A Neural Network Based Approach for ESM/Radar Track Association

Yifeng Zhou, Jim Mickeal and Barbara Ford
Defence R&D Canada - Ottawa
3701 Carling Avenue, Ottawa, Ontario
Canada K1A 0Z4
yifeng.zhou@drdc-rddc.gc.ca

Abstract – *In this paper, a neural network based ESM/radar track association algorithm is presented. The algorithm consists of a feed-forward neural network and a probability combiner. The neural network classifier is trained using the radar bearing measurements as well as their time stamps to approximate the a posteriori probabilities. The ESM bearing measurements along with their time stamps are fed to the trained network to provide a sequence of a posteriori probabilities. The probability combiner combines the local a posteriori probabilities to provide the global a posteriori probabilities that an ESM track belongs to each individual radar track. The track association logic associates the ESM track with the radar track that has the maximum a posteriori probability. The approach is able to eliminate the complex track time alignment process that is required by other techniques. It also alleviates the requirement for the Gaussian assumption about the measurements. Computer simulations are used to demonstrate the performance and effectiveness of the proposed algorithm.*

Keywords: track association, sensor fusion, ESM, radar, neural networks, classifier.

1 Introduction

The problem of multi-sensor data fusion has received considerable interest in recent years [1][2]. Radar and electronic support measures (ESM) sensors are two of the most important sensor types used in sensor fusion applications. Radar is an active device that provides the kinematic information about a target On the other hand, an ESM sensor is passive that intercepts the target emissions and performs emitter classification and identification for electronic defence purposes. The outputs of an ESM sensor contain the target attributes (emitter type, platform category and their threat level *etc.*) as well as their bearings. Integrating the kinematics and attributes of targets will benefit both the radar and ESM sensors. For example, introducing target kinematics into ESM analysis helps to enhance the performance of classification and identification. Similarly, utilizing target attribute information in radar tracking can improve the track continuity and tracking of maneuvering targets. It is known that ESM sensors have a range advantage over most radars because the received signal by the ESM sensor only travels one way while the radar signals have to travel two ways [3]. ESM sensors can be deployed to augment surveillance and provide early warning for radar to reduce initiation time for long range targets. The bearing measurement of ESM sensors can also be used to help reduce the vulnerability of radar to jamming. In general, ESM/radar sensor integration can increase the likelihood of target acquisition, improve situation awareness and provide more accurate and complete composite target track files.

This paper focuses on the problem of ESM/radar track association, which is a pre-requisite step for ESM/radar sensor fusion. By association we mean the grouping of tracks observed by radar and ESM sensors that are supposedly from the same emitter. The problem of ESM/radar track association is unique in that the only common parameter between radar and an ESM sensor is the bearing measurement. Often, the ESM bearing measurements are relatively inaccurate than the radar measurements. The measurements from radar and ESM sensors are not synchronized, and may not even cover the same time interval. This is due to the fact that the reporting of an ESM sensor mainly depends on the emission, illumination, antenna scanning type of the target emitters, and the performance and structure of the receiver. The difficulty may be further compounded by possible ESM track fragmentation, which is often caused by a dense signal environment and the existence of complex (agile) signals.

The problem of ESM/radar track association has not received much attention in the past, and publications on it are limited. Trunk and Wilson [4] are among the first ones to consider the problem. In their formulation, DF tracks are given with different numbers of measurements, and they attempt to associate each DF track with either no radar track or one of the tracks. The approach is based on an assumption that the DF measurements are independent and Gaussian distributed with zero mean and equal *a priori* probabilities. The algorithm uses a combination of Bayes and Neyman-Pearson procedures. Time alignment procedures are required to make the maximum use of the data. The *a posteriori* probabilities of a DF track that would associate with each radar track are calculated. The decision rule selects the association that has the largest probability. Different thresholds are set to accommodate the real scenario in which radar and DF tracks are constantly initiated, updated and dropped. The algorithm has two disadvantages. First, it requires a prediction or interpolation for measurements that are not time aligned. Secondly, the algorithm requires definitions of different threshold settings using Monte Carlo simulations. These requirements are time-consuming and would increase the algorithm complexity making it difficult for real-time implementation. There have been other efforts to improve the performance or computation complexity of the Trunk-Wilson algorithm such as the fuzzy logic approach by Smith [5] and the triple-threshold radar-to-ESM correlation algorithm by Wang *et al.* [6]. However, these algorithm still can not avoid the difficulties that are encountered by the Trunk-Wilson algorithm.

Another type of ESM/radar track association algorithms utilize the bearing-only track techniques for the ESM bearing measurements [6][7]. In these approaches, bearing-only tracking techniques are applied to ESM bearing measurements for tracking. The association logic is a hypothesis test based on the most recent state estimates by the radar and ESM sensor. For asynchronous sensor measurements, estimates are predicted from one of the sensors to the time of the estimates of the other. These types of algorithms are track dependent, i.e., their performance relies on the appropriate modeling of the target dynamics. Their application is also restricted by the observability conditions. In order to ensure that the model parameters are observable, the platform carrying the ESM sensor must have a motion with a non-zero derivative of higher order than that of the target [8].

