalem, Israel

Abstract—The field of Movement Ecology is experiencing a period of rapid growth in availability of data, and like many other fields is turning to data science for tools and methods to cope with the new challenges and opportunities that this presents. One rich and interesting source of data is the bio-logger. These small electronic devices are attached to animals free to roam in their natural habitats, and report back readings from multiple sensors, including GPS and accelerometer bursts. A common use of this accelerometer data is for supervised learning of behavioral modes. However, there is a need for unsupervised analysis tools as well, due to the inherent difficulties of obtaining a labeled dataset, which in some cases is either infeasible or does not successfully encompass the full repertoire of behavioral modes of interest. Here we present a matrix factorization based clustering method that allows either a soft or a hard partitioning of acceleration measurements, as well as a straight-forward way of drawing insight into the complex movements themselves. The method is validated by comparing the partitions with a labeled dataset, and is further compared to standard methods highlighting the advantages of the new method.

I. INTRODUCTION

The aim of the field of Movement ecology is to unify research of movement of organisms and aid in the development of a general theory of whole-organism movement [1]. Recent technological advances in tracking tools and especially the appearance of cheap and small GPS devices [2], have driven the field into a period of rapid growth in knowledge and insight [3], and have led to the emergence of various methods of analyzing movement patterns [4]. These advances have motivated the development of integrative conceptual frameworks unifying cognitive, biomechanical, randomness and optimality paradigms to study movements of all kinds by all types of organisms [1]. Nevertheless, movement data, however accurate, is unlikely to suffice for inference on the links between behavioral, ecological, physiological, and evolutionary processes driving the movement of individuals, and link these subjects which have traditionally been researched separately in their respective fields. Thus, promoting movement ecology research and the desirable unification across species and movement phenomena requires the development of additional data sources: sensors and tools providing simultaneous information

about the movement, energy expenditure and behavior of the focal organisms, together with the environmental conditions they encounter en route [5].

One such tool, which has been introduced into the field of movement ecology, is the accelerometer-biologger (ACC). These sensors allow the determination of the acceleration of the tagged animal's body, and are used as a means of identifying moment-to-moment behavioral modes [6], and estimating energy expenditure [7]. ACC loggers typically record in 1-3 dimensions, either continuously or in short bouts in a constant window [8]. Their output is used to infer behavior most commonly through supervised machine learning techniques, and energy expenditure using the Overall Dynamic Body Acceleration (ODBA) or related metrics [9], [7]. Combined with GPS recordings, acceleration sensors add fine scale information on the variation in animal's behavior and energy expenditure in space and time (see [10] for a recent review). ACC-based analysis has been used to compute many measures of interest, including behavior-specific body posture, movement and activity budgets, measures of foraging effort, attempted food capture events, mortality detection, classifying behavioral modes and more [10]. These measures have facilitated movement-related research for a wide range of topics in ecology and animal behavior [4], [10], [11], [12] as well as other fields of research such as animal conservation and welfare [11], [13] and biomechanics [14], [15].

In recent years there has been considerable interest in the analysis of behavioral modes from ACC data using supervised learning techniques. The protocol for using ACC data for supervised learning of behavioral modes consists of several steps. First, a calibration procedure is performed in a controlled environment. Before deployment, the response of each tag to $\pm 1G$ acceleration on each axis is recorded, in order to fit the tag-specific linear transformation from the recorded values (mV) to the desired units of acceleration. Next, the calibrated tags are given a recording schedule and mounted on the focal animals. The data is later retrieved either by RF (radio) methods, or by physically reacquiring the device.

In order to train supervised machine learning models, a labeled dataset is collected through field observations. This time and labor intensive stage requires the researcher to

observe the animal, either in its natural habitat or in captivity, and relate the actual behavioral modes to the time-stamp of the ACC recordings. Since some behavioral modes tend to be less common, or preformed predominantly at specific times, recording a sufficient number of such behavior-measurement samples may be tricky. Furthermore, for nocturnal species, observations may not be feasible. In the final stage, models are trained using the labeled data, and the entire dataset is then labeled.

These methods have been applied to data from many species, and for a diverse range of behavioral modes. However, there are several drawbacks to the supervised approach. Observations, even if perfectly accurate, may not be adequately representative of the behavioral pattern throughout the period of the research (which is desirably the lifetime of the animal), for several reasons: field work is inherently confined to a specific time and place, only some of the animals are observed, and the presence of the observer may in some cases have an impact on the behavior of the observed animals. The very need for observations limits the scope of such research projects to observable species and thus to labs with the necessary resources (both in money and manpower) to carry out all the steps listed above.

