Predicting Migration Location of White Stork

Monica Patel
260728093
monica.patel@mail.mcgill.ca

Njeri Muguthi
260588566
njeri.muguthi@mail.mcgill.ca

Krtin Kumar
260713550
krtin.kumar@mail.mcgill.ca

***Abstract*— White storks are migratory birds that have a diverse migratory pattern that often causes varied wintering locations among juvenile birds within the species. Birds that winter in areas with a higher human population tend to behave differently from those that don't. Since they are affected by their ecological system, it would be beneficial to map out their widespread migration patterns. A better understanding of what areas are affecting the birds in their migration routes would help in knowing where to concentrate conservation efforts.**

We develop several algorithms to help us to help us predict migratory patterns of these birds. First we cluster the birds in order to find out which birds are traveling together so we can map out maps for these clusters. Since we are working with time series data we then use LSTM and compare it with a regression model to predict future paths.

I. Introduction

Human interference is getting broader and encroaching the lives of animals in their natural state. For white storks it has been shown that birds overwinter in areas where there is a large human population causing effects such as shorter foraging trips, closer wintering grounds or a complete suppression of migration by the birds. Given the adverse effects this can have in the life of a naturally migrating bird, there is the issue of how to better understand their migratory paths so we have concentrated conservation efforts.

Working with migratory data from Movebank we choose a set of features that are known to affect migration patterns. The features are: acceleration of the bird, pressure, wind speed and temperature. We will try to solve the problem in two ways. First given historic complete migration paths, we will predict how the migration path of a new bird will look like, we will call this as $Task1$. Second, given historic incomplete paths, we will predict how the future paths may look like, we will call this as $Task2$. The core problem to in both the task is same, of finding future paths is predicting locations at the next time step. But the two tasks will differ on the portion of data we choose to train and test. The biggest challenge is that we have to keep track of birds that are traveling together or have similar paths, especially for $Task1$. Each cluster will then be used to train a separate model. Once we have features for each cluster, then we can train a model and then use it to make our predictions.

II. Related work

Significant amount of work done in movement ecology aims at measuring and characterizing movements of foraging animals and measuring potential consequences of these movements.[6]. Methods being used for these purposes primarily deals with modeling animal movement and does not account for the effects environmental factors have on migration. Animal movements are often modeled using co-related random walks[4] or Hidden Markov models[5]. The neural network (NN) has been applied successfully to many problems involving time series prediction and modeling of non-linear systems. This method is also capable of accepting accepting time series environmental features like pressure and temperature and help predict the movement sequences. Our work primarily takes inspiration from [7] where elephant movement is modeled using RNN. But RNN in above work only models movement data. We propose to model temperature and pressure effects on movement with movement data. We use a data set which was originally used in the study of correlation between migration decisions taken by White Stork and energy expenditure, over the life of White Stork[1]. They used acceleration data to make estimate of energy value using overall body dynamic acceleration (ODBA). We chose this data set particularly because we suspect that acceleration data could be a good indicator of bird path.

III. Data Structure and Processing

III-A. Data Description

Our data-set consists of life time data of 70 White Storks from 2013-2014. The birds were fitted with trackers at 8 different locations in different countries. The data consists of Global Positioning System (GPS) coordinates and 3-axis acceleration data. GPS data is in the form of longitudes and latitudes, while acceleration is in the form of raw analog data. Both GPS and acceleration were recorded at an interval of 5 minutes, acceleration values were recorded for 3.8 seconds at each 5 minute interval, at a frequency of 10.54Hz. In addition to acceleration and GPS coordinates, we used environment data from European Center for Medium-Range Weather Forecasts[9] (ECMWF), using Movebank[8]. We used temperature 2 meter above ground (or sea) level, wind velocity (magnitude and it's two components u_x and v_y) 10 meter above ground (or sea) level and mean sea level pressure. Table I summarizes the data used by us along with their units. Acceleration data was mapped to GPS data using one to one mapping as both have the same duration, data from Movebank was interpolated using bi-linear interpolation.

