SeeGull: Sea Gull Path Predictions in the Canadian and Alaskan Regions

Project Category: Application

William Cai
Department of Mathematics
Stanford University
willcai1@stanford.edu

Hagop Chinchinian
Department of Electrical Engineering
Stanford University
hagop@stanford.edu

Russell Tran
Department of Computer Science
Stanford University
tranrl@stanford.edu

1 Introduction

Animal movements usually contain information about various animal behaviors. Based on spatiotemporal changes in the environments that the animals are in, animals instinctively change their behaviors. Therefore, analyzing and predicting animals' movement patterns according to spatiotemporal change offers an opportunity to understand their unique traits and acquire ecological insights into animals [1]. Moreover, having the ability to predict animals' movement also helps prevent potential anthropogenic threats to the animals [2] and help with the understanding the spread of infectious diseases through animal populations [3]. Additionally, for the particular case of gull path prediction, being able to predict a gull's location may enable servicing or replacement of the gull's tag.

Since the data of the gull from the Canadian and Alaskan regions have never been analyzed with other environmental factors before, this project dives into the previously unexplored topic of gull path prediction in the Canadian and Alaskan regions which entails (1) activity classification based on positional data and (2) forecasting the future location of a gull using a long short-term memory (LSTM) neural network architecture given its prior locations and weather data.

Time-stamped latitudinal and longitudinal positions in sets of $P = 5$ of the gulls were provided as inputs for the activity classification task. Testing with various classification algorithms revealed that logistic regression performed on 10 features extracted with principal component analysis provided the best trade-off between performance on the testing data and number of features. Five prior time-stamped latitudinal and longitudinal positions of the gull as well as approximations of the air temperature, air pressure, wind direction, and wind speed were provided as inputs for the future location prediction task, and a long short-term memory neural network was used to determine the next latitudinal and longitudinal position of the gull.

2 Related Work

Previous work on predicting the movements of blue whales in the California region[2], black carpenter ants (Camponotus pennsylvanicus) in a man-made nest[3], and other types of animals [3][4][5] have demonstrated promising potential of using machine learning algorithms in animals' movements prediction. In particular, long short-term memory architectures were used for predicting gull movement in Europe in [3]. Long short-term memory architectures were also used for prediction of aircraft trajectories [6]. In addition, the multi-domain-data-driven approaches that combine geographical data, oceanic data, wind data, and other types of data also shows huge potential in those research [3][2][4][5] and set up a data fusion paradigm when tackling the animal movement prediction problem. A previous study that categorizes whale behavior as either "transiting" or "foraging" using a switching state-space model to perform the classification [7] inspired this project to apply various classification algorithms and long short-term memory (LSTM) neural network to predict the paths of sea gulls in the Canadian and Alaskan regions.

3 Dataset

Positional information for 37 gulls obtained via GPS and ARGOS was obtained from a US Geological Survey database through Movebank [8]. Presently, only the ARGOS data is being analyzed. Using latitude to estimate the location of the gulls, it was determined that the majority of both the GPS and ARGOS datapoints consisted of locations in Alaska and Canada, with only a minority of the datapoints consisting of locations south of the United States-Canada border. It was thus decided that location data outside of Alaska and Canada would be excluded for the purposes of the long short-term memory (LSTM) neural network for sea gull path prediction, as there is likely insufficient positional data to both train and test the neural network. However, the few datapoints south of the United States-Canada border may be useful for the activity classification task, particularly for examples of long-distance, uni-directional travel.

4 Activity Classification

4.1 Task Definition

The activity classification task is a multi-class classification problem that involves determining which of four activities a single gull is performing over P recorded positions of the gull: Category 1 involves little to no movement; category 2 involves minor movement in multiple directions around the same general geographic area; category 3 involves unidirectional travel over short to moderate distances; and category 4 involves unidirectional travel over long distances.

4.2 Labeling and Training/Test Data

$P = 5$ was chosen arbitrarily to provide reasonable granularity without having to consider individual point-to-point paths. A single individual performed labeling of the data using the listed classification criteria.

Data from 2 gulls was used for testing, and data from 5 gulls was used for training. Of the two gulls that were used for testing, the first gull's geographic extent ranged from Alaska all the way south to California, and the second gull's geographic extent ranged from Alaska all the way south to the Oregon/Washington area. Of the five gulls that were used for training, two of the gulls' geographic extent ranged from Alaska to California, two of the gulls' geographic extent ranged from Alaska to Oregon/Washington, and one of the gull's geographic extent covered Alaska. The wide geographic range of the gulls in the training data ensures that all four of the identified activity categories are observed in the training data.

