Maritime Anomaly Detection using Gaussian Process Active Learning

Kira Kowalska
Advanced Technology Centre
BAE Systems
Bristol, UK
Email: kira.kowalska@baesystems.com

Leto Peel
Computer Science Department
University College London
London, UK
Email: l.peel@cs.ucl.ac.uk

July 17, 2012

Abstract

A model of normal vessel behaviours is useful for detecting illegal, suspicious, or unsafe behaviour; such as vessel theft, drugs smuggling, people trafficking or poor sailing. This work presents a data-driven non-parametric Bayesian model, based on Gaussian Processes, to model normal shipping behaviour. This model is learned from Automatic Identification System (AIS) data and uses an Active Learning paradigm to select an informative subsample of the data to reduce the computational complexity of training. The resultant model allows a measure of normality to be calculated for each newly-observed transmission according to its velocity given its current latitude and longitude. Using this measure of normality, ships can be identified as potentially anomalous and prioritised for further investigation.

The model performance is assessed by its ability to detect artificially generated AIS anomalies at locations around the United Kingdom. Finally, the model is demonstrated on case studies from artificial and real vessel data to detect anomalies in unusual tracks.

1 Introduction

A wealth of information on vessel behaviour has become available due to the compulsory use of the Automated Identification System (AIS) for most international voyaging and passenger ships. AIS messages are transmitted from vessels reporting their position, speed, heading, and other details, such as their destination and ship identifier. Although the main purpose of the self-reporting AIS system is safety in navigation and collision avoidance, it is also an abundant source of data for maritime surveillance. Surveillance authorities are interested in using this data to uncover threats to security and illegal activities. A lot of current research is aimed at determining the best way to exploit this wealth of AIS data, in order to improve situation awareness in the maritime domain. The main research goal is to be able to identify anomalies.

Detecting anomalous behaviour is important as it can be indicative of nefarious activities, such as piracy, drug smuggling, arms trading, people trafficking and illegal immigration. In some cases, anomalous behaviour could pose a safety risk. The large numbers of vessels on the seas makes inspecting vessel tracks for anomalous behaviour time consuming and error prone for human analysts. Furthermore, with more satellite-based Automatic Identification System (AIS) receivers coming online (as reported in [1]), the amount of data will only increase. Automated tools are essential to reduce the cognitive workload of the analysts.

This requires development of a model representing normal vessel behaviour. The model can be then used to identify anomalous behaviour by vessel's deviation from normality. One of two routes can be taken for automatically detecting anomalous behaviour: a top-down approach, which requires the explicit definition of models a priory; and a bottom-up approach, which builds models of normal behaviour from data. Top-down approaches usually involve the codification of behaviours by Subject Matter Experts (SMEs) based on their experience and domain knowledge. For instance, [1] ran a workshop with SMEs to elicit abnormal behaviours that were codified as rules. [2] represented that expert knowledge as an ontology and developed a system to perform reasoning using Description Logics to classify vessels of interest and detect anomalies. A fuzzy expert system introduced by Jasinevicius and Petrauskas [3] that takes into account the vessel type, persons on board and the riskiness of the cargo when detecting abnormal behaviour. Other examples are Bayesian networks [4, 5], which rely on conditional probabilities as a measure of normality. If probability of an event, such as a vessel's track, falls below a chosen threshold, it is flagged as abnormal. Although Bayesian networks can be easily understood and validated due to their representation of causal relationships, they require characterization a priori of all network variables and edges, which is usually a very challenging task. On top of that, once the network complexity is fixed, its expressive power is limited by the initial network structure, so it does not deal with data that changes over time very well. It is a common issue of top-down approaches.

The fact that there is a large amount of AIS data available, corresponding to routine behaviour, motivates the use of bottom-up techniques for building models of normal behaviour. For example, adaptive kernel density estimation is used in [6] to model normal vessel tracks. Vessels in regions of low density are flagged as anomalies. A clear advantage of the method is that there is no need to specify a model or any parameters a priori. [7] developed a prototype system to detect unsafe, illegal and threatening vessel activity that learns from operator-labelled track reports and their responses to automatically generated alerts. A modified Fuzzy ARTMAP neural network classifier is employed to learn models of vessel behaviour. Longitude, latitude, speed and course are input to the system and the classes employed are normal, anomalous and unknown. [8] detected suspicious or anomalous behaviour by dividing the tracks into "motifs" (such as straight line, u-turn or loop) and examples of anomalous behaviour are used to train the classifier. [9] took a Bayesian network approach where a model of normal vessel behaviour was learnt and anomalies were then detected by the system by comparing the data to the model. Other bottom-up techniques use an unsupervised learning approach, such as work by Rhodes, Bomberger et al. using artificial neural networks [10], or the work by Laxhammar [11], where Gaussian Mixture Model is used as cluster model and a greedy version of the Expectation-Maximization algorithm as clustering algorithm. Although Gaussian Mixture Model allows learning from multimodal data, i.e. more than one normal behaviour in one position, finding an optimal number of mixture components is not trivial and is prone to over-fitting. Moreover, applying Gaussian Mixture Model to all possible locations would be too computationally expensive, so it requires placing a uniform grid over the area of interest and then dealing with each square separately. That imposes an artificial discretization of otherwise continuous positions and might not reflect real data clusters very well. The problem of position discretization also exists in [12], where associative neural networks are used to learn normal vessel behaviour and detect anomalies. Most of the methods previously introduced focus on modelling normal ship "tracks". Although this allows detection of anomalous behaviour embodied by the track as a whole, it cannot be easily used in real-time surveillance. An alternative approach considers ship behaviour independent of time.