In this paper we present a framework for unsupervised analysis of behavioral modes from ACC data. The rest of the paper is structured as follows: The next section describes related work both in Movement Ecology and in matrix factorization for clustering. In section III we introduce the features and model. Finally, section IV presents the results on a large real-world dataset and comparison to other methods.

II. PREVIOUS WORK

Previous work on behavioral mode analysis from ACC data focused predominantly on supervised learning, with an emphasis on constructing useful features and finding the right classifiers. While this line of work proved very successful, both in terms of classifier performance and of scientific discovery that it was able to drive, it still suffers from the inherent limitations of supervised learning, compounded by the very high cost of obtaining labeled data for behavioral observations of wild animals. It remains the case that for some animals (nocturnal or sea species for instance), obtaining a labeled dataset is currently infeasible. Thus, in order to use all available ACC data for behavioral mode analysis in the field of Movement Ecology, an unsupervised framework is essential.

To the best of our knowledge, there has been only a single published attempt at such a method [16] which essentially applied K-means to the ACC data. We consider K-means as a baseline to compare our method against (see Results section below).

Matrix factorization has been studied extensively in the context of clustering [17], [18] and connections have been shown to various popular clustering algorithms such as K-means and spectral clustering [19]. Our proposed method is essentially soft assignment matrix factorization clustering,

together with a theoretical justification based on the combined nature of the signals and features under consideration.

III. METHODOLOGY

A. Feature generation

The intuition behind the features of ACC signals that we use here is that a behavioral mode is distinguishable by the distribution of short-scale movements that it is comprised of. Considering these movement primitives as word analogs, we construct what is essentially a movement bag-of-words (m-bow) representation - a histogram of movement type counts. Since behavioral modes tend to have distinct time scales, different from one-another, we construct histograms relating to multiple scales in order to capture more of the relevant information in the signal. The multiple time-scale histograms are then concatenated to obtain one long feature vector per input signal.

Since the signal is a continuous measurement, we do not have direct access to the short-scale movement primitives that exist in it. Hence, we define the notion of a *patch*. A patch of width l of an ACC signal is any l consecutive measurements in it. Thus, for a signal of length T there are $T - l + 1$ distinct patches of length l . Since these patches contain continuous-values they tend to be distinct. Our aim is to discretize this space in order to bin similar patches together and form a compact representation of the count-values of similar patches.

In order to transform the raw signal into the feature vector, we first need to construct a dictionary of movement primitives. Association of patches to these primitives will then allow the binning we use as the final feature vector. The dictionary is generated for each scale separately, by applying K-Means to the set of all patches of that scale in the data. Thus, our movement primitives are the centroids of these short-scale movements that exist in the signal. Following the terminology for such features in Natural Language Processing and in Computer Vision, we call these the *patch codebook* (Algorithm 1). The resulting centroids are referred to as *patch words*. Next, for each ACC signal, the patches it contains for each scale are extracted and assigned to the histogram bin of the nearest patch word in the patch codebook. The final representation is a concatenation of such histograms across scales (Algorithm 2).

B. Mixture property of patch features

In order to motivate the proposed model (next section), we present the mixture property of patch features. We assume that our signals have the property that a large enough part of a sample from a certain behavioral mode will have distribution of patches that is the same as the distribution in the entire sample. The meaning of this assumption is that each behavioral mode has a distribution of patches that characterizes it.

Intuitively, if a signal S is constructed by taking the first half of a signal S_a and the second half of an equal length signal S_b then the distribution of patches in S will be approximately an equal parts mixture of those in S_a and in S_b . The reason for this is that a patch in S is either (a) completely contained

Algorithm 1 Multiscale codebook generation

input: $\{s_i\}_{i=1}^p$ the set of acceleration measurements l_1, \dots, l_m the scales to use k_1, \dots, k_m the size of each codebook**output:** m sets of centroids of sizes k_1, \dots, k_m and scales (dimensionality) l_1, \dots, l_m **for** $j = 1$ to m $S_i \leftarrow$ empty list**foreach** ACC signal s in $\{s_i\}_{i=1}^p$ **for** $t = 1$ to $s.Length - l_j$ add $s[t : t + l_j]$ to S_i **end****end** $C_i \leftarrow KMeans(S_i; k_j)$ **end****return** C_1, \dots, C_m