III-B. Data Processing

The data processing was done in two steps, first we used clustering to group birds which had similar paths. This was

Name	Interval	Frequency	Duration	Unit
GPS	5 min	-	Trigger	Long/Lat
Acceleration	5 min	10.54 HZ	3.8 sec	mV
Temperature	6 hrs	-	Trigger	Kelvin
Wind Velocity	6 hrs	Unknown	3 sec	m/s
Pressure	6 hrs	-	Trigger	Pascal

TABLE I

NAME IS THE TYPE OF DATA USED; INTERVAL IS DATA CAPTURE TIME; FREQUENCY IS NUMBER OF SAMPLES TAKEN IN THE GIVEN DURATION; TRIGGER MEANS ONLY SINGLE SAMPLE AT EACH INTERVAL

Feature	Type	Regression	LSTM
X Acceleration	Continuous	Yes	Yes
Y Acceleration	Continuous	Yes	Yes
Z Acceleration	Continuous	Yes	Yes
Temperature	Continuous	Yes	Yes
Pressure	Continuous	Yes	Yes
Wind Speed	Continuous	Yes	Yes
X wind speed	Continuous	Yes	Yes
Y wind speed	Continuous	Yes	Yes
Time stamp	Continuous	Yes	No

TABLE II

FEATURES USED IN LSTM AND REGRESSION MODELS

specially necessary for solving $Task2$, as we had to made sure, when our model makes prediction for a different bird, then that bird should belong to the same class (i.e. similar path). Once, we had clusters of birds our next task was to properly encode features, for training the models. We shall call this as $rawfeatureencoding$ (RFE). A slightly, modified version of RFE will be used for regression, which we will describe in section IV-B.

III-B.1. ***Clustering:*** *:* To group similar bird paths, we first plotted the graph on a 2-D coordinate system. For each image thus obtained (number of images is 70, equal to number of birds), we extracted SIFT descriptors using openCV[10]. We then performed k-means clustering on the SIFT descriptors to form 2000 clusters, which will be our features for grouping images with similar paths. The number of SIFT descriptors were chosen manually based upon best performance using visual inspection. This was possible to do in our case, as we had a data-set of only 70 birds. In section VIII we will discuss how we can automate this task for future studies.

III-B.2. ***Raw Feature Encoding:*** *:* Our input features consisted of Temperature, Pressure, time stamp of each event, wind velocity in x and y axis, wind speed and x,y,z axes accelerations. Time stamp was converted to Unix time stamp in seconds. For acceleration along each axis and for each time step, readings were taken for 3.8 seconds at 10.54 Hz, which means we had 40 observations each in milli-volts. Ideally, we would have liked to convert the values to m/s^2, but the authors of paper[1] didn't disclose the calibration settings for each sensor which is required for the conversion. So, we decided to take the average of raw values itself for each axis as values for our features. Finally, we normalized all our features using z-score normalization, which is $(x - x_{mean})/stdev(x)$.

LSTM vs Regression Features: The above features were used directly in LSTM model, but we used a special encoding scheme in order to use a regular regression model. The features were encoded using the Bi-gram assumption, that is output at T+1 (Y_{T+1}), depends only on Y_T, X_T and X_{T+1}. This could be a strong assumption and could make our model biased, but would give us a good baseline predictor. Hence the features for regression were chosen as $X_{T+1} - X_T$ and Y_T. By choosing this we are making another assumption that there is linear relationship between the two time steps. Table II summarizes the set of features used and which model uses it. Note, Time stamp was indirectly used in LSTM as it is an inherent part of how LSTM model is built.

IV. METHODOLOGY

Our objective is to solve path prediction problem, in two ways. $Task1$ is to learn on complete paths of historic data, and then make prediction of the path for a new bird. $Task2$ is train on historic incomplete paths of all the birds, and then predicting their future paths. $Task2$ can be useful to predict future bird locations given previous 'n' time steps, while $Task2$ is more general and will require only the first time step. Both the tasks are similar in terms of the capability of the learned models, but we will choose our evaluation methods in a slightly different way, so that our model solves the intended task. For $Task1$, we will train and test on different birds but both will include the complete time steps, this will allow our model to generalize better for $Task1$. For $Task2$, we will train and test on the same birds and split data for each bird at n^{th} time step.

IV-A. Clustering

Our data set consists of birds from different origins, this can create a problem for our learned models. If birds having different paths are trained then, the model will try to find a particular pattern which fits both of them. This could create problems, as such a pattern won't exists and thus will add error to our model. It will be more relevant if we can separate these paths and train different models. To achieve this we used clustering, idea is to perform image clustering on the paths, the algorithm used will be explained in section V-A

IV-B. Regression

We use a standard regression model and not an auto-regression model for handling time series data. In order to use a standard regression model we encode our features in a certain way as explained in section III-B.2. This way we can aim to train a biased model, which assumes that the next prediction only depends on the previous data point. In section VIII we will discuss, how we can improve such a model using boosting in order to reduce bias.