In this work, a model of normality is created from historical AIS data using Gaussian Processes (GPs), thus codified expert knowledge is not required. An advantage with GPs is that the model is non parametric so it is not necessary to build in features of anomalous behaviour. A limitation of this approach is that although GPs provide a flexible and robust approach to anomaly detection, they are not typically suitable for large datasets due to high computational complexity in training ($O(n^3)$) and prediction ($O(n)$), where n is the number of samples in the training set.

This limitation is addressed by Active Learning, which allows selection of an optimal training sample from AIS data. The sample accurately represents the entire set but is relatively small. It is an alternative approach to our previous work [13], where computational complexity of GPs was minimized by a Kd-Tree approximation for training and prediction. The two methods could potentially be used together, although it was not done so in this paper.

2 Approach

2.1 Data Preprocessing

AIS data from six weeks starting on January 16th 2011 was used, which provides information on each vessel's type t, position x (latitude, longitude) and velocity y (speed, heading) along the English Coastline. As Gaussian Processes are being used, the behaviours modelled are required to be unimodal, i.e. one normal behaviour in one position. However, different vessel types tend to have different "normal" behaviours. For example, while heading towards a fishing zone is normal for a fishing boat, it should be flagged as anomaly for a tug boat. Another example would be that low speed is normal for a cargo ships but not for speed vessels. Therefore, the first step in the reduction of data modality is the separation of AIS data according to vessel types. As a result, for each vessel type t, speed usually becomes unimodal, heading, however, tends to be at least bimodal. This is because at a given position x, vessels tend to move in two opposite directions along the same shipping lane. Therefore, to create unimodal heading data, we restrict heading values to (0, 180) range:

$$heading = \mathbf{mod}(heading, 180). \tag{1}$$

This approach expects only one shipping lane in one position. If more shipping lanes were to be considered in one position, the method would require data clustering, e.g. a mixture of Gaussian processes [14], but it is outside the scope of this work. At this stage, we deal with unimodal y data with two components: speed, which is a scalar measurement, and heading, which is an angle in the range (0, 180). These two measurements define a vessel's velocity in a Polar coordinate system. A possible approach to model vessel behaviour would be to use two GPs: one to model speed and the other to model heading. However, heading contains a discontinuity, which needs to be carefully considered for a GP regression model. The approach in this work is to convert velocity to Cartesian coordinate system, as shown in Figure 1.

Following the conversion a discontinuity still remains. Specifically, the values of y_j velocity component approach opposite values in the limits of the heading range, which could negatively affect the accuracy of GP regression. Therefore, to resolve the issue, we use absolute values of y_j only.

$$y_j = abs(speed \cos(heading)) \tag{2}$$

As a result of constraints introduced in (1) and (2), we deal with velocities in the first quadrant of the Cartesian plane only (Figure 2). The behaviour data y is now expressed in terms of y_i and y_j velocity components.

2.2 Gaussian Processes

GP regression models are constructed using training data $D = \{y, x\}_{1,\dots,n}$ for the prediction of velocity y^* at a new unseen position x^*. The prediction of y^*