Algorithm 2 Multiscale codebook feature generation

input: C_1, \dots, C_m the output of the codebook generation algorithm (Algorithm 1) l_1, \dots, l_m the scales used to build the codebook s an ACC signal**output:**The representation of signal s **for** $j = 1$ to m $S \leftarrow$ empty list**for** $t = 1$ to $s.Length - l_j$ add $s[t : t + l_j]$ to S **end** $h_j \leftarrow$ empty list**foreach** patch p in S add $\arg \min_{k'} dist(p, C_j[k'])$ to h_j **end****end****return** $histogram(h_1)|histogram(h_2)|\dots|histogram(h_m)$

in S_a and will then be distributed like patches in S_a or (b) completely in S_b and distributed like patches in S_b or (c) start in S_a and continue into S_b , in which case we know little about the distribution and consider these patches as noise. The key point now is that the number of patches of type (c) is at most twice the length of the patch, and thus can be made small in relation to the total number of patches which is in the order of the length of the signal. More formally:

Let S be an ACC signal composed of a concatenation of t_1 consecutive samples during behavioral mode a and t_2 consecutive samples during behavioral mode b . Denote $p_{mode}(s)$ the probability of a patch s of length l in behavioral mode $mode \in \{a, b\}$. Let s be a patch drawn uniformly from S , then:

$$p(s) = Pr(A)p(s|A) + Pr(B)p(s|B) + Pr(C)p(s|C)$$

$$\geq Pr(A)p_a(s) + Pr(B)p_b(s)$$

$$= \frac{t_1 - l}{t_1 + t_2} p_a(s) + \frac{t_2 - l}{t_1 + t_2} p_b(s)$$

$$= \frac{t_1}{t_1 + t_2} p_a(s) + \frac{t_2}{t_1 + t_2} p_b(s) - \epsilon$$

where events A, B, C denote the patch being all in S_1 , all in S_2 and starting in S_1 and ending in S_2 respectively, and:

$$\epsilon = \frac{l}{t_1 + t_2} [p_a(s) + p_b(s)] \quad (1)$$

ϵ can be made arbitrarily small by making $t_1 + t_2$ large with a constant l , meaning that for patches small enough in relation to the length of the entire signal, the distribution of patches of the concatenated signal is a mixture of the distributions of the parts, with mixing coefficients of the part lengths. We note that this result can easily be extended to a concatenation of any finite number of signals, as long as each one is of sufficient length related to the patch width.

Since behaviors of real-world animals start and stop, and a measured ACC signal is likely to be a concatenation of signals representing different behavioral modes (typically 1-3), the above property inspires a model that is able to capture such mixtures in an explicit fashion. Furthermore, the resulting mixture coefficients may provide some insight into the nature of the underlying behaviors and the relationships between them – which often appear alongside each other and which are more temporally separated, for instance.

C. The proposed model

Let k denote the number of behavioral modes under consideration, and p the dimension of the representation of ACC observations. Following the mixture property presented in the previous section, we assume that every sample is a convex combination of the representation of a “pure” signal of the various behavioral modes. Further, we assume the existence of a matrix $F \in \mathbb{R}^{p \times k}$. The **factor matrix**, such that the $i - th$ column of F is the representation of a pure signal of the $i - th$ behavioral mode, which we will call the factor associated with the $i - th$ behavioral mode. Let s be an ACC sample, then:

$$s = F\alpha + \epsilon \quad (2)$$

where ϵ is some random vector of the appropriate dimensions. In other words, we say that the sample s is a linear combination of the factors associated with each of the behavioral modes with some remainder term. For the full dataset, we then have:

$$S = FA + \epsilon \quad (3)$$

where F is the same matrix, A 's columns are the factor loadings for each of the samples denoted α in 2, and ϵ is a random matrix of the appropriate dimensions. Since our features are non-negative histograms, and we would like the factor loadings to be non-negative (The reason we constrain the factor loadings to be non-negative is that we would like to interpret these values as the *extent* to which the signal is associated to the behavioral modes, and as such these should naturally be non-negative numbers), we constrain the matrices F, A to have non-negative values. We solve for F, A using a least squares criterion:

$$F, A = \operatorname{argmin} \|FA - S\|_F \text{ s.t. } F_{i,j}, A_{i,j} \geq 0 \quad \forall i, j \quad (4)$$

which can be solved, for instance, using alternating non-negative least squares (Algorithm 3) [17], or projected gradient methods [20].