IV-C. LSTM

The Long short memory model adopted here has memory block connected through layers. LSTM network [2] is chosen for the proposed work over Elman RNN for it can deal with longer term seq and handles vanishing gradient problem

using back-propagation Through Time. The network consists of two hidden layers and one drop out layer with input to last hidden layers connect via many-to-many connection and last hidden layer is connected to dense output layers using many-to-one connections. We make use of stacked hidden layers because they can make better use of parameters by distributing them over the space through multiple layers. We use 20 % dropout for data-driven regularization in model. [3] proposes many different places where dropout should be applied we chose *inside-option* from the paper to apply drop out, where drop out acts on the output values of the LSTM within the recurrence loop.

V. Algorithm and Cross-Validation

Validation Techniques

Cross-validation technique used for $Task1$ and $Task2$, models differ slightly. For $Task1$ if a cluster had n birds, we kept one bird as test and on the n-1 birds we performed leave one out cross validation (This was done for each cluster separately). Important point to note here is that we performed leave one out cross-validation on the birds and not individual data points as such an evaluation criteria better suits the problem for $Task1$. For $Task2$ we kept 20 % of time series data as the test set. For validation we again divided the time series data by k % and varied k from 5 % to 40 % in intervals of 5 to calculate validation performance (The average for all the cases). Figure V shows the plot of varying Standard Deviation (in meters) and Error (in meters), with varying values of k these results are from linear regression model.

Error Evaluation

To evaluate error between longitudes and latitudes, we estimated the distance between the test points and the prediction points using Haversine distance formula. Equation 1 shows the Haversine formula, where $varphi$ is latitude and λ is longitude.

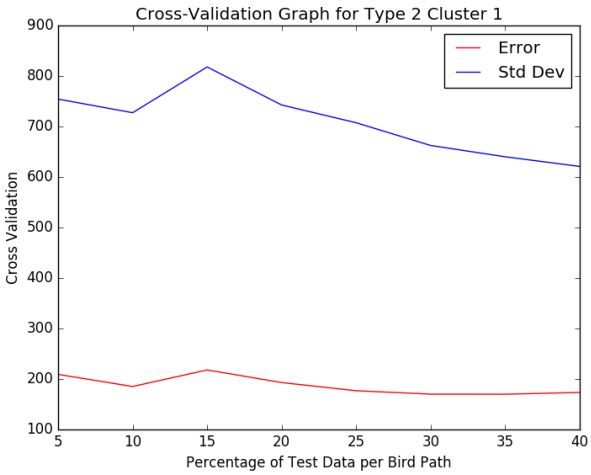
$$d = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right) \tag{1}$$

V-A. Clustering

Using the features obtained as explained in section III-B.1, we performed Agglomerative clustering from Sci-kit[11]. Agglomerative clustering belongs to the family of Hierarchical Clustering, which builds nested clusters by splitting or merging them. Further, we used cosine similarity kernel (Equation 2) and maximum linkage for our algorithm settings. We chose these parameters as they performed the best in our case, since the data set contained 70 points (each bird), we were able to manually check the label and check the performance of the algorithm. To choose the number of clusters, we chose the one which best performed on our data set based upon precision and recall. We didn't use any cross-validation technique, since this is an unsupervised algorithm and clustering here was used as a data filtering technique and not for prediction of our main problem.

$$K(x, y) = \frac{xy}{|x||y|} \tag{2}$$

Fig. 1. Regression Validation Error (Values in Meters)



V-B. Linear Regression:

Cross-validation and evaluation was performed as explained above. We used a simple linear-regression model from sci-kit[12], it had no hyper-parameters to cross-validate and served as a good baseline predictor. We trained two separate models for longitude and latitude.