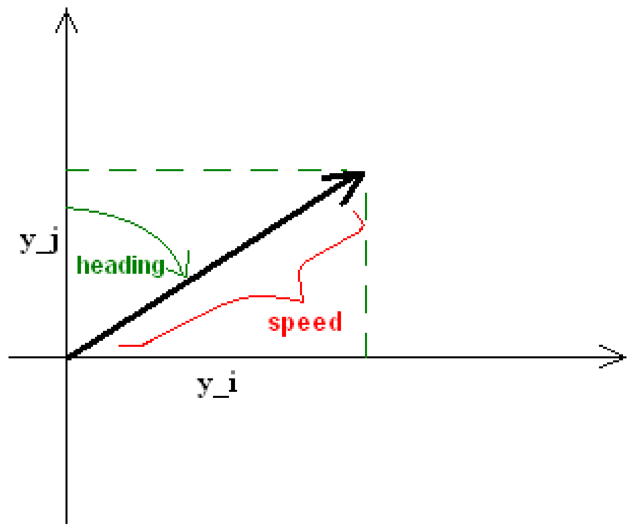

Figure 1: Conversion from polar to Cartesian coordinate system

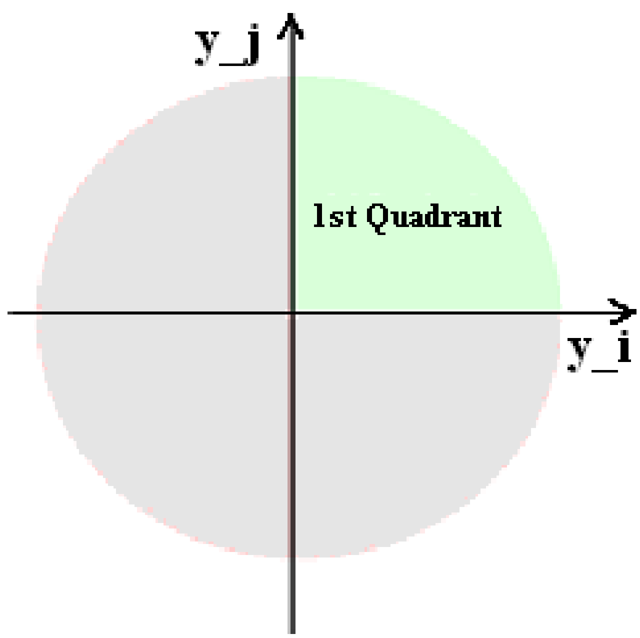

Figure 2: The velocity vectors of the vessels are pre-processed to lie only in the quadrant indicated.

is achieved through the construction of two separate GPs to predict y_i^* and y_j^* velocity components, as described in Section 2.1, for each ship type. There are 16 ship types in the AIS database, therefore, there are 32 GP regression models. After the training, the GP models are capable of predicting y_i^* and y_j^* velocity components for an unseen position x^*. Gaussian Process is specified by its mean function $m(x)$ and covariance function $K(x, x')$. The prior mean is set to zero. Therefore, before the training step, we expect zero velocity everywhere. What relates one observation to another is the covariance function. Our choice is the "squared exponential" stationary covariance function:

$$K(x, x') = \sigma_f^2 \exp\left[\frac{-(x - x')^2}{2l^2}\right] \tag{3}$$

where σ_f^2 is the signal variance and l is the characteristic length scale which tells us how much separation of x and x' affects their covariance. The length scale has the empirical effect of affecting the smoothness of the predictive function.

After applying the covariance function to the training set, velocity in regions with no vessel activity (in the training sample) will be characterized by zero posterior mean (equal to prior) and high variance. Therefore, any vessel behaviour in these regions will be seen as unusual and flagged as an anomaly, irrespective of their velocity. It allows detection of anomalous vessels, which venture outside established shipping tracks. Each reading y^* can be thought of as a noisy output of an underlying function $f(x^*)$ (refered to hereinafter using the shorthand f^*) which is inferred from the posterior distribution:

$$p(f^*|x^*, D) \sim \mathcal{N}(\bar{f}^*, \mathbb{V}(f^*)) \tag{4}$$

where the mean and variance are defined as:

$$\bar{f}^* = k^{*T} M^{-1} \mathbf{y} \tag{5}$$

$$\mathbb{V}(f^*) = K(x^*, x^*) - k^{*T} M^{-1} k^* \tag{6}$$

where

$$M = (K + \sigma_N^2 I), \tag{7}$$

$$k^* = [K(x^*, x_1), ..., K(x^*, x_n)]^T, \tag{8}$$

and σ_N^2 is the noise variance. The value of y^* differs from the underlying value f^* by additive noise. The noise is assumed to be an independently and identically distributed Gaussian distribution with zero mean and variance (also referred to as the likelihood parameter),

$$y = f(x) + N(0, \sigma_N^2) \tag{9}$$

We do not know *a priori* the values of the covariance and likelihood hyperparameters, so they are first initialized to random values and then optimized using conjugate gradient method. To avoid being trapped at poor local minima, multiple initialisations are performed and the hyperparameter set that maximizes

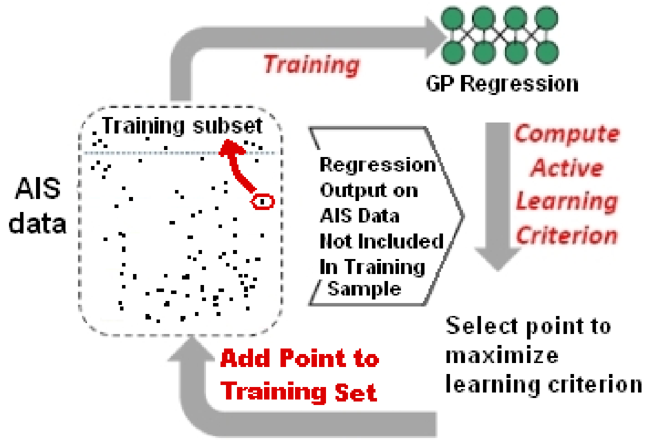

Figure 3: A visual representation of the Active Learning cycle.

the marginal likelihood is chosen. The computational cost of inverting M is reduced by adopting Cholesky decomposition.

2.3 Active Learning

Active learning is used to select a subset of AIS data that would serve as a good representation of the entire set and enable accurate GP regression. It significantly reduces the computational cost of GP regression by decreasing the size of the training dataset from n to m where $m < n$. It is an iterative process. In each iteration, GP training and regression is performed and based on the regression outputs, one point from the available AIS readings is selected to be included in the training set that would improve GP performance in the following iteration. The Active Learning cycle is shown in Figure 3.

