
Bayesian Predictive Modelling: Application to Aircraft Short-Term Conflict Alert System

V. Schetinin

Computer Science Dept.
University of Bedfordshire
Luton, LU1 3JU, UK

L. Jakaite

Computer Science Dept.
University of Bedfordshire
Luton, LU1 3JU, UK

W. Krzanowski

College of Engineering,
Mathematics and Physical Sciences
University of Exeter
Exeter, EX4 4QF, UK

Abstract

Bayesian Model Averaging (BMA), computationally feasible using Markov Chain Monte Carlo (MCMC), is a well-known method for reliable estimation of predictive distributions. The use of decision tree (DT) models for the averaging enables experts not only to estimate a predictive posterior but also to interpret models of interest and estimate the importance of predictor factors that are assumed to contribute to the prediction. The MCMC method generates parameters of DT models in order to explore their posterior distributions and to draw samples from the models. However, these samples can often over-represent DT models of an excessive size, which in cases of real-world applications affects the results of BMA. When this happens, it is unlikely for a DT model that provides Maximum a Posteriori probability to explain the observed data with high accuracy. We propose a new technology in order to estimate and interpret predictive posteriors. In our experiments with aircraft short-term conflict alerts, we show how this technology can be used for analysing uncertainties in detections of conflicts.

1 INTRODUCTION

In many cases of engineering applications, such as air-traffic control, estimation of uncertainty in predictions is of crucial importance, e.g. (Majeske, 2012; Ayuso, 2012). For such applications, the methodology of Bayesian Model Averaging (BMA) has been shown to provide the most accurate estimates of uncertainty. The BMA methodology has been made computationally feasible with the use of Markov Chain Monte Carlo (MCMC) approximation, e.g. (Green, 1995; Robert, 2009).

The use of decision trees (DTs) models within BMA is preferable for applications when experts aim to interpret

models of probabilistic inference and evaluate factors that cause uncertainty in predictions. DTs are hierarchical structures of splitting and terminal nodes that recursively split data. The size of a DT model is determined by the number of its terminal nodes (Chipman, 1998; Denison, 2002).

There are two phases during MCMC approximation. At the first, so-called burn-in, phase the MCMC generates the parameters of a DT in order to explore areas of its maximal likelihood on the given set of observed data. At the second, so-called post burn-in phase, samples of a DT model are collected for averaging according to the Bayesian methodology. It has also been shown that the most accurate results of BMA are achieved when prior information on DT models is available for the MCMC approximation (Chipman, 1998; Denison, 2002).

For interpretation purposes, a single DT which provides the Maximum a Posteriori probability (MAP) could be selected from a set of DT models that were accepted during the post burn-in phase (Domingos, 1998). The other approach to finding a single explanatory model is based on the idea of clustering DT models in a two-dimensional space that is represented by size and fitness of DT models (Chipman, 1998).

According to the Bayesian methodology, samples collected during the post burn-in phase have to be diverse in order to achieve the best accuracy of approximation of predictive density. However, in practice the desired diversity of DT models cannot be achieved in reasonable computing time when prior information on the models is absent or incomplete (Domingos, 2000; Denison, 2002).

Possible reasons of this are as follows. First, the likelihood distribution could be multimodal, which limits MCMC in exploring the full posterior distribution (Robert, 2009). Second, MCMC is limited in exploring all possible DT structures because of the hierarchical structure of DT models (Denison, 2002). A side effect of this results in sampling DT models that contain an excessive number of nodes. Consequently, the ensemble of DT models collected

during the MCMC sampling, as well as any single DT model that is selected for interpretation of the ensemble, will underperform. To mitigate the negative effect, a technique has been suggested for selecting a single DT model which has been tested in a clinical application (Schetinin, 2007).