V-C. LSTM:

We trained separate models for *Latitude* and *Longitude* as described in [7].Merging can be taken up as future work. Data from the GPS sensor for white stock are taken very frequently taken and most of the data does not contribute to long term migration path prediction. We run exponential weighted moving average(EWMA) smoothing on data for smoothing the data before feeding it to the network. This helps network to converge appropriately and avoid oscillation noise. Alpha for EWMA is calculated using half-life of the time series data using following equation 3 where half life is calculated using $\log(2)/autocorrelation(series)$. Smooth data with 5 steps delay and temperature data is passed to the LSTM network given in 2. Adamax optimizer is used to update weights for the network. Since network aims at predicting a continuous valued task linear activation function were used for the neuron. The network uses state fullness of LSTM, where network is trained for one epoch and resets. Batch size (= 32) for the input was fixed exploring already available literature.

$$\alpha = 1 - \exp\left(log(0{,}5)/halflife\right) \tag{3}$$

VI. Results

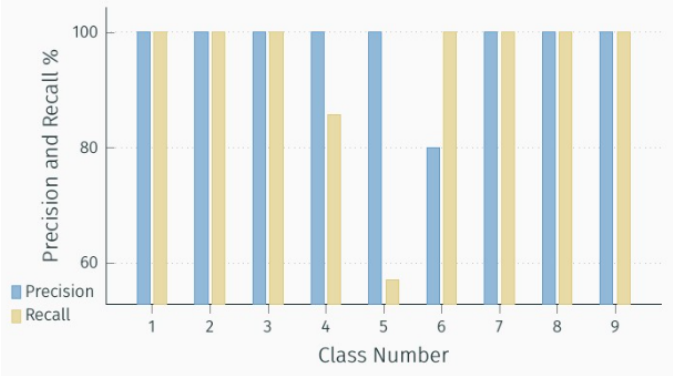
VI-A. Clustering

We obtained an accuracy of **94.3 %** on our data set, the precision and recall are shown in Fig 3. Out of the 9 clusters only **3 clusters** were chosen for training models (Fig: 3 and 4 shows examples of accepted and rejected clusters). This is because other clusters had small amounts of data or had few birds in them. This is possible if either there was something

Fig. 2. LSTM model

Keras.layers.cores.masking	Input	(None,5,2)
	output	(None,5,2)
Keras.layers.recurrent.LSTM	Input	(None,5,2)
	output	(None,100)
Keras.layers.cores.dropout	Input	(None,100)
	output	(None,100)
Keras.layers.recurrent.LSTM	Input	(None,100)
	output	(None,100)
Keras.layers.cores.Dense	Input	(None,100)
	output	(None,1)

Fig. 3. Clustering Precision and Recall



wrong with the sensors or the birds didn't survive. According to Flack A.[1] it is the later reason.

VI-B. Linear Regression

For both $Task1$ and $Task2$ we evaluated our model in two ways, first is predicting the next position (Discrete) and second, is continuous prediction for the entire test set based upon a single instance of bird path. We see that $Task1$ performs better for distance error for discrete prediction (Table III) and continuous prediction (Table IV) for all the clusters. For standard deviation this holds true for all clusters except for cluster 2 in continuous evaluation, where $Task2$ performs better. As expected continuous evaluation performs worse than discrete evaluation, this also reflected in Fig 7 and 9, we see that the continuous path predicts correctly in the beginning and then almost averages out the path, which was expected from linear regression. The discrete path however continuous to predict everything very accurately.

VI-C. LSTM

Results for LSTM have been summarized in Table V, these are continuous predictions and when compared to continuous