The incremental re-training of the GP is made computationally feasible due to Cholesky Factor Update [15]. The method exploits the incremental structure of the problem, and recycles most of the work put into computing the Cholesky decomposition at the last iteration to produce the new Cholesky decomposition. This reduces the complexity of each iteration from $O(m^3)$ to $O(m^2)$. This means that despite having to incrementally train the GP, the computational complexity is still much less than training on the complete training set. In active learning a learning criterion is used to determine the selection order. In this paper, three different learning criteria are compared. The first one is the posterior variance:

$$criterion_1 = var(y^*). \tag{10}$$

Variance gives information about the uncertainty of prediction at each location. The goal is to minimize the uncertainty of prediction, so the location in the set of available points with the maximum posterior variance is selected and moved with the corresponding target value to the training set. The second learning criterion is the distance between the posterior mean and the actual value.

$$criterion_2 = |y^* - \bar{y}^*| \tag{11}$$

This criterion helps select vessel locations, where regression is least accurate. The third learning criterion considers both the posterior mean as well as the posterior variance. Specifically, we select the next point according to:

$$criterion_3 = |y^* - \bar{y}^*| \cdot \sqrt{var(y^*)} \tag{12}$$

The active learning approach employed here could easily be extended to allow an online update of the training set. Initially, the set would be created through the active learning scheme and available AIS data. Later, as new AIS data arrives, the points in the set could be exchanged for new points, if they were more informative for training. This means the size of the training set would remain unchanged. This could be achieved by a principled 'windowing' of data, which removes the least informative data points from the training set [15].

2.4 Anomaly Detection

It is assumed that the data from the AIS dataset represents normal behaviour. Anomalous behaviour is therefore artificially generated in this work by taking an actual vessel information from AIS database and moving the vessel to a new position, i.e. its speed and heading remain unaffected but its position changes. This type of anomaly could be characteristic of AIS spoofing. AIS spoofing is the false reporting of the ship's real AIS transmission and could be an indication of the ship's involvement in illegal activities. Direction of vessel movement is chosen from a uniform random distribution. Distance of movement is expressed in degrees of latitude/longitude and is drawn from a Gaussian distribution with chosen parameters μ_s and σ_s (spoofing parameters). Anomaly is detected based on a local anomaly score, which measures the deviation of the actual observation from the predictive distribution at each vessel position. Two anomaly scores are used in this work. The first score is the squared residual between the actual observation and the mean prediction at the given position. The squared residual does not take predictive uncertainty (variance) into account. The second anomaly score is the predictive log-likelihood and does take into account the variance. The likelihood score is given by:

$$score = \frac{1}{2}\log(2\pi\sigma^{*2}) + \frac{(y^* - \bar{y}^*)^2}{2\sigma^{*2}}, \tag{13}$$

where predictive variance $\sigma^{*2} = var(y^*)$. To formulate a global anomaly score, we compute the sum of scores calculated for each velocity component, that is $scoreglobal = scorey_i + scorey_j$.

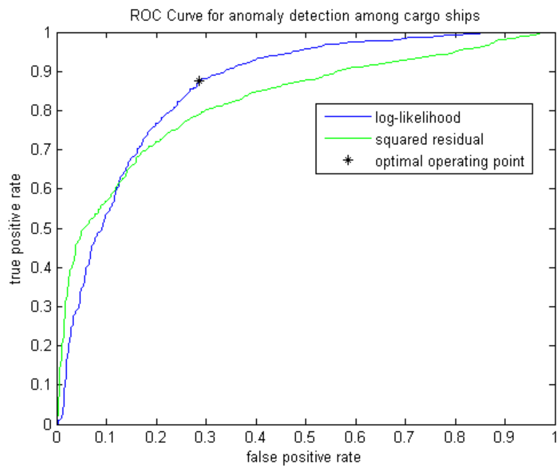

Figure 4: A ROC curve for the classification of anomalous cargo ship behaviours using the GP regression model. The point on the ROC curve representing the optimal threshold (assuming uniform cost of misclassification) is indicted.

3 Results

3.1 GP Regression

2000 training points for a given vessel type were chosen randomly from the AIS database of ships around the UK. A separate test set was created, with 1000 points representing normal behaviour, taken directly from AIS database, and 1000 points that simulate AIS spoofing with spoofing parameters $\mu_s = 4$ and $\sigma_s = 1$. Figure 4 shows accuracy of classification of the unlabelled test set for different threshold values. Probabilistic log-likelihood has the greatest area under the ROC curve, so it is used to establish the optimal threshold value for anomaly detection assuming a uniform cost of misclassification. Figure 5 shows where the cases of misclassifications most often occur. The classification accuracy in Figure 5 is very good in regions where there is no actual AIS training data, i.e. outside of standard shipping lanes. However, in regions where both "normal" data points and anomalies can be found, there are cases of misclassification. It occurs when a data point is moved (during the process of artificially generating anomalies) in any direction by an amount very close to zero. This is because although it is not the true behaviour it is still a normal behaviour for its new position and that is how it is classified by the anomaly detector. For example if a vessel reported that it was further along a shipping lane than it actually was, then this type of anomaly would not be detected. Figure 6 shows how the fraction of misclassified ships change as distance of displacement increases.