In this paper, we explore the potential of the Bayesian approach for an air-traffic control problem known as Short-Term Conflict Alert (STCA) detection, where it is critically important to analyse uncertainty intervals in the detection of conflicts. The approach is verified on real data that have been made available by the UK National Air Traffic Services, (NATS, 2002). First we show that Bayesian modelling of the STCA system can explain 89% of decisions that the STCA system has made on these data. We demonstrate that the Bayesian approach allows us to estimate uncertainty in detection of conflicts, which is necessary for specifying possible areas of improvement of the STCA system. The use of DT models allows us to estimate the importance of predictor variables in terms of their contributions to the conflict detection. Finally we show how DT models can be used to find conditions under which the STCA system makes false detections. To achieve this goal we propose a technique for selecting a single DT model from an ensemble of models collected during BMA.

The rest of the paper is organized as follows. Section 2 introduces the STCA problem and describes the data that are used in our experiments. Section 3 briefly introduces the methodology of BMA and MCMC approximation with DT models. The details of the proposed technique and experiments are described in Sections 4 and 5. Finally Section 6 concludes the paper.

2 PROBLEM OF SHORT-TERM CONFLICT ALERT

STCA systems are used in airports to warn dispatchers when the distance between two aircraft, landing or taking off, is critically short in a given alert zone (Prandini, 2000; Brooker 2005). The STCA system is therefore expected to detect conflicts as accurately as possible in the presence of uncertainty in the data that are provided by the airport operation service. In this context, it is of crucial importance to estimate predictive posterior probability distributions of decisions made by the STCA system. The availability of a model that can accurately model the detection of conflicts will allow experts to analyse factors of uncertainty in the detection of conflicts.

The primary information about aircraft movements comes from airport radar. Fig. 1 shows the traces of two aircraft in the 3-dimension system of coordinates X , Y , and Z . The first two coordinates define the position of an aircraft on the X - Y lateral plane with a scale factor, s , that is deter-

mined by the airport radar. Their negative values specify the radar position on the lateral plane. The third coordinate Z is height in feet. The alert cycles here are marked by the filled (red) circle, while the normal cycles are shown by the unfilled circles. In the lateral plane X - Y , the aircraft start their flights at positions indicated here by 1 and 2.

This figure shows that after the 18th radar cycle the system detects a series of 5 alarm cycles during which both aircraft pilots, being warned by the operator, attempt to resolve the conflict. The distance between the aircraft critically decreases from 2100 to 1200. The following 5 cycles are false negative errors as the distance keeps decreasing to 900, and the system is expected to continue detecting the alarm. The system triggers the alarm only at the 28th cycle when the distance decreases to a minimum of 500. In this case the series of 5 false negative error cycles cannot be explained without analysis of factors of uncertainty.

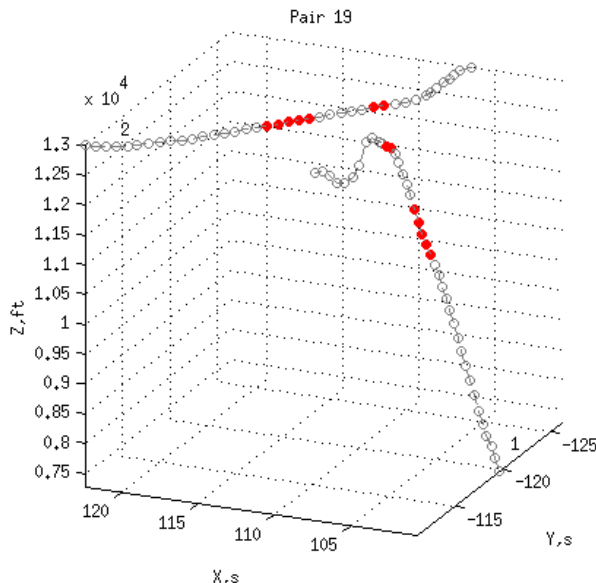


Figure 1: Example of Conflict Prediction Problem

In this paper we aim to model the STCA system in order to find possible solutions to the problem. For the modelling we use primary data about aircraft positions and velocities, which are received by the system as part of the flight information. All flight information is updated each radar cycle, in our case every 6 seconds.

