CS229 Project Report
Detect Leaders in Cow Group Movement using Pairwise Distances

Yang Li

Introduction

Project Background

Studies in animal social behavior show that group-living animals, such as cows, travel together in a collective pattern known as spontaneous group movement [2] Such movement exhibits the leader-follower phenomenon, that is, some individual are more likely than others, to initiate group movement that causes others to follow. We call these cows leaders of the group movement. An useful tool in cattle management is to automatically identify the group movement leaders by analyzing cow motion data. The goal of this project is thus to find out whether we can model the leader probability of individual cows based on the pattern of recent group movements.

We define the leading cows of a group movement to be occupants of the front positions in the direction which the group is moving. Although it is easy to determine a cow's relative position within the group given its accurate geographical location tracked using GPS, it is often infeasible to deploy GPS devices to cattle on a large scale due to cost, power and signal limitation. An alternate way of data collection proposed by students in ETH Zurich University[1], is dispersing the environment with sparse landmarks and equipping the cows with inexpensive sensors that records contact events with other sensors and landmarks[1]. Previous work by Stanford students Daniel Chen and Johnathan Jiang showed that it is possible to estimate the distances between sensors from contact information. Knowing the cow-to-cow distance, we define the *neighbour distance distribution* of Cow i to be the distribution of distances between Cow i and other cows within the group. We can describe this distribution using a Gaussian model, $p(d_i) \sim N(\mu_i, \sigma_i^2)$ where μ_i and σ_i^2 are empirical estimations of mean and variance. We expect leading cows to have very different μ_i and σ_i from non-leading cows. Figure 1 compares the neighbour distance of Cow 2 and Cow 11 at frame 700. In this case, we have $\mu_2 = 8.09,\ \mu_1 1 = 4.80,\ \sigma_2^2 = 21.42$ and $\sigma_{11}^2 = 7.18$. Base on this intuition, this project applies several supervised learning models to estimate a cow's leader probability from its neighbour distance distribution.

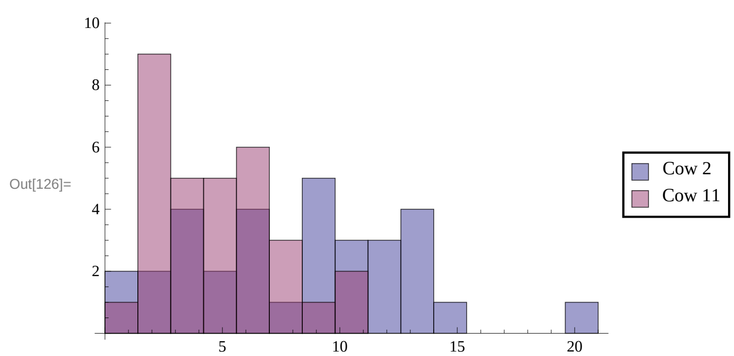

Figure 1: The neighbour distance of a leading cow (left) and a non-leading cow (right)

[1]Data courtesy of Dr. Silvia Santini, Institute for Pervasive Computing, ETH Zurich.

Dataset Description

Our original dataset contains the GPS traces of 34 cows and clustering results from previous work.

GPS Traces. In an experiment conducted by previous researchers, the exact locations of 34 cows were tracked using GPS continuously for 2 days at 2Hz frequency. For the prototyping purpose, we used a filtered dataset sampled at 5 seconds intervals, which contains 36262 frames and have size 30MB.

Clustering Results. In previous research, single linkage clustering were used to partition the cows into clusters in each frame. According to the clustering results, we can see that most cows move in large groups, with individuals leaving and joining a group at random occasions. In most frames, there is one or two large clusters of cows and a few outcasts.

For this project, we primarily looked at the first 733 frames of data, during this time all but 2 cows are within a single cluster, labelled C1. To simulate the senario of working with pure sensor collected data, we limited the training data to pairwise distances of 32 cows at every frame, synthesized from the GPS traces.

Methods

Feature design

Let m be the number of frames that the training data spans. The design matrix for modelling the leader probability of cow i is an m by 2 matrix. Each row vector $x_i^{(t)}$ represents a training example characterizing Cow i's distance and relative speed with respect to its neighbors at time t (t = 1,..m.)

$$x_i^{(t)} = \left(\mu_i^{(t)} \ \sigma_i^{2(t)} \gamma_i^{(t)} \ \tau_i^{2(t)}\right)$$

- $\mu_i^{(t)} \ \sigma_i^{2(t)}$ - mean and variance of Cow i's neighbor distance distribution at time t.
- $\gamma_i^{(t)} \tau_i^{2(t)}$ - mean and variance of the speed correlation of Cow i and other cows over a time window $[t-r, t+r]$, where the positive integer r represents the window radius. The relative speed between Cow i and Cow j at time t is approximated using the forward difference: $s_{i,j}^{(t)} = d(i,j)^{(t)} - d(i,j)^{(t-1)}$.