4.3 Feature Mappings

Latitude and longitude are not inherently useful for the activity classification task, and as such, various feature mappings were introduced. All of these features are computed from the latitude and longitude: (1) distance between the first and last point in the set of P points (1-dimensional), (2) initial and final bearings corresponding to first and last point in the set of P points (1-dimensional for each bearing), (3) pairwise distances (4-dimensional for $P = 5$), (4) pairwise initial and final bearings (4-dimensional for $P = 5$ for each bearing), (5) average pairwise distance (1-dimensional), (6) standard deviation of pairwise distances (1-dimensional), (7) maximum pairwise distance (1-dimensional), and (8) minimum pairwise distance (1-dimensional). These respective mapped features are combined into a vector and treated as a single input.

4.4 Principal Component Analysis (PCA)

Principal component analysis is applied to the training dataset to identify features that are most crucial to the multi-class classificaton problem. There are originally 19 features, and the authors explored reducing the number of features while maintaining adequate performance.

4.5 Definitions of Algorithms Used in Activity Classification

(1) Support Vector Machine (SVM) with the following kernel functions: RBF kernel, polynomial kernel, and sigmoid kernel. The SVM is a set of supervised learning methods used for classification. Its objective is to find a hyperplane in an N-dimensional space (N is the number of features) that distinctly classifies the data points. It is effective in high dimensional spaces and versatile with different types of kernel functions.
(2) Naive Bayes. The Naive Bayes is a set of supervised learning methods that applies Bayes' theorem with the assumption that every pair of features are conditionally independent.
(3) Logistic Regression: a set of supervised learning methods that is used for the classification problems. Logistic Regression uses the sigmoid ($\sigma(z) = \frac{1}{1+e^{-z}}$) function to transform the hypothesis of a linear regression to give outputs that have values between 0 and 1.

4.6 Results

Multiclass classification was attempted on the transformed data using support vector machines with an RBF kernel, support vector machines with a polynomial kernel, support vector machines with a sigmoid kernel, logistic regression, and naive Bayes regression. The average F1 scores and the weighted average F1 scores of difference algorithms are shown in the table below. After determining that logistic regression performed best, the authors experimented with different dimensionality reductions on the input data followed by logistic regression classification on the reduced data. Reducing to 10 dimensions does not significantly affect performance, whereas reducing to 9 dimensions does significantly affect performance. Therefore, it was concluded that applying PCA with 10 dimensions allows for maximal dimensionality reduction while not significantly affecting performance. PCA identified the following top 10 features (listed from most important to less important): net distance magnitude, maximum pairwise distance, standard deviation of pairwise distances, first pairwise distance, last pairwise distance, third pairwise distance, average pairwise distance, second pairwise distance, last pairwise initial bearing, last pairwise final bearing.

Classification Algorithm	Average F1 score	Weighted average F1 score
Support Vector Machine, one versus rest, RBF kernel	51%	46%
Support Vector Machine, one versus rest, polynomial kernel	59%	55%
Support Vector Machine, one versus rest, sigmoid kernel	36%	32%
Naive Bayes	67%	62%
Logistic, all 19 features	77%	75%
Logistic, PCA reduced to 9 features	66%	66%
Logistic, PCA reduced to 10 features	76%	74%

Table 1: Performance of various classification algorithms on testing data

A confusion matrix for the logistic classifier with 10 features was further generated to better visualize performance.

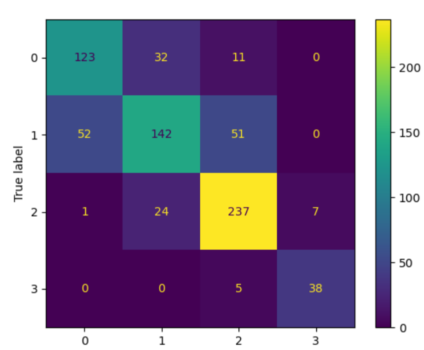

Figure 1: Confusion matrix for Logistic regression, PCA-reduced to 10 dimensions

5 Path Prediction

5.1 Task Definition

Given T prior locations of the gull in the last T hours as well as weather data across the geographical area at the same respective times, predict the next position of the gull.