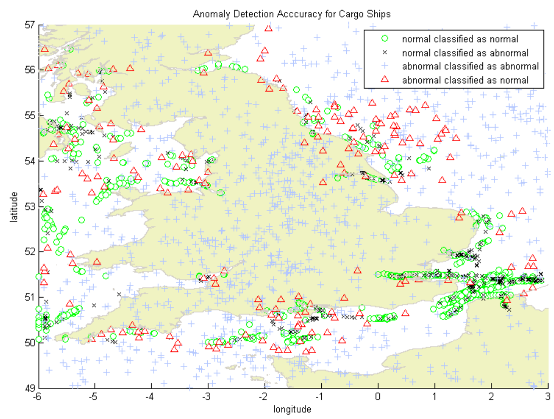

Figure 5: Classification of test points around the UK using the GP model trained on cargo vessels.

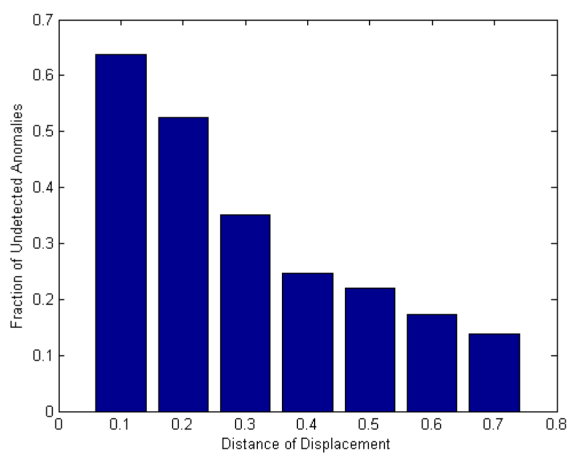

Figure 6: Proportion of simulating AIS spoofing anomalies which are not detected as a function of their displacement. It can be seen that the greater the AIS mesurements are displaced the more likely they are to be identified as anomalies.

3.2 GP regression with active learning

This section describes the results of experiments with different active learning criteria. For the following experiments we choose an initial training set of 50 points and add more points through the Active Learning scheme with different learning criteria. The accuracy of anomaly detection is tested as the size of the training set increases using a labelled validation set. The results are shown in Figure 7. For all the criteria, the accuracy of anomaly detection converges to approximately 80% when all data points are in the training set. When the learning criterion is variance (Criterion1), accuracy of anomaly detection increases steadily and much more rapidly than with the other criteria. Therefore, if we wanted to stop the active learning process earlier to improve efficiency, variance would produce better results. The reason for this is that Criterion2 and Criterion3 take into account the distance between the posterior mean and the actual value. This means that they start by adding the most unusual behaviours to the training set first. This might potentially help for regression as it tries to add the examples which are furthest away from prediction. However, it is less useful for anomaly detection where we are trying to model "normal" behaviour, as adding the most unusual behaviors has the effect of increasing the predicted variance. When the learning criterion is variance, there is no emphasis on collecting data outliers; instead the aim is to cover the input space as well as possible by adding points from the least explored regions first. This also means that the samples are made independantly of the predictive error and therefore gives a more accurate representation of the true data variance. As expected, it is much more effective to select a training set through Active Learning (with Criterion1) than to select the set randomly. It is shown in Figure 8, where the size of the training set was gradually increased until 70% of all data points were in the training set and the accuracy of anomaly detection for a labelled reference set was measured as Area under ROC Curve (AUROC). The plot shows the mean result of 30 independent runs.

As a last remark, we note that the accuracy of anomaly detection increases more rapidly if the size of the available AIS dataset that could be included in the training set is bigger. This results from the fact that the active learning process has more points to choose from and can make more optimal decisions when selecting new points to add to the training set.

3.3 Case Studies

Finally, we present the proposed anomaly detection scheme applied to a couple of case studies, one artificial and the other from real vessel data. The artificial scenario focusses on AIS data near Southend-on-Sea situated by the Thames confluence. The test trajectory for a tanker is shown in Figure 9 as dots, whose colours indicate their anomaly scores. The tanker starts its journey in the river Thames and moves east along the river until it reaches the open sea, then it speeds up and heads north. Once it reaches a shipping lane, it turns southwest and sails back to the Thames confluence. The circles represent real AIS