Fig. 4. Accepted Cluster

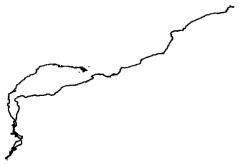


Fig. 5. Rejected Cluster

Fig. 6. Task 2 Path Comparison for Regression

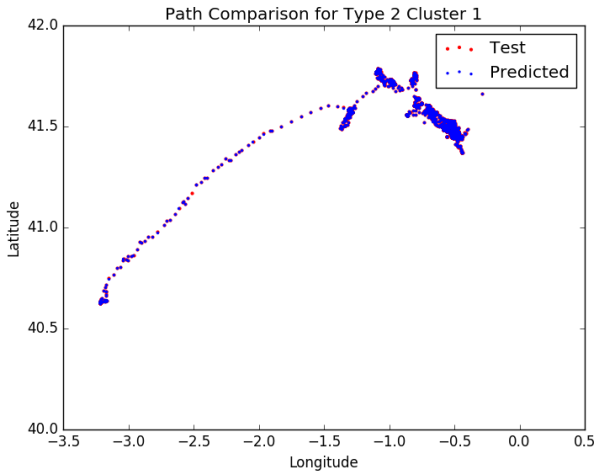


Fig. 7. Task 1 Path Comparison for Regression

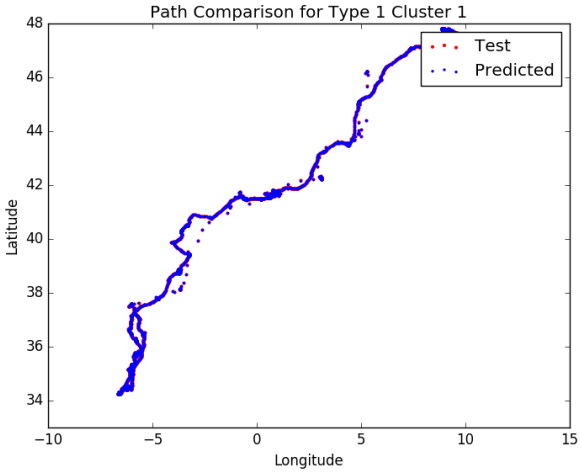


Fig. 8. Task 2 Continuous Path Comparison for Regression

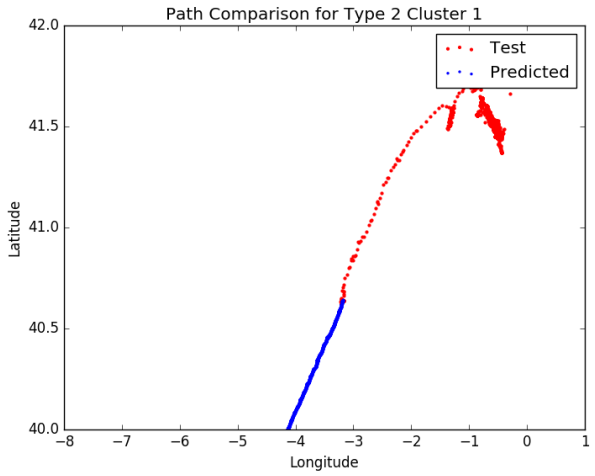
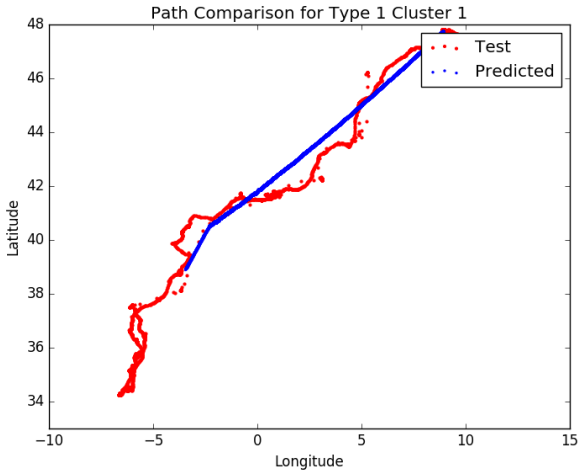


Fig. 9. Task 1 Continuous Path Comparison for Regression



Task	Cluster	Regression					
		Error			Std Dev		
		Train	Valid	Test	Train	Valid	Test
Task 1	1	177.7m	166.1m	165.9m	736.8m	651.3m	1234.4m
	2	334.1m	332.5m	415.9m	1491.6m	1747.9m	1102m
	3	704.5m	895.9m	606.2m	1532.4m	1453.2m	1885.8m
Task 2	1	168.1m	186.9m	253.7m	661.5m	709.1m	1302.8m
	2	342.9m	341.4m	454.2m	1405.5m	1024m	1737.2m
	3	640.7m	743.6m	766.5m	1588.4m	1106.9m	2524.1m

TABLE III

REGRESSION RESULTS FOR TASK 1 AND TASK 2

Task	Cluster	Error	StdDev
Task 1	1	878.5 km	539.4 km
	2	1320.1 km	1241.9 km
	3	2682.2 km	1379.9 km
Task 2	1	1505.6 km	697.0 km
	2	3005.0 km	889.2 km
	3	3860.5 km	2733.5 km