5.2 Data Description

The map of the area of interest was divided into a 10 by 10 grid. The bounds of the map were determined by the minimum and maximum longitude and latitudes of the weather stations from which weather data (air temperature, air pressure, wind speed, and wind direction) was obtained. Latitude and longitude coordinates of the gull locations were translated into the 10 by 10 grid using a one-hot representation, with a 1 denoting that the gull was at the particular cell of the grid and a 0 denoting that the gull was not at the particular cell of the grid. In order to create a time series, the position of the gulls was sampled every hour, with the location nearest in time used if a gull location was not available in a particular hour.

The data was converted into a matrix $D \in \mathbb{R}^{|G|\times|T_a|\times|P|}$, where $|G| = 37$ is the number of gulls, $|T_a|$ is the total number of times in the time series and defined as the number of hours between the first and last time a gull position was recorded, and $|P| = 500$ denotes the dimension of the vector encoding information about the gull position and weather at a particular point in time. Note that this latter vector include a 100-vector for each parameter represented in the form of the 10 by 10 grid; since there are 5 vectors (gull position, air temperature, air pressure, wind speed, and wind direction), $|P| = 5 \times 10 \times 10 = 500$.

The chronologically earliest 70% of the data was used for training; the next 5% of the data was set aside for validation but not used; and the chronologically latest 25% of the data was used for testing.

5.3 Neural Network Architecture

The input matrix is then defined as $I \in \mathbb{R}^{|G|\times|T|\times|P|}$, where $|G|$ denotes the number of gulls, $|T|$ denotes the number of prior positions used to predict an output, and $|P|$ denotes the dimension of the vector encoding gull position and weather information for a particular gull at a particular time. The output matrix is defined as $O \in \mathbb{R}^{|G|\times|P_0|}$, where $|P_0| = 100$ denotes the number of cells in a grid. The index of the maximum element of the output matrix is meant to correspond to the present location of the gull within the 10 by 10 grid. For an arbitrary time point i, the input matrix can be selected as the $i-5, i-4, i-3, i-2, i-1$th elements of the second axis of the data matrix D and the output matrix can be selected as the ith element of the second axis of the data matrix D.

The neural network consists of a single-layer LSTM followed by a linear layer. The output of the neural network a matrix $\in \mathbb{R}^{|G|\times|T|\times|H|}$ where $|H| = 6$, and the output of the linear layer is a matrix $\in \mathbb{R}^{|G|\times|T|\times|P_0|}$. The last layer is then selected [9], resulting in a matrix $\in \mathbb{R}^{|G|\times|P_0|}$.

To train the neural network, gradient descent with the ADAM optimizer is run with data from $|G_s| = 20$ gulls randomly chosen from the $|G| = 37$ gulls, five prior positions for each of the gulls, and one current position for each of the gulls the on every iteration for a single randomly chosen point in time. Cross-entropy is used as the loss metric.

5.4 Definitions of Algorithms Used in Path Prediction

A Long Short-Term Memory (LSTM) network architecture was adopted when tackling the Path Prediction problem. LSTMs are a type of modified recurrent neural network (RNN) capable of learning long-term dependencies. Like other RNNs, LSTMs have a chain of repeating modules of neural network. However, the structure of the repeating modules in the LSTM is different from other RNNs. A common LSTM module is composed of a cell, an input gate, an output gate, and a forget gate. The cell remembers values over arbitrary time window and the three gates control the flow of information into and out of the cell.

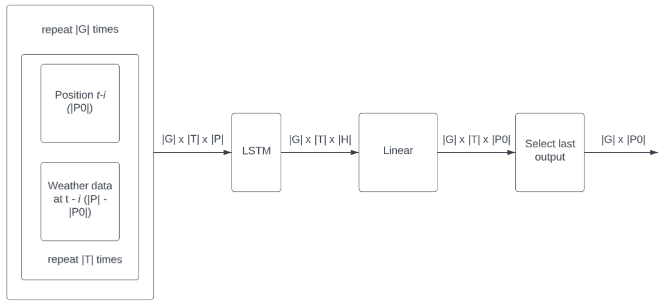

Figure 2: Neural Network Architecture

5.5 Results

After training for 3000 epochs, it was observed that the training loss did not converge but instead oscillated significantly. Additionally, the cross entropy loss for the training data involved repeated values in the range of 3.3-3.5, which is rather high. Further debugging is needed to ensure convergence of the LSTM model.