Algorithm 3 alternating non-negative least squares

input:

$S \in \mathbb{R}^{pN}$ the matrix containing the representations of N ACC samples

output:

A factorization $F \in \mathbb{R}^{pk}, A \in \mathbb{R}^{kN}$ of S such that $S \approx FA$

$F, A \leftarrow$ random initialization

repeat until convergence

$$F \leftarrow \min_F \|FA - S\|_F^2 \text{ s.t. } F_{i,j} \geq 0 \quad \forall i, j$$

$$A \leftarrow \min_A \|FA - S\|_F^2 \text{ s.t. } A_{i,j} \geq 0 \quad \forall i, j$$

end

return F, A

IV. RESULTS

In this section we present experiments designed to compare our method to alternatives, and derive insights about the data. The results are then discussed in the next section.

Data for these experiments consists of 3D acceleration measurements from bio-loggers which were recorded during 2012. Each measurement consists of 4 seconds at 10Hz per axis, giving a total of 120 values. A ground truth partitioning of the data was obtained using 3815 field observations each of which was assigned one of 5 distinct behavioral modes (Walking, Standing, Sitting, Flapping, Gliding). The labeling was extending to the entire dataset using standard supervised machine learning methods (see [8], [5] for more details regarding the methodology, and an open-access web-app specializing in supervised learning for movement data of this sort). Experiments were conducted using stratified sampling of 100,000 measurements (20,000 per behavioral mode).

Matrix factorization was performed using the scikit-learn [21] python software library (see [20] for method details). In all experiments the results were stable across repetitions, leading to essentially zero standard deviation, and therefore the reported results correspond to single repetitions.

Our method is compared to the following:

(a) Random partitioning - each sample is assigned a value drawn uniformly from the set of possible partitions $\{1, 2, \dots, k\}$, hence we expect an 80% 0 – 1 loss.

(b) Uniform partition - each sample is assigned the same distribution of $\frac{1}{k}$ per partition, over the k partitions, again leading to an expected 80% 0 – 1 loss.

(c) Kmeans - the sample are partitioned using Kmeans.

(d) Gaussian Mixture Models (GMM) - The samples get coefficients per partition, for each of the k partitions using a GMM.

where (a) and (b) are used as controls, (c) and (d) are used as representative hard and soft clustering methods, respectively.

The data is then divided into two equal parts designated as train and test. Using the training-set we assign to each partition the label (out of the 5 labels in the ground-truth data) that has the highest weight in that partition. The score a sample s gets for label l is the sum of the scores it got in the partitions that were designated with label l . The following evaluation is performed on the test-set only. Resemblance to the ground-truth is measured using log-loss (Figure 1) and 0 – 1 loss (Figure 2). Table 1 shows mean label association using non-negative matrix-factorization (NNMF) with $k = 30$. Data is presented after row normalization to facilitate between-row comparison.

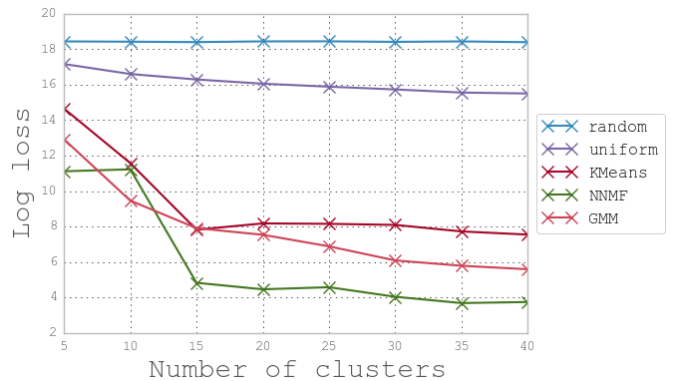


Figure 1. Log loss of soft-assignment to each of the ground-truth classes using each of the methods under consideration. (NNMF: non-negative matrix factorization, GMM: Gaussian mixture model)

V. DISCUSSION

The most striking result of the experiments above is that while the matrix-factorization method performs well compared to the other methods with respect to the log-loss metric (Figure 1), it is not quite as good with respect to the 0-1 loss (Figure 2). In order to better understand this phenomena, we take a closer look at the data. Consider an observation where the animal takes a single step during the 4-second acceleration measurement window, and stands still for the rest of it. In order not to underestimate the amount of walking, we label this sample as Walking. From the mixture property of the features we use (see Methodology section), when using the matrix factorization approach we would expect to get a Walking factor

Table I
MEAN LABEL ASSOCIATION PER GROUND-TRUTH BEHAVIORAL MODE. NNMF WITH 30 FACTORS. NORMALIZED ROWS.