In our research we use these data as follows. The positions are used for calculating the distances d_x , d_y , and d_z between aircraft 1 and 2 along axes X , Y , and height axis Z , respectively. Velocities $V_{x,*}$, $V_{y,*}$ and $V_{z,*}$ of the aircraft are given on axes X , Y , and Z . We assume that distance between aircraft 1 and 2 is important information for detecting conflicts in the airport environment when aircraft change positions in X , Y and Z during landing or

taking off. For this reason distance d is calculated in a 3-dimensional space as $d = \sqrt{d_x^2 + d_y^2 + d_z^2}$. We assign here a scale factor $s = 1\text{ft}$, as s has not been specified for the flight data available for our research. The secondary information about times T_1 and T_2 in the lateral plane for the aircraft 1 and 2 could be also taken into account.

The above assumptions allow us to generate the 12 input variables listed in Table 1. Here, negative values reflect the positions of aircraft in the radar coordinate system.

In our research we use operational data about traces of aircraft pairs. A trace is represented by a sequence of radar cycles as described above. Each cycle in the sequence represents the aircraft movements and is labelled as normal or alert. We aim to use these data for modelling the STCA system within the Bayesian framework in order to quantitatively estimate the uncertainty in detection of conflicts. We assume that this uncertainty is dependent first on the flight parameters, such as aircraft distances and velocities, and second on the accuracy of the radar data. The use of DT models will allow us first to estimate the importance of predictor variables and second to specify conditions under which the system makes false decisions of conflicts. For interpretation of the results of BMA, we will finally select a single DT model in order to find new insights into false detections.

Table 1: Predictor Variables

VARIABLE	NOTATION	MIN	MAX
x_1	d_x	-48.11	51.70
x_2	d_y	-52.93	35.78
x_3	d_z	-10297	8760
x_4	d	2.40	10297.06
x_5	$V_{x,1}$	-691	584
x_6	$V_{y,1}$	-806	473
x_7	$V_{z,1}$	-83.10	96.66
x_8	$V_{x,2}$	-444	599
x_9	$V_{y,2}$	-527	426
x_{10}	$V_{z,2}$	-95.05	11.53
x_{11}	T_1	0	9
x_{12}	T_2	0	9

3 BAYESIAN MODEL AVERAGING OVER DECISION TREES

DTs are known as hierarchical models consisting of splitting and terminal nodes. The DT models are said to be binary if the splitting nodes divide data points into two disjoint subsets. The terminal node assigns a data point to one of the possible classes, the probability of which is dominant (Breiman, 1984). This section is mainly focused on details

of MCMC implementation of BMA over DT models.

3.1 MCMC IMPLEMENTATION

Except for trivial cases the Bayesian methodology of averaging over DTs can be feasibly implemented with MCMC approximation. For the approximation, the parameters, θ , of a DT candidate are drawn from the given proposal distributions. A candidate is accepted or rejected according to the Bayes rule calculated on the given data \mathbf{D} . For the m -dimensional input vector x , data \mathbf{D} and parameters θ , the predictive posterior distribution $p(y|x, \mathbf{D})$, $y \in \{1, \dots, C\}$, is

$$p(y|x, \mathbf{D}) = \int p(y|x, \theta, \mathbf{D})p(\theta|\mathbf{D})d\theta \quad (1)$$

$$\approx \frac{1}{N} \sum_{i=1}^N p(y|x, \theta^{(i)}, \mathbf{D}),$$

where $p(y|x, \theta, \mathbf{D})$ is the posterior distribution given a model with parameters θ and data \mathbf{D} ; $p(\theta|\mathbf{D})$ is the posterior distribution of parameters θ conditioned on data \mathbf{D} ; N is the number of samples taken from the posterior distribution, and C is the number of classes.

In practice, DT models are learnt from data and so their dimensionality (or number of nodes) is variable. The Reversible Jump (RJ) extension of MCMC makes possible the approximation over such models (Green, 1995). Given priors and a sufficient number of samples, the RJ MCMC technique explores the posterior distribution and takes samples of model parameters.