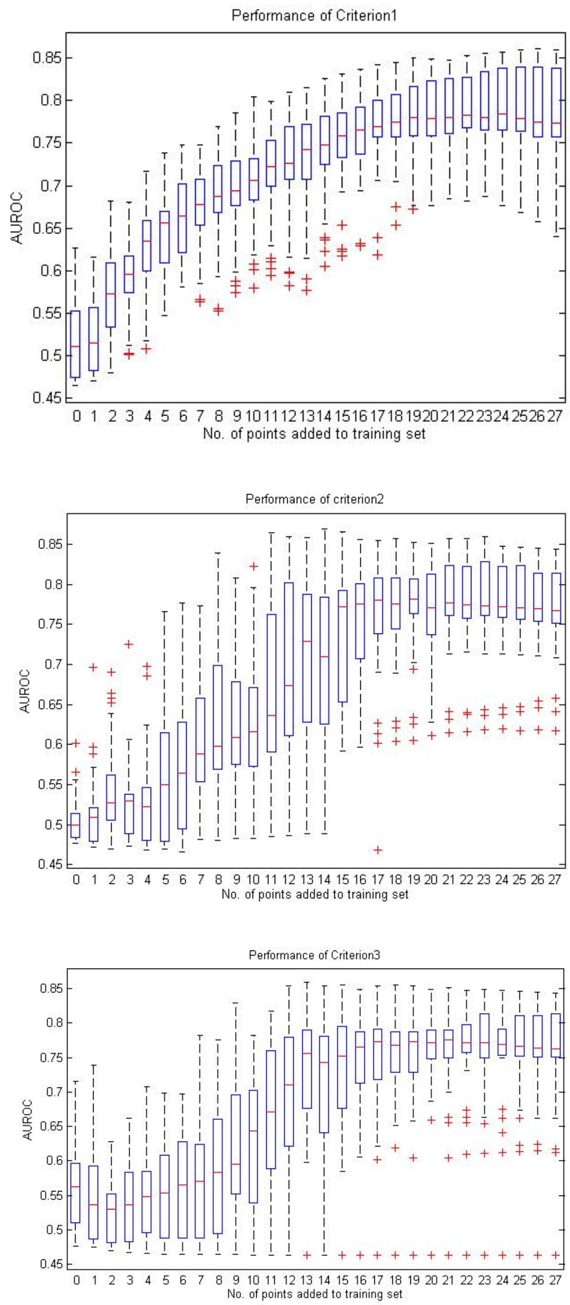

Figure 7: Box plots of accuracy of classification as the size of the training set is iteratively increased using the given criterion (a) variance (criterion 1), (b) residuals (criterion 2), and (c) combined variance and residuals (criterion 3). The results shown are taken over 30 independant runs.

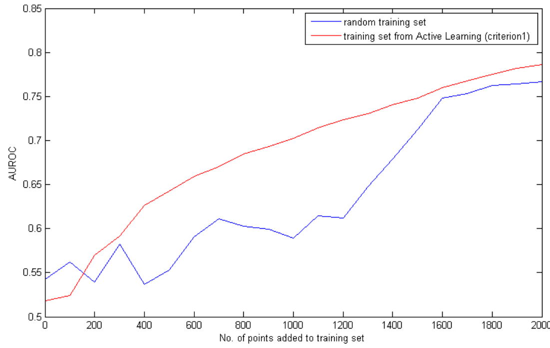

Figure 8: A comparison of mean accuracy of anomaly detection as the training set size is increased using active learning vs. random sampling.

data for vessels moving along the river and the sea nearby. The vessels move along well-established lanes which are vessel type specific. This can be noticed when comparing Figure 9(a), which shows all AIS data, and Figure 9 (b), which shows tankers only. Tankers do not normally move along the shipping lane in the south-west direction towards the river confluence, therefore, such behaviour of the test tanker is detected as anomalous (increased anomaly score) for GP trained on tanker data only, as in Figure 9(b).

The anomaly scores can be used to establish anomaly threshold for the best trade-off between sensitivity and specificity depending on the application.

The second case study was located around Cowes, Isle of Wight using data gathered by BAE Systems for a campaign aimed at developing tracking techniques for difficult maritime targets. The data was collected on 2012-10-11 and 2012-10-12 and the effort was concentrated on gathering Radar and EO sensor measurements of small, fast boats. Most of the targets in the data gathering exercise were targets-of-opportunity. However, to introduce some ground truth, Rigid-Hulled Inflatable Vehicles (RHIBs) were equipped with GPS data loggers and tasked to perform both 'normal' and 'abnormal' activities within the sensors' field-of-regard. The GPS data loggers give position, heading and speed, as well as identity, with 10 second time resolution. It was this GPS data from the RHIBs that was used in this case study as the target tracking problem is beyond the scope of this work. However, with an appropriate tracking solution, this case study demonstrates that our approach would be applicable to this type of data. The normal data to train the GP model was taken from North-East/South-West Channel runs from both days of the campaign and is indicated in Figure 10 by circles. Three tracks of 'unusual' behaviour was used as test data. The three test tracks correspond to:

1. (Figure 10a) A vessel heading north approaching a departing AIS-enabled vessel and loitering in its wake. Initially, it is moving very slowly waiting for the AIS-enabled vessel to depart, then it speeds up trying to approach