Ground truth / Assignment	Flapping	Gliding	Walking	Standing	Sitting
Flapping	51.25%	13.66%	13.37%	4.33%	17.39%
Gliding	0.75%	49.98%	8.49%	3.95%	36.84%
Walking	2.41%	19.71%	43.92%	20.56%	13.41%
Standing	0.86%	13.30%	1.04%	74.93%	9.88%
Sitting	0.01%	30.88%	0.15%	10.46%	58.50%



Figure 2. 0-1 loss of hard-assignment to each of the ground-truth classes using each of the methods under consideration. For the soft-assignment partitioning methods, hard-assignment is achieved using argmax. (NNMF: non-negative matrix factorization, GMM: Gaussian mixture model)

proportional to the time spent doing so in the measurement windows. Thus, for a sample with some walking (say, less than 50%) we get a miss in the 0-1 loss metric, but a better score in the log-loss which is more sensitive to assignment of very low probabilities to the correct class.

Table 1 sheds more light on the aforementioned result by showing average assignment of factors for each of the ground-truth classes. Flapping samples indeed received the highest weight, on average, on Flapping factors (51.25%), but the Gliding and Walking factors get over 13% each. This may be due to the fact that Storks indeed glide between wing flaps, and may have walked prior to taking off during the observations which are inherently biased to behavior close to the ground (where the observer is). This may also point to the tendency of the field observers to assign the more active behavior to mixed samples (In which case a sample where the bird flaps for a part of the duration of the measurement would be assigned to Flapping, in the same sense that a step or two would qualify an otherwise stationary sample as Walking). We note that the Sitting factors received factor weights higher than expected in all other behavioral modes, meaning there might be a need for a column normalization. We defer this to future research.

VI. CONCLUSIONS

In this paper we describe a matrix factorization approach to behavioral mode analysis from accelerometer data and demonstrate it’s qualities using a large Movement Ecology dataset. While clustering with matrix factorization is by no means a new idea, the novelty here is in the integration with

patch features that theoretically motivate the method in the context of time-series sensor readings for behavioral mode analysis.

The main contribution of this paper is in presenting a framework that will allow for a widespread use of behavioral mode analysis in Movement Ecology, and related fields where determining movement patterns from remote sensor readings is necessary. Further, we introduce the multiscale patch features that may be applicable for many continuous sensor readings, and show that a linear mixture model is justified when using such features.

One issue we did not address in this paper is the scenario of multiple sensor readings. This is of particular interest since most devices produce readings from more than one type of sensor (ex. Gyroscopes and Magnetometers). One possible direction for future work would be to develop a framework for integrating readings from multiple sensors and jointly clustering the readings into behavioral mode partitions. This can be done, for instance, using separate factor matrices: F_1, \dots, F_k for k sensor types, and a single shared factor loading matrix A . Denoting the features matrices S_1, \dots, S_k we now look for matrices A, F_1, \dots, F_k such that $S_i \approx F_i A$ for $i = 1, \dots, k$.

ACKNOWLEDGMENT

This work was supported in part by a grant from the Israel Science Foundation (ISF) to Prof. Daphna Weinshall. We would also like to thank three anonymous reviewers for their helpful comments and suggestions, as well as the numerous people who were instrumental in the collecting and storage of the data.

REFERENCES

- [1] R. Nathan and W. Getz, “A movement ecology paradigm for unifying organismal movement research,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 49, pp. 19 052–19 059, 2008. [Online]. Available: <http://www.pnas.org/content/105/49/19052.short>
- [2] M. Hebblewhite and D. T. Haydon, “Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology.” *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 365, no. 1550, pp. 2303–12, Jul. 2010. [Online]. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2894965&tool=pmcentrez&rendertype=abstract>
- [3] M. Holyoak, R. Casagrandi, R. Nathan, E. Revilla, and O. Spiegel, “Trends and missing parts in the study of movement ecology.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 49, pp. 19 060–5, Dec. 2008. [Online]. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2614715&tool=pmcentrez&rendertype=abstract>