The exploration of DT models of variable size has been efficiently made by using the following moves (Denison, 2002):

Birth moves randomly split the data points falling in one of the terminal nodes by a new splitting node with a variable and rule drawn from the corresponding priors.

Death moves randomly pick a splitting node with two terminal nodes and assign it as a single terminal with the united data points.

Change-split moves randomly pick a splitting node and assign it a new splitting variable and rule drawn from the corresponding priors.

Change-rule moves randomly pick a splitting node and assign it a new rule drawn from a given prior.

The first two moves lead to a change in the dimensionality of parameters. The other moves explore the distribution within the current dimensionality. In particular, the change-split move makes “large” jumps which potentially increase the chance of sampling from a maximal posterior. By contrast, the change-rule move makes “local” jumps in order to explore the details of an area of interest.

As the birth and death moves change the dimensionality, the Bayesian rule includes a ratio R to achieve the condition for reversibility of Markov Chain. For the birth moves, R is written as follows:

$$R = \frac{q(\theta|\theta')p(\theta')}{q(\theta'\theta)p(\theta)}, \quad (2)$$

where $q(\theta|\theta')$ and $q(\theta'\theta)$ are the proposal distributions, θ' and θ are $(k+1)$ and k -dimensional vectors of DT parameters, respectively, and $p(\theta)$ and $p(\theta')$ are the probabilities of the DT with parameters θ and θ' , respectively.

The above probability $p(\theta)$ is defined by a DT structure as follows (Denison, 2002):

$$p(\theta) = \left(\prod_{i=1}^{k-1} \frac{1}{N(s_i^{var})} \frac{1}{m} \right) \frac{k}{S_k} \frac{1}{K}, \quad (3)$$

where $N(s_i^{var})$ is the number of possible values of s_i^{var} that can be assigned as a new splitting rule, S_k is the number of possible structures of a DT with k terminal nodes, and K is the maximal number of terminal nodes.

The proposal distribution is defined as follows:

$$q(\theta|\theta') = \frac{d_{k+1}}{D_{Q1}}, \quad (4)$$

where $D_{Q1} = D_Q + 1$ is the number of splitting nodes whose both branches are terminal nodes.

The MCMC sampler will accept birth and death moves with rates R_b and R_d as follows:

$$R_b = \frac{d_{k+1}}{b_k} \frac{k}{D_{Q1}} \frac{S_k}{S_{k+1}}, \quad (5)$$

$$R_d = \frac{b_k}{d_{k-1}} \frac{D_Q}{(k-1)} \frac{S_k}{S_{k-1}}. \quad (6)$$

If the prior on the number of splitting nodes is given properly, most samples are expected to be drawn from the posterior that is related to areas of interests. If such a prior is unavailable, a DT model will grow excessively and most of the samples will be drawn from posterior distributions that are calculated for oversized DT models. As a result, the estimates of the predictive distribution will be biased (Denison, 2002).

3.2 SWEEPING STRATEGY OF MCMC

In practice, priors on DT structures are often unavailable, and the MCMC sampler cannot efficiently control DT structures, which leads to poor mixing. However, the DT structure can be better controlled with a sweeping strategy of the MCMC approximation as proposed in (Schetinin, 2007). The main idea behind this strategy is to assign the prior probability of splitting DT nodes dependent on the

range of values within which the size of a new data partition will exceed $2p_{min}$, where p_{min} is the minimal number of data points allowed in a partition. This prior is adapted to the range of a data partition. The new splitting threshold $q_{j'}$ proposed for variable j and partition i is drawn from a uniform distribution: $q_{j'} \sim U(x_{min}^{i,j}, x_{max}^{i,j})$.

When the change move is applied to a node that is close to the DT root, distributions of data points in its terminal nodes can be greatly changed, and one or more terminal nodes can contain fewer data points than p_{min} . If there is one such node, this node is swept from the DT and the move is counted as a death move. In cases when there is more than one such node, the move is deemed unavailable.