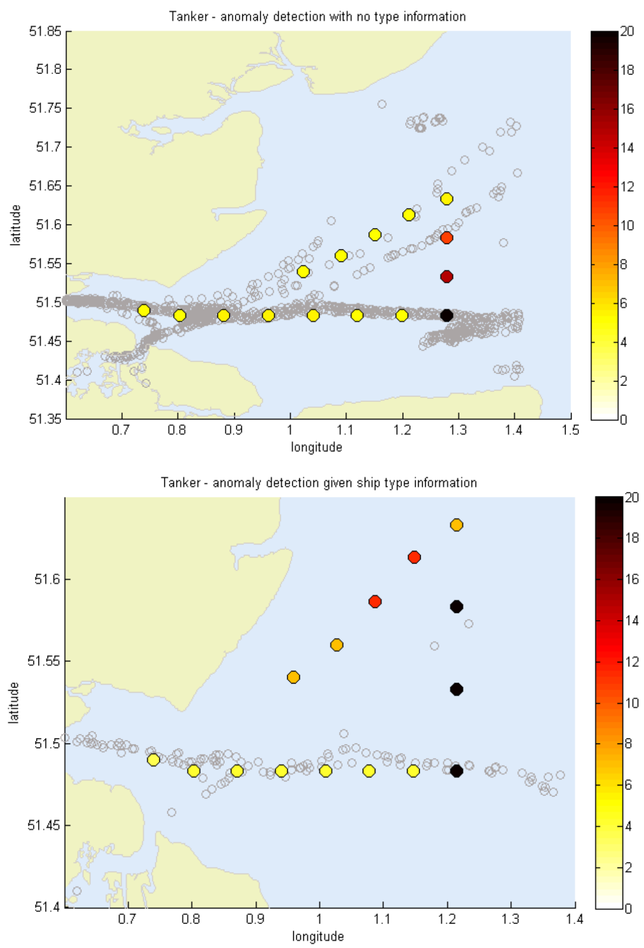

Figure 9: Anomaly detection using training data comprised of (a) all AIS data, and (b) AIS data for tankers only, near Southend-on-Sea. The empty cirlces represent the locations of the vessels used for training while the filled circles represent the generated test points; the corresponding fill colour indicated the level of abnormality as detected by the GP.

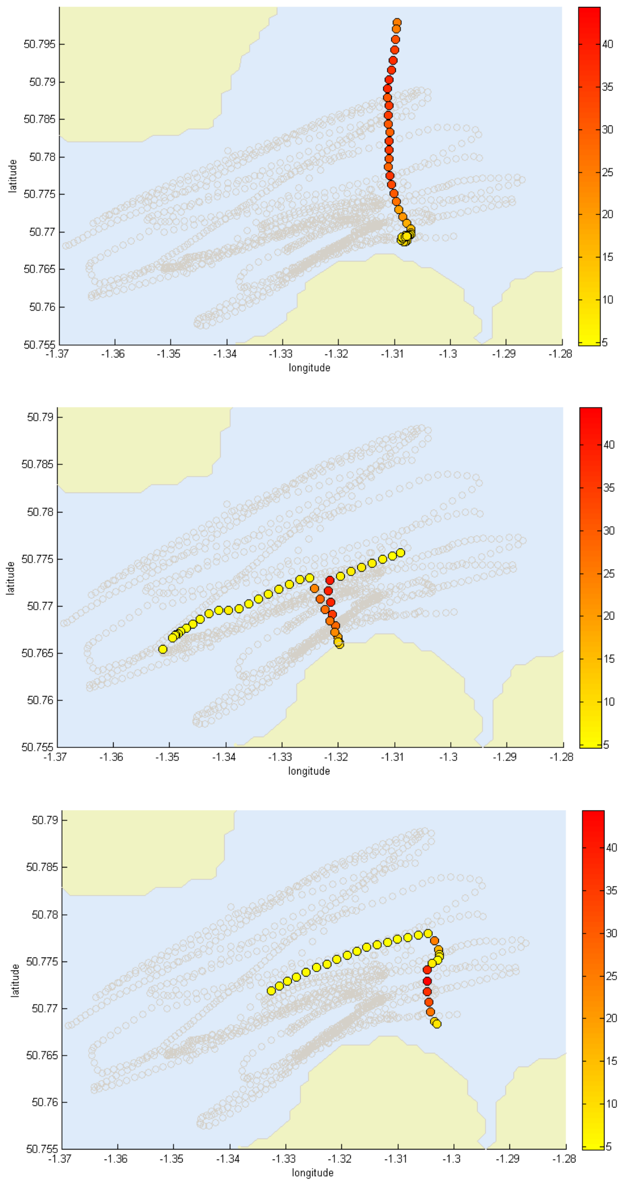

Figure 10: Detecting anomalies in ‘unusual’ tracks around the Isle of Wight representing different behaviours (a) drug smuggling, (b) people smuggling, and (c) terrorism

the vessel. This behaviour represents drug smuggling where the AIS-enabled vessel is some cargo ship or other big vessel that moves slowly and drops some packages into the sea. The packages are then collected by the vessel following behind.

2. (Figure 10b) A vessel that during a NE/SW Channel run breaks off to head to shore, before returning to it's original path. This behaviour indicates people smuggling, i.e. a vessel follows a normal track, then speeds to the shore to pick up some people and quickly returns to its original path.

3. (Figure 10c) A vessel that heads at high speed towards a car ferry, then goes around the ferry and heads south. This is a terrorist scenario. The vessel follows a normal track, then starts speeding towards the ferry to drop some explosives. After dropping the explosives, it speeds towards the shore.