TABLE IV

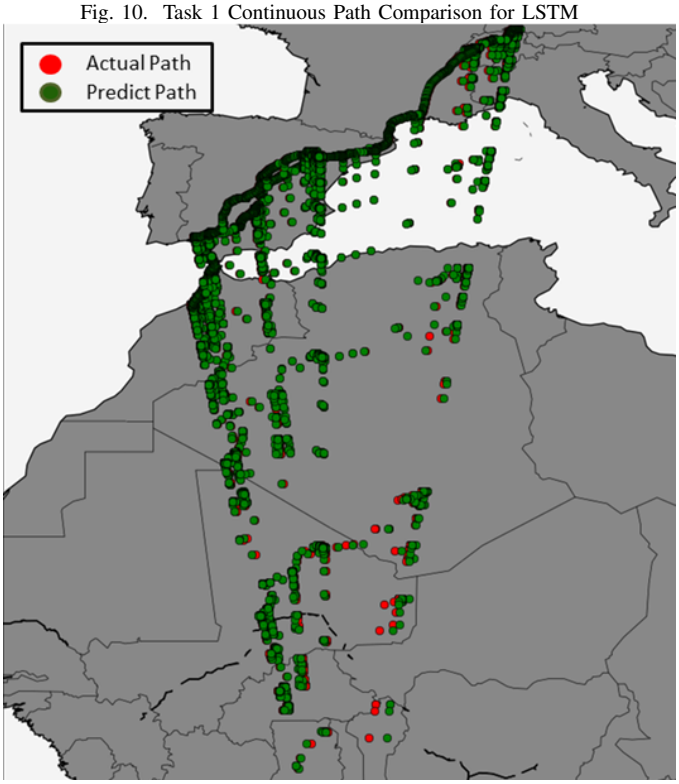
REGRESSION CONTINUOUS PREDICTION OF TASK 1 AND TASK 2

Task	Cluster	Regression			
		Error		Std Dev	
		Train	Test	Train	Test
Task 1	1	50.6km	30.4km	289.2km	244.8km
	2	95.1km	76.5km	585.4km	218.5km
	3	200.7km	111.5km	601.4km	373.9km
Task 2	1	48.0km	46.4km	259.7km	258.3km
	2	36.3km	83.1km	551.8km	344.4km
	3	67.7km	140.3km	623.5km	500.5km

TABLE V

LSTM RESULTS FOR TASK 1 AND TASK 2 CONTINUOUS

Fig. 10. Task 1 Continuous Path Comparison for LSTM

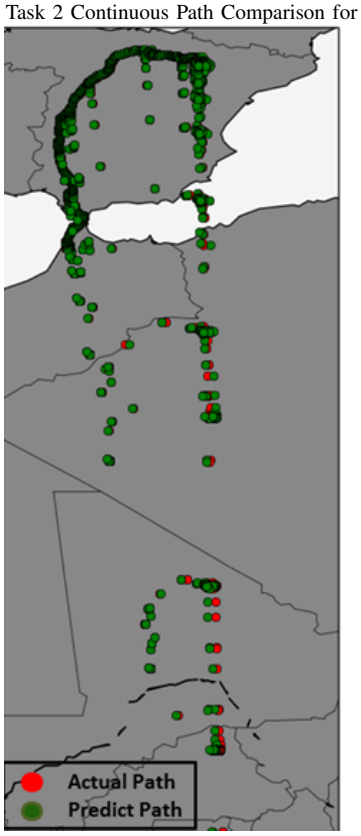


prediction from our baseline predictor of linear regression we see an improvement of about 90 % on average over all the clusters on test results. We see improvement in the standard deviation of the error in distance also of about 50 %, which means the model is not only predicting more accurately but also has a Gaussian curve with smaller width.

VII. DISCUSSION

We see that discrete prediction for linear regression works well, the accuracy falls sharply with continuous prediction. This, is because for continuous prediction we use the value predicted by our linear model as the input to the next time step. Since, our model is linear we will always get a linear prediction, thus if the actual bird path is linear this model will work well. In order to make our model more robust we need to use a non-linear model. Our intuition after the results from the baseline were proven correct when we evaluated our LSTM model. Which showed a huge improvement over the baseline as it makes less assumptions about the hypothesis set and thus is less biased. We can also see from the plotted

Fig. 11. Task 2 Continuous Path Comparison for LSTM



figures that 11 and 10 that the predicted path is no longer linear and closely follows the actual path. We observe that $Task1$ performs better compared to $Task2$, one of the reason could be that $Task1$ test set consists of only a single bird with 65 thousand data points. So the amount of variance in $Task1$ data set is lower compared to $Task2$. We see in Fig 1 that as we increase the amount of training data our distance error saturates and standard deviation reduces, which is an intuitive result because as we increase the number of samples variance should decrease.