4 SELECTION OF A SINGLE DECISION TREE MODEL

As discussed in the Introduction, experts need to interpret an ensemble of DT models collected during MCMC sampling as a single DT. Although such a model will likely explain the observed data less accurately, experts will have an opportunity to look at new insights into data. For selection of a single DT model from an ensemble, the MAP and the Maximum a Posterior Weight (MAPW) techniques have been proposed as described in (Domingos, 1998; Chipman, 1998). A drawback of these techniques is that a DT model can be selected from any oversized DT models which are present in the ensemble and as a result this model will under-perform. The idea of a new approach is based on quantitative estimates of classification confidence as described in (Krzanowski et al, 2006). Classifiers that were included in the ensemble produce different outputs for a given input, and each of them is considered as having voted for positive or negative output. The counts over all votes will therefore reflect the difficulty (or confidence) of assigning a given input to a class of interest.

Within this approach, we can define an ensemble of N DT models and then count the number N_i of the classifiers that assign a given input to classes i , $i = 1, \dots, C$. Therefore for a given class i and a given input, the consistency of the ensemble is calculated as a ratio $\gamma = \frac{N_i}{N}$. Its value has a maximum of 1.0, when all the classifiers assign a given input to one class. The minimum value of confidence is $1/C$, when the classifiers assign the input to the all C classes with an equal probability. So for a given input the classification confidence of the ensemble is estimated by the ratio γ whose value is proportional to the accuracy of classification.

We can then define a threshold confidence ratio $\gamma_0 : 1/C \leq \gamma_0 < 1$, for which the cost of misclassifications is considered acceptably small on the given labelled data. The outcome of the ensemble is said to be confident if $\gamma \geq \gamma_0$.

Having counted the number of confident and correct out-

comes on the observed labelled data set, we can select a single DT that covers the maximal number of the labelled data instances that were classified as confident and correct while the number of misclassifications of the remaining examples is kept minimal. Then the DTs with a maximal coverage are selected from the ensemble, and finally a single DT model that has a minimal number of splitting nodes is chosen.

The main steps of the selection technique are as follows.

1. Given an ensemble of DT models, select a set of DT models, $S1$, that cover a maximal number of the data instances classified as confident and correct with a given confidence level γ_0 .
2. Find the instances that were correctly classified by the DT ensemble and denote these instances as $D1$.
3. Among the set $S1$ find DT models that provide a minimal misclassification rate on the data $D1$. Denote the found set as $S2$.
4. Among the set $S2$, find DT models whose size is minimal. A set of such DT models, $S3$, includes at least one DT model.
5. Randomly select a DT model from the set $S3$.

The above procedure finds a single DT model of interest that covers a maximal number of the data instances classified as confident and correct with a given confident level γ_0 . The resultant model is selected to be of minimal size, which reduces the risk of overfitting unlike existing techniques.

5 EXPERIMENTS

In this section we describe experimental results obtained with the proposed BMA technology on real STCA data. First we show that using the BMA technology we can achieve an accuracy of modelling the STCA system around 89%. Second we estimate the importance of predictor variables that are used for modelling the system. Third we demonstrate the proposed technique for selecting a single DT model that is required for interpenetration and finding conditions under which the STCA system can improve accuracy of detections. Finally we show an example of estimating uncertainties in detection of conflicts, which allows us to demonstrate the ability of the proposed technology to identify areas of possible improvement of the STCA system.

5.1 STCA DATA

In our experiments we used 2,526 radar cycles that represent traces of 66 aircraft pairs that were landing or taking

off at the Heathrow, June 1998. The traces were selected with high alertness. The number of cycles in a trace was dependent on the aircraft velocities, and their average number was around 40. Each trace was split into two parts, training and testing, to evaluate the performance within the repeated random sub-sampling validation over 5 runs.