The results of these 3 scenarios are shown in Figure 10, where the colour of the test points represents the level of abnormality detected using the model. It can be seen that the model picks out the most unusual parts of the track. This GP approach to anomaly detection only considers the behaviour of a vessel at a given instant. In order to apply this model to detect anomalous tracks the scores can be aggregated for all the points in the track by taking the sum of the log likelihood. This approach was applied to this case study using 22 of the normal tracks along with 22 'abnormal' tracks which were variations on the scenarios descrbed above. The results of this experiment are given in Figure 11 in the form of a ROC curve, where the area under the ROC curve is 0.8678.

4 Conclusion

This paper has demonstrated Gaussian Processes being used for maritime anomaly detection. Combined with Active Learning, to enable the selection of a reduced training set, Gaussian Processes can be applied to large AIS datasets for accurate anomaly detection. Its accuracy is further increased by the proposed data pre-processing stage. Further work could focus on extending these techniques for online anomaly detection, where models of normality are updated as new data arrives.

Acknowledgment

This work was funded by BAE Systems Advanced Technology Centre.

References

[1] M. Davenport, "Maritime anomaly detection workshop report and analysis," DRDC-Valcartier, Tech. Rep., 2008.

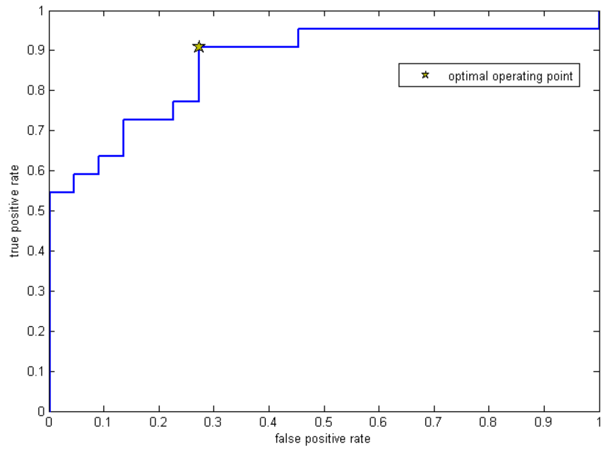

Figure 11: A ROC curve for the classification of anomalous tracks using aggregated scores from the GP regression model. The point on the ROC curve representing the optimal threshold (assuming uniform cost of misclassification) is indicted.

[2] J. Roy and M. Davenport, "Exploitation of maritime domain ontologies for anomaly detection and threat analysis," in *nternational Waterside Security Conference (WSS)*, 2010, pp. 1–8.

[3] R. Jasinevicius and V. Petrauskas, "Fuzzy expert maps for risk management systems," in *US/EU-Baltic International Symposium*, 2008.

[4] S. Mascaro, A. Nicholson, and K. Korb., "Anomaly detection in vessel tracks using bayesian networks," in *8th Bayesian Modelling Applications Workshop*, 2011.

[5] R. Lane, D. Nevell, S. Hayward, and T. Beaney, "Maritime anomaly detection and threat assessment," in *13th International Conference on Information Fusion*, 2010.

[6] B. Ristic, B. L. Scala, M. Morelande, and N. Gordon, "Statistical analysis of motion patterns in ais data: Anomaly detection and motion prediction," in *11th International Conference on Information Fusion*, 2008.

[7] B. Rhodes, N. Bomberger, M. Seibert, and A. Waxman, "Seecoast: Automated port scene understanding facilitated by normalcy learning," in *IEEE Military Communications Conference*, 2006, pp. 1–7.

[8] X. Li, J. Han, and S. Kim, "Motion-alert: automatic anomaly detection in massive moving objects," in *IEEE Intelligence and Security Informatics Conference*, 2006.

[9] F. Johansson and G. Falkman, "Detection of vessel anomalies - a bayesian network approach," in *3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, 2007.

[10] B. Rhodes, N. Bomberger, M. Seibert, and A. Waxman, "Maritime situation monitoring and awareness using learning mechanisms," in *Military Communications Conference*, 2005.

[11] R. Laxhammar, "Anomaly detection for sea surveillance," in *11th International Conference on Information Fusion*, 2008.

[12] B. Rhodes, N. Bomberger, and M. Zandipour, "Probabilistic associative learning of vessel motion patterns at multiple spatial scales for maritime situation awareness," in *10th International Conference on Information Fusion*, 2007.

[13] J. Will, L. Peel, and C. Claxton, "Fast maritime anomaly detection using kd-tree gaussian processes," in *2nd IMA Maths in Defence Conference*, 2011.

[14] C. E. Rasmussen and Z. Ghahramani, "Infinite mixtures of Gaussian process experts," in *In Advances in Neural Information Processing Systems 14*, 2002, pp. 881–888.

[15] M. A. Osborne, A. Rogers, S. J. Roberts, S. D. Ramchurn, and N. Jennings, "Bayesian gaussian process models for multi-sensor time-series prediction," in *Inference and Learning in Dynamic Models.* Cambridge University Press, 2010.