- [4] P. E. Smouse, S. Focardi, P. R. Moorcroft, J. G. Kie, J. D. Forester, and J. M. Morales, "Stochastic modelling of animal movement." *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 365, no. 1550, pp. 2201–11, Jul. 2010. [Online]. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2894957&tool=pmcentrez&rendertype=abstract>
- [5] R. Nathan, O. Spiegel, S. Fortmann-Roe, R. Harel, M. Wikelski, and W. M. Getz, "Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: general concepts and tools illustrated for griffon vultures," *The Journal of experimental biology*, vol. 215, no. 6, pp. 986–996, 2012.
- [6] K. Yoda, K. Sato, Y. Niizuma, M. Kurita, C. Bost, Y. Le Maho, and Y. Naito, "Precise monitoring of porpoising behaviour of Ad{é}lie penguins determined using acceleration data loggers," *Journal of Experimental Biology*, vol. 202, no. 22, pp. 3121–3126, 1999.
- [7] R. P. Wilson, C. R. White, F. Quintana, L. G. Halsey, N. Liebsch, G. R. Martin, and P. J. Butler, "Moving towards acceleration for estimates of activity-specific metabolic rate in free-living animals: the case of the cormorant," *Journal of Animal Ecology*, vol. 75, no. 5, pp. 1081–1090, 2006.
- [8] Y. S. Resheff, S. Rotics, R. Harel, O. Spiegel, and R. Nathan, "AcceleRater: a web application for supervised learning of behavioral modes from acceleration measurements," *Movement Ecology*, vol. 2, no. 1, p. 25, 2014.
- [9] A. C. Gleiss, R. P. Wilson, and E. L. C. Shepard, "Making overall dynamic body acceleration work: on the theory of acceleration as a proxy for energy expenditure," *Methods in Ecology and Evolution*, vol. 2, no. 1, pp. 23–33, 2011.
- [10] D. D. Brown, R. Kays, M. Wikelski, R. Wilson, and a. Klimley, "Observing the unwatchable through acceleration logging of animal behavior," *Animal Biotelemetry*, vol. 1, no. 1, p. 20, 2013. [Online]. Available: <http://www.animalbiotelemetry.com/content/1/1/20>
- [11] M. Takahashi, J. R. Tobey, C. B. Pisacane, and C. H. Andrus, "Evaluating the utility of an accelerometer and urinary hormone analysis as indicators of estrus in a Zoo-housed koala (*Phascolarctos cinereus*)," *Zoo biology*, vol. 28, no. 1, pp. 59–68, 2009. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19358319>
- [12] O. Spiegel, R. Harel, W. M. Getz, and R. Nathan, "Mixed strategies of griffon vultures (*Gyps fulvus*) response to food deprivation lead to a hump-shaped movement pattern," *Movement Ecology*, vol. 1, no. 1, p. 5, 2013. [Online]. Available: <http://www.movementecologyjournal.com/content/1/1/5>
- [13] S. Cooke, "Biotelemetry and biologging in endangered species research and animal conservation: relevance to regional, national, and IUCN Red List threat assessments," *Endangered Species Research*, vol. 4, no. January, pp. 165–185, Jan. 2008. [Online]. Available: <http://www.int-res.com/abstracts/esr/v4/n1-2/p165-185/>
- [14] A. Hindle, D. Rosen, and A. Trites, "Swimming depth and ocean currents affect transit costs in Steller sea lions *Eumetopias jubatus*," *Aquatic Biology*, vol. 10, no. 2, pp. 139–148, Aug. 2010. [Online]. Available: <http://www.int-res.com/abstracts/ab/v10/n2/p139-148/>
- [15] W. I. Sellers and R. H. Crompton, "Automatic monitoring of primate locomotor behaviour using accelerometers." *Folia primatologica; international journal of primatology*, vol. 75, no. 4, pp. 279–93, 2004. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/15316155>
- [16] K. Q. Sakamoto, K. Sato, M. Ishizuka, Y. Watanuki, A. Takahashi, F. Daunt, and S. Wanless, "Can ethograms be automatically generated using body acceleration data from free-ranging birds?" *PloS one*, vol. 4, no. 4, p. e5379, Jan. 2009. [Online]. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2671159&tool=pmcentrez&rendertype=abstract>
- [17] Y.-X. Wang and Y.-J. Zhang, "Nonnegative matrix factorization: A comprehensive review," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 25, no. 6, pp. 1336–1353, 2013.
- [18] T. Li and C. Ding, "The relationships among various nonnegative matrix factorization methods for clustering," in *Data Mining, 2006. ICDM'06. Sixth International Conference on*. IEEE, 2006, pp. 362–371.
- [19] C. H. Ding, X. He, and H. D. Simon, "On the equivalence of nonnegative matrix factorization and spectral clustering," in *SDM*, vol. 5. SIAM, 2005, pp. 606–610.
- [20] C.-b. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural computation*, vol. 19, no. 10, pp. 2756–2779, 2007.
- [21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.