5.2 MCMC IMPLEMENTATION

The BMA was run with a uniform prior on DT models as there was no information about possible DT structures. The minimal number p_{min} was set equal to 5. The proposal probabilities for the death, birth, change-split and change- rules were set to 0.1, 0.1, 0.2, and 0.6, respectively. The numbers of burn-in and post burn-in samples were set to 100,000 and 10,000, respectively. The sampling rate was set equal to 7. The proposal variance was set to 4.0 in order to achieve an acceptance rate of updating the Markov chain around 0.52, which indicates an efficient MCMC implementation. With these settings, the BMA performance within the random sub-sampling validation is $88.6 \pm 1.3\%$.

Fig. 2 depicts samples of log likelihood values (upper plots), the numbers of DT nodes (middle plots) and the distributions of DT nodes for the burn-in (left) and post burn-in (right side) phases.

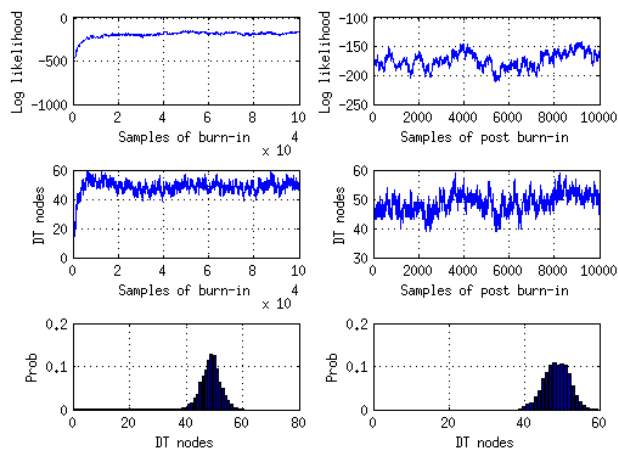


Figure 2: Samples of Log Likelihood and DT Sizes

We can see that in the burn-in phase the Markov chain started with log likelihood value around 1000 converges to the stationary value that oscillates around 175. In the post burn-in phase the log likelihood continues to oscillate between 200 and 150. The lower plots show that the average number of DT nodes was around 46.

5.3 FEATURE IMPORTANCE

During the post burn-in phase DT parameters are changed within the given priors on the proposal distribution, and as a result the accepted DT models include different predictor variables. The frequencies of use of these variables reflect the information about their importance - we assume that a variable with a greater frequency makes more important contribution to the classification.

The frequencies were calculated within the random sub-sampling validation and are shown for all 12 variables in Fig. 3. Table 2 lists these variables in the order of their importance. In this table we see that the three most important features are x_8 , the speed of the second aircraft on the X -axis, x_1 , the distance between aircraft pair on the X -axis, and x_9 , the speed of the second aircraft on the Y -axis.

By contrast, the variables x_{11} and x_{12} , which give the times T_1 and T_2 since the last correlated plot in the lateral plane for the aircraft, are used with a much lower frequency, and we conclude that they make the smallest contribution.

Table 2: Importance of Variables in Terms of Frequency of Use

VARIABLE	FREQUENCY
x_8	0.168
x_1	0.137
x_9	0.120
x_6	0.110
x_4	0.095
x_5	0.090
x_3	0.078
x_2	0.061
x_{10}	0.050
x_7	0.042
x_{11}	0.001
x_{12}	0.008

5.4 SINGLE DECISION TREE MODELS

The total number of DT models that were collected during the MCMC post burn-in phase was 10,000. In theory, Bayesian averaging over an ensemble of models should outperform any single model that is taken for interpretation purposes. In our case, we expect to find a single DT model whose performance is maximally close to that obtained with the ensemble average. Such a DT model is required for interpretation purposes and for specifying conditions under which the STCA system makes wrong decisions.