In the case of logistic regression, $x_i^{(t)}$ is augmented to $(1 \ \mu_i^{(t)} \ \sigma_i^{2(t)})$ for fitting the constant term. Without the loss of generality, we will drop the subscript i in the rest of this section.

In most of the design matrices, the value of $(\sigma^{(t)})^2$ is much larger the value of $\mu^{(t)}$, it is therefore necessary to normalize the feature vectors to have mean 0 and standard deviation 1. For instance, let $\bar{x}_k$ and τ_k be the mean and standard deviation of the kth feature vector, then the normalized feature vector is $x_k := \frac{x_k - \bar{x}_k}{\tau_k}$.

Target labelling

The training target y_i^t is a binary value that indicates whether cow i is at a leading position in frame t. The ground truth labels are generated geometrically using the "first-k algorithm". In this algorithm, each cow is represented by a 2D point. Define $v^{(t)}$, the velocity of the cluster at frame t to be the difference of the cluster mean between frame $t-1$ and frame t. We then ranked all points in the cluster by their relative positions in the direction of the cluster velocity (See Figure 2.) Then the leaders are the k points with highest ranks. Here we chose k = 5 to ensure we have sufficient

”positive” examples for training, while preserving the sensitivity requirement of the problem. In particular, $y_i = 1$ implies Cow i occupies one of the first 5 positions of the group at time t.

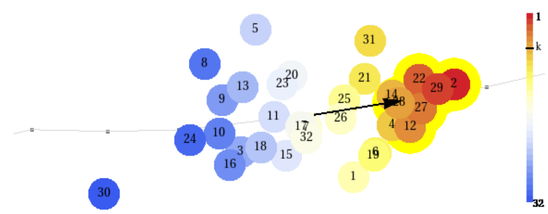

Figure 2: Leader classification of cows in cluster C1 at frame t = 696. *Each cow is represented by a labelled dot. The Gray line is the trajectory traced by the cluster mean, and the black arrow represents to the cluster velocity. The points are colored by position ranking. The top 5 ranked points are highlighted in yellow.*

Logistic Regression

With logistic regression (LR), we modelled the likelihood of a cow being at a leading position given parameter θ using the familiar sigmoid function $p(y|x;\theta) = h(x^i) = \frac{1}{1+e^{-\theta^T x}}$. We proceeded to maximize with respect to θ, the log-likelihood: $l(\theta) = \sum_{i=1}^{m} y^i logh(x^i) + (1-y^i)log(1-h(x^i))$.

Using Newton's method, we obtained the optimal parameter $\theta \in \mathbb{R}^5$. The probability that cow i is at the leading position given a new distance distribution $x = [\mu\ \sigma]$ can be evaluated as $p(y|x;\theta) = h_\theta(x^i)$ at x.

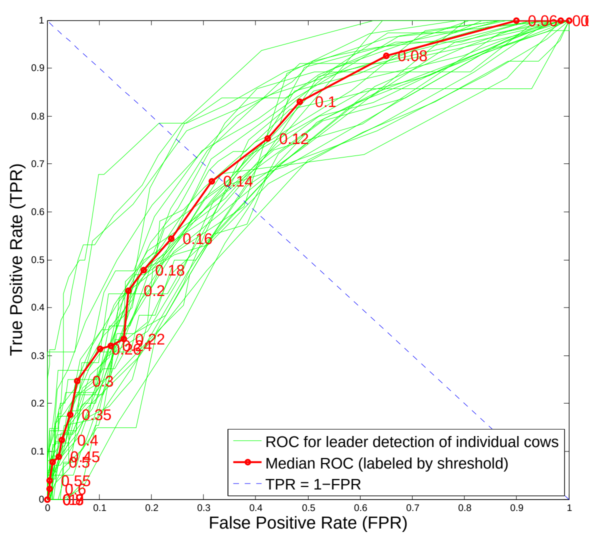

Figure 3: ROC curve for Data source GPS traces of 34 cows in 2 days leader detection using Training data Pairwise distances of 32 cows during LR.

In the leader detection problem, the size of the leader class is much smaller than the size of the non-leader class. This leads to a problem when we use $p(y|x;\theta) > 0.5$ as the leader detection criterion: LR often finds a trivial classifier that labels all cows as non-leaders. We therefore biased the criterion by a threshold parameter T, such that $y_T = \begin{Bmatrix} 1 & p(y|x;\theta) \geq T \\ 0 & p(y|x;\theta) < T \end{Bmatrix}$.

We used ROC curves[3], shown in Figure 3, to determine the best T for all models. The training examples were randomly partitioned into a training set of size 512 (70% of the original training set) and a test set of size 220. For each model, we repeatedly trained and tested our model with T ranging from 0 to 1 at step size 0.02. Then the true positive rate (TPR) was plotted against false positive rate (FPR) over the threshold space, forming a ROC curve. The median ROC curve for all models is highlighted in red.