VIII. Future Work

VIII-A. Clustering

Our usage of clustering has not been very robust. This is because we were trying to solve a very specific problem. But we can generalize this and make the algorithm more robust to different data sets. To achieve this we need to fix two things, first is automatic selection of the number of clusters, we can achieve this if we have domain knowledge, about the maximum deviation of similar bird paths. This will allow us to train clustering algorithm with increasing number of clusters and stop when the maximum standard deviation is reached. Second, we chose image clustering because it was easy for us to evaluate the performance, but this can be replaced by an algorithm, which compromises on interpretability and leverages more important features such as time.

VIII-B. Regression

Our linear regression had two type of biases first was bi-gram assumption and second the hypothesis set. In order to solve these issues we can use a boosting algorithm. The different biased models for boosting could be different linear regression models, which make different biasing assumptions. To do this we can choose 'n' linear models which trains on the difference in values of $T_{current}$-T_n, with varying n. Thus, each models make bi-gram assumptions but between different time steps. In order to choose 'n', we propose to choose a dynamic n rather than a static one. In each iteration of boosting the n depth that our model should go to will depend on acceptance probability 'p'. This technique is called Russian Roulette and has been used successfully in different domains for similar purposes.

VIII-C. LSTM

Model can be further improved to give faster results by merging the models of the Latitude and Longitude predictions. Further improvement can be done on better predicting the longer path by just giving input a single seed point. Various other task specific optimizer can be used, like particle swarm optimization[7] which explicitly make use of animal velocity. Model can be made more robust and flexible to take input even the animal random movement data of short term and make prediction about the long term without

IX. Statement of Contributions

- **Defining the problem:** Done as a team.
- **Developing the Methodology:** Done as a team.
- **Performing the Data Analysis:** Done as a team
- **Coding the Solution:**
  LSTM - Monica Regression - Krtin Clustering - Krtin
- **Writing the Report:**
  Done as a team
  "We hereby state that all the work presented in this report is that of the authors unless otherwise referenced"

References

[1] Flack A, Fiedler W, Blas J, Pokrovski I, Kaatz M, Mitropolsky M, Aghababyan K, Fakriadis Y, Makrigianni E, Jerzak L, Shamina Flack A, Fiedler W, Blas J, Pokrovski I, Kaatz M, Mitropolsky M, Aghababyan K, Fakriadis Y, Makrigianni E, Jerzak L, Azafzaf H, Feltrup-Azafzaf C, Rotics S, Mokotjomela TM, Nathan R, Wikelski M, 2016, Costs of migratory decisions: a comparison across eight white stork populations. Science Advances 2(1): e1500931.

[2] Hochreiter, Sepp, and Jrgen Schmidhuber. "Long short-term memory."Neural computation 9.8 (1997): 1735-1780.

[3] Bluche, T., Kermorvant, C., & Louradour, J. (2015, August). Where to apply dropout in recurrent neural networks for handwriting recognition?. In Document Analysis and Recognition (ICDAR), 2015 13th International Conference on (pp. 681-685). IEEE.

[4] Johnson, D. S., London, J. M., Lea, M. A., & Durban, J. W. (2008). CONTINUOUSTIME CORRELATED RANDOM WALK MODEL FOR ANIMAL TELEMETRY DATA. Ecology, 89(5), 1208-1215.

[5] Rakhimberdiev, E., Winkler, D. W., Bridge, E., Seavy, N. E., Sheldon, D., Piersma, T., & Saveliev, A. (2015). A hidden Markov model for reconstructing animal paths from solar geolocation loggers using templates for light intensity. Movement ecology, 3(1), 1.

[6] Frelich, L. E. (2014), How to Become a Forest Ecologist In Only 40 Years. The Bulletin of the Ecological Society of America, 95: 207210. doi:10.1890/0012-9623-95.3.207
[7] Palangpour, P., Venayagamoorthy, G. K., & Duffy, K. (2006, July). Recurrent neural network based predictions of elephant migration in a South African game reserve. In The 2006 IEEE International Joint Conference on Neural Network Proceedings (pp. 4084-4088). IEEE.
[8] https://www.movebank.org
[9] http://www.ecmwf.int
[10] http://opencv.org
[11] http://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html
[12] http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
[13] https://github.com/jeshaitan/migration-lstm