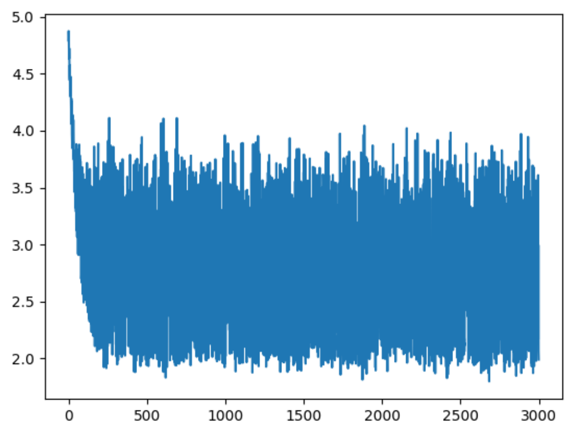

Figure 3: Convergence of training loss

6 Conclusion and Future Work

Overall, the majority of the classification algorithm (except the SVM with sigmoid kernel) shown better performance than random guessing. We also noted that logistic regression with 10 features extracted by PCA has the best performance among all the classification algorithms. While the results from the activity classification task is relatively promising, further work is needed for the path prediction task to ensure that the training loss converges and does not oscillate.

7 Contributions

Hagop Chinchinian downloaded all data, wrote data processing code, helped assess the different classification algorithms, and configured the LSTM pipeline. William Cai laboriously labeled the datasets for the activity classification task, helped assess the different classification algorithms, contributed to the code for principal component analysis of the input data, and configured the LSTM pipeline for the path prediction algorithm. Russell Tran contributed to the earlier stages of the project, including project ideation, devising feature mappings for the activity classification task, and devising a standardized scoring procedure for the activity classification task. The authors did not collaborate with any non-author. The project is solely being submitted for CS 229.

References

[1] Jehyeok Rew, Sungwoo Park, Yongjang Cho, Seungwon Jung, and Eenjun Hwang. Animal movement prediction based on predictive recurrent neural network. In *Sensors Volume 19 Issue 20 10.3390/s19204411*, 2019.

[2] Karin A. Forney Evan A. Howell Elizabeth Becker Aimee L. Hoover Ladd Irvine Monica DeAngelis Steven J. Bograd Bruce R. Mate Helen Bailey Elliott L. Hazen, Daniel M. Palacios. Whalewatch: a dynamic management tool for predicting blue whale density in the california current. In *Journal of Applied Ecology*, volume 54, pages 1415–1428, November 2016.

[3] Dhanushi A. Wijeyakulasuriya, Elizabeth W. Eisenhauer, Benjamin A. Shaby, and Ephraim M. Hanks. Machine learning for modeling animal movement. In *PLoS One*, volume 15, July 2020.

[4] Tsubasa Hirakawa, Takayoshi Yamashita, Toru Tamaki, Hironobu Fujiyoshi, Yuta Umezu, Ichiro Takeuchi, Sakiko Matsumoto, and 9Ken Yoda. Can ai predict animal movements? filling gaps in animal trajectories using inverse reinforcement learning. In *ECOSPHERE*, volume 9, October 2018.

[5] M.P. Herzog J. Howar K.D. Hyrenbach J.E. Zamon D.G. Ainley J.A. Wiens K. Morgan L.T. Ballance N. Nur, J. Jahncke and D. Stralberg. Where the wild things are: Predicting hotspots of seabird aggregations in the california current system. In *Ecological Applications*, volume 21, pages 2241–2257, 2011.

[6] Shan Tian Lan Ma. A hybrid cnn-lstm model for aircraft 4d trajectory prediction. In *IEEE Access*, volume 8, pages 134668 – 134680, June 2020.

[7] Behavioural estimation of blue whale movements in the northeast pacific from state-space model analysis of satellite tracks. In *Endangered Species Research*, volume 10, pages 93–106, November 2009.

[8] Ahlstrom C.A. van Toor M.L. Woksepp H. Chandler J.C. Reed J.A. Reeves A.B. Waldenström J. Franklin A.B. Bonnedahl J. Ramey, A.M. and D.C. Douglas. Tracking data for three large-bodied gull species and hybrids. In *U.S. Geological Survey data release, doi:10.5066/P9FZ4OJW*, June 2020.

[9] Gabriel Loye. Long short-term memory: From zero to hero with pytorch. June 2019.