We approximated the optimal threshold T to be the intersection of the ROC curve with the line $FPR = 1 - TPR$. This can be interpreted as setting the number of false positives equal to the number of false negatives. In our problem, we found $T = 0.14$, which non-surprisingly coincided with the threshold value that maximizes the F1 score[2].

Linear SVM

The second model we attempted is l_2 regularized linear support vector machine (SVM). Regularization is necessary since the models are clearly non-separable. Since our problem is low dimensional and has small scale, we did not need to worry about the additional computation cost of l_2 regularization over l_1. The model fitting was done using the LIBLINEAR package, solving optimization problem $min_{\ w} \ \ \frac{1}{2}w^T w + C \sum_{i=1}^{m} max(0, 1 - y_i w^T x_i)^2$ for optimal margin parameter w. Constant C (with default value 1) is the penalty parameter. The predicted outcome is 1 if $w^T x >= B$, where B is the bias constant.

We used a 2-step grid search method to find the (C, B) pair that maximizes the $F1$ score of the model. In the coarse step, we searched within the range $0 \leq C \leq 3$ and $-3 \leq B \leq 3$ at step size 0.5. We then performed a finer search with step size 0.1. The final parameters we obtained is $C = 0.10$ and $B = 0.41$.

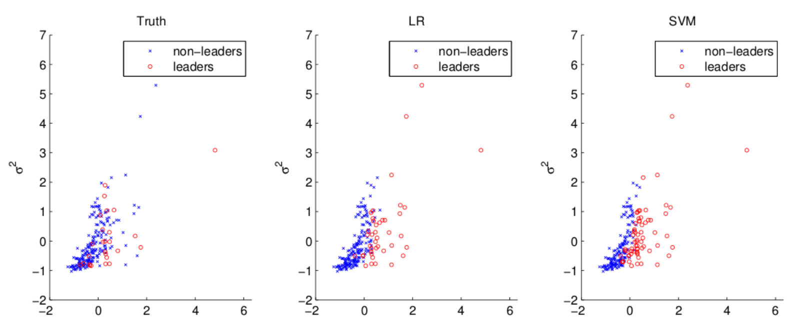

Figure 4: Comparison between true labels, LR predicted labels and SVM predicted labels for Cow 12.

Results

We used 5-fold cross-validation to evaluate the LR and SVM models. In each trail, the number of training examples was 549, and the number of test examples were 183. The mean and variance of

[2]Here we used $F = 0.5 * R * P/(P + R)$

cross validation errors, precisions and recalls over all 64 models are summarized in Table 1. We found no significant difference in the average test errors and F-scores between the two methods.

In addition, we experimented with mixing all training examples together to train a global leader detection model for all cows. The training and testing size became 16416 and 7008, and the model evaluation results are shown in Table 2. The results shows little difference from the average of individual models. Hence increasing training examples did not seem to improve fitness of the model.

~	Test Error		Precision		Recall		F-Score	
	mean	std	mean	std	mean	std	mean	std
LR	0.43	0.19	0.24	0.08	0.72	0.24	0.35	0.10
SVM	0.36	0.04	0.25	0.09	0.70	0.11	0.37	0.08

Table 1: Test Error, Precision and Recall of LR and SVM Models for individual cows

	Test Error	Precision	Recall	F-Score
SVM	0.35	0.26	.71	0.38

Table 2: Test Error, Precision and Recall of SVM Model for all cows

Conclusion

In this project, we tackled the cow leader detection problem through learning the pairwise distance and speed correlation of cows. We designed a feature extraction scheme, fitted our model using logistic regression and linear SVM, and performed cross validation to evaluate the fitted models. It was surprising to see that LR models are on average competitive to SVM models, while both methods were insufficient to detect leaders with high sensitivity under reasonable specificity. One inherited challenge of our problem is the low prior probability of leaders in our training examples. Comparison between the F-scores of individual models with the prior probability in the respective training sets reveals that, detection models trained on data with higher prior probability are better fitted than those trained on data with low prior probability.

In order to deploy leader detection in practice, additional features need to be considered. Future work may incorporate cow-to-landmark distances, a piece of information that we did not employ under the scope of this class project. As a final remark ,this project has been part of a research project that studies motion patterns in the trajectory data of mobile objects. The cow leader detection problem presented here can be generalized to other scenarios that involve the leader and follower phenomenon, such as crowd movement and trends on social network. Machine learning techniques will most likely play important roles in studying these topics.

References

[1] Brano Kusy, Silvia Santini, and Leo Guibas. Trajectory estimation using network traces of mobile devices. Unpublished, 2011.

[2] A. Ramseyer, A. Boissy, B. Thierry, and B. Dumont. Individual and social determinants of spontaneous group movements in cattle and sheep. *animal*, 3(09):1319–1326, 2009.

[3] Wikipedia. Receiver operating characteristic. `http://en.wikipedia.org/wiki/Receiver_operating_characteristic`.