Having identified mistaken decisions made by the system on the given data, we can use the selected DT model to specify terminal nodes into which these decisions fall. The

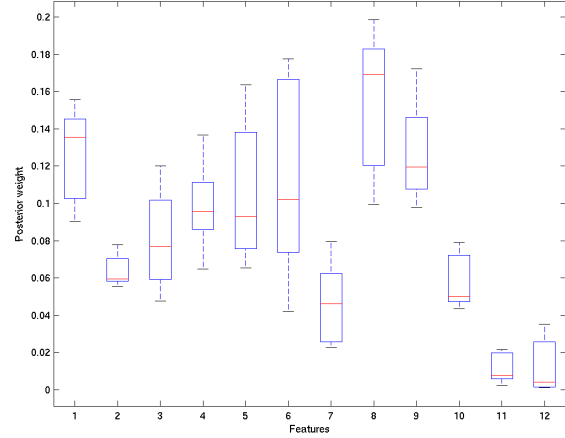


Figure 3: Feature Importance within Random Sub-Sampling Validation

terminals of interest can be converted into a set of n rules in the form *if $x_i \geq q_i$ then ...*, $i = 1, \dots, n$, which is tractable by experts.

The desired model can be found by applying the technique described in Section 4 as a Sure Correct (SC) DT model. This model is compared with two other DT models that were selected by the existing techniques, MAP and MAPW, discussed in Section 4. The comparison is made in terms of misclassification rate within the random sub-sampling validation and shown in Fig. 4. We see that the SC DT model more often outperforms the other two models. The average accuracy is 87.6.

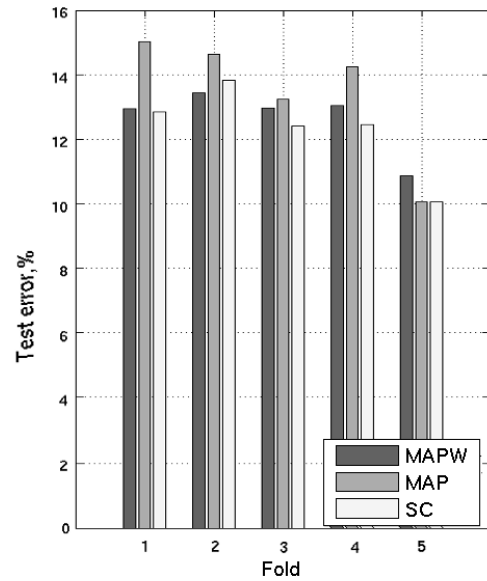


Figure 4: Test errors of SC, MAP, and MAPW Decision Tree Models within Random Sub-Sampling Validation

For comparison we also used the CART technique and

found that the average accuracy was 87.0, which is competitive with the above SC DT model. The CART technique has been run with the Gini diversity index as splitting criterion, using the same number ($p_{min} = 5$) of data points allowed in terminals.

5.5 ESTIMATING UNCERTAINTIES IN DETECTION OF CONFLICTS

Fig. 5 shows the uncertainty intervals estimated by the proposed BMA technology for the aircraft pair whose traces are plotted in Fig. 1. The upper plot shows the distance d over radar cycles. The alert cycles here are marked by the plus sign. We see that the STCA system missed detection of 5 alerts between 23d and 27th cycles. Furthermore, the aircraft positions between 38th and 40th cycles become closer and remain within the distance that triggered the alert at the 18th cycle. This probably means that the system missed detection of new alerts.

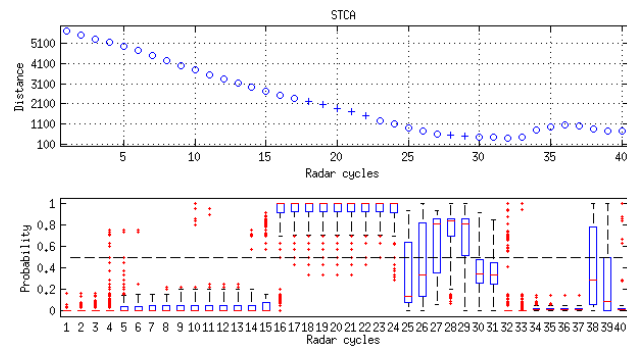


Figure 5: Estimates of Uncertainty Intervals for an Aircraft Pair

The lower plot shows the estimates of uncertainties in decisions made by the proposed Bayesian technology. Boxes here show the summary of predictive posterior probability distributions of alerts. The median probability values exceed the threshold 0.5 between the 16th and 24th cycles. The following 3 cycles are detected with large uncertainty, which indicates a high risk of making wrong decisions. Between the 33rd and 37th cycles the aircraft move away from each other and the probability of conflict decreases. However, between the 38th and 40th cycles they move closer again and we can observe that the uncertainty intervals become larger. This example demonstrates the ability of the proposed technology to provide essential information about risks of making wrong decisions.

6 CONCLUSION

The MCMC technique proposed for Bayesian averaging over DT models was applied to the STCA problem. In this work we aimed at modelling the STCA system within

a Bayesian framework. The use of DT models was introduced in order to provide a possible interpretation of factors that can affect the reliability of STCA decisions. In these experiments, no prior information about possible DT structures was available.

A single DT model was selected from the ensemble of DT models that were collected during MCMC approximation. A DT model can be selected as one providing the Maximum a Posteriori probability. However, we have shown that such a DT model tends to be over-sized and so can underperform. A new technique that is based on estimating the consistency of DT models included in an ensemble was implemented and tested on the STCA data. The experiments show that this approach outperforms the existing techniques in terms of predictive accuracy.

Thus we can conclude that the proposed Bayesian technology can be used to find possible ways of improving accuracy of STCA detection. In a more general context, the proposed technology is capable of providing experts with the full probabilistic information that is required for interpretation of decision making where safety is of crucial importance.

Acknowledgements

The authors are grateful to the anonymous reviewers for useful and constructive comments on the paper. This research was partly supported by the Engineering and Physical Sciences Research Council (EPSRC), GR/R24357/01.

References

- A. Ayuso, L. Escudero, and F. Martn-Campo (2012) A mixed 0-1 nonlinear optimization model and algorithmic approach for the collision avoidance in ATM: velocity changes through a time horizon. *Computers and Operations Research* 39(12) 3136-3146.
- L. Brieman, J. Friedman, R. Olshen, and C. Stone (1984). *Classification and Regression Trees*, Belmont, CA, Wadsworth.
- P. Brooker (2005). Airborne collision avoidance systems and air traffic management safety. *Journal of Navigation* 1 1-16.
- H. Chipman, E. George, and R. McCulloch (1998). Bayesian CART model search, *Journal of American Statistics* 93 935-960.
- H. Chipman, E. George, and R. McCulloch (1998). Making sense of a forest of trees. In S. Weisberg, (ed.), *Symposium on the Interface*. Interface Foundation of North America.
- D. Denison, C. Holmes, B. Malick, and A. Smith (2002). *Bayesian Methods for Nonlinear Classification and Regression*. Wiley.
- P. Domingos (1998). Knowledge discovery via multiple models. *Intelligent Data Analysis* 2 187-202.
- P. Domingos (2000). Bayesian Averaging of classifiers and the overfitting problem, *International Conference on Machine Learning*, 223-230. Stanford, Morgan Kaufmann.
- P. Green (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika* 82 711-732.
- W. Krzanowski, et al. (2006). Confidence in classification: A Bayesian approach. *Journal of Classification* 23(2) 199-220.
- K. Majeske, and T. Lauer (2012). Optimizing airline passenger prescreening systems with Bayesian decision models, *Computers and Operations Research* 39(8) 1827-1836.
- NATS. London Area Control Centre (2002). Short Term Conflict Alerts. *Controller briefing note*. National Air Traffic Services, London
- M. Prandini, J. Hu, J. Lygeros, and S. Sastry (2000). A probabilistic framework for aircraft conflict detection. *IEEE Transactions on Intelligent Transportation Systems* 1(4) 199-220.
- C. Robert, and G. Casella (2009). *Introducing Monte Carlo methods with R*. Springer.
- V. Schetin et al. (2007). Confident Interpretation of Bayesian decision trees for clinical applications. *IEEE Transaction on IT in Biomedicine* 11(3) 312-319.