In this paper, a neural network based ESM/radar track association algorithm is proposed. The algorithm consists of a feed-forward neural network and a *a posteriori* probability combiner. The neural network implements as a classifier and its outputs are trained to approximate the *a posteriori* probabilities that the inputs belong to radar tracks. The network is trained using the radar bearing measurements as well as their time stamps. The ESM bearing measurements from a track together with time stamps are fed to the trained network to provide a set of *a posteriori* probabilities. The probability combiner computes the global *a posteriori* probabilities that the ESM track belongs to each individual radar track. The track association logic associates the ESM track with the radar track that has the largest *a posteriori* probability. The proposed association algorithm is able to eliminate an extra complex track time alignment or interpolation process required by other techniques. It also alleviates the requirement for the Gaussian assumption about the measurements, an assumption that may not hold in practical applications. Computer simulations are used to demonstrate the performance and effectiveness of the proposed algorithm.

2 Problem Formulation

Consider a radar and an ESM sensor that are distributed at different locations. Without loss of generality, we select the system coordinate with the location of the ESM sensor as the origin. The radar measurements are assumed to have been converted from the local radar coordinate to the system coordinate. Assume that K_e target bearing tracks that are reported by the ESM sensor. Let

$$\theta_e^{(k)}(t_n) = \bar{\theta}^{(k)}(t_n) + n_e^{(k)}(t_n) \tag{1}$$

for $k = 1, 2, \ldots, K_e$ and $n = 1, 2, \ldots, N_e(k)$, where $\bar{\theta}^{(k)}(t_n)$ denotes the actual bearing of the kth target at the t_n, $N_e(k)$ is the number of measurements of the kth track and $n_e^{(k)}(t_n)$ is the measurement noise. The measurement noise is assumed to be zero mean. Let K_r denote the number of targets that are observed by the radar. The bearings observed by the radar in the system coordinate can be calculated by

$$\theta_r^{(k)}(t_n) = \arctan\left[\frac{y_k(t_n)}{x_k(t_n)}\right], \tag{2}$$

where $y_k(t_n)$ and $x_k(t_n)$ are the x and y coordinates of the kth target measured by the radar at t_n, respectively. It is known that radar usually reports the azimuth and ranges of targets. The error in the range and bearing measurements can be described by a Gaussian distribution. The Gaussian distribution is often motivated by the central limit theorem and has the advantage of mathematical convenience. Note that, when the radar and ESM sensor are co-located, no conversion is required for the radar measurements, and the bearing measurements by the radar can be modeled by a Gaussian distribution. However, when the radar and ESM sensor are not co-located, the bearing measurements by the radar in the system coordinates are calculated according to (2). Since the conversion is a highly nonlinear process, the radar bearing measurement uncertainty in the system coordinate can no longer be described by a Gaussian distribution.

The ESM/radar track association problem can be stated as: given the bearing measurements

$$\theta_e^{(k)}(t_n), \quad k = 1, 2, \ldots, K_e; \; n = 1, 2, \ldots, N_e(k)$$

$$\theta_r^{(k)}(t_n), \quad k = 1, 2, \ldots, K_r; \; n = 1, 2, \ldots, N_r(k)$$

decide whether an ESM track should be associated with a radar track or with no radar. In the latter case, the ESM bearing track is either a fragmented track or the radar does not detect the target that is observed by the ESM sensor.

3 Neural Network Based ESM/Radar Track Association

The proposed neural network based ESM/radar track association algorithm consists of two parts: a neural network classifier and a *a posteriori* probability combiner. Each output of the neural network corresponds to a radar track, which we denote as a class. The approach contains two stages: training and classification. In the training stage, the network outputs are trained using the radar bearing measurements together with their time stamps to model the *a posteriori* probability of a class membership of the inputs. In the classification stage, the ESM bearing track measurements with their time stamps are fed to the network to produce a sequence of local *a posteriori* probabilities. The probability combiner combines the local *a posteriori* probabilities to provide the global *a posteriori* probabilities that the ESM track would associate with each individual radar track. The association is decided based on the largest *a posteriori* probability.

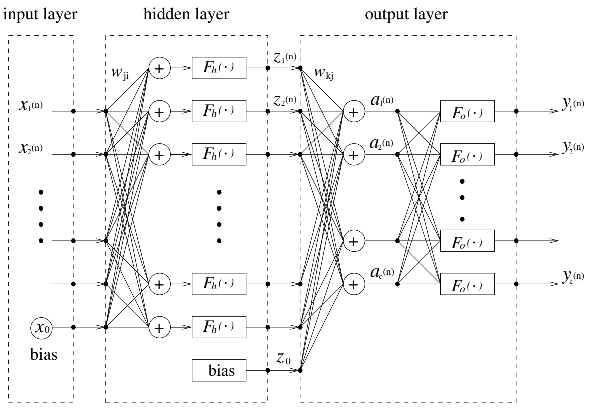

Fig. 1: Structure of a two-layer feed-forward neural network.

3.1 Neural network classifier

A neural network consists of many neurons that are interconnected. A neuron is an information processing unit and is fundamental to the operation of a neural network. Neural networks have been used in a wide variety of applications because of their nonlinearity nature and adaptivity [9]. The input-output mapping ability and evidential response of a neural network make it particularly suitable for pattern recognition problems. In addition, the massively parallel nature of the neural networks makes it potentially fast for certain computation tasks. This same feature also makes the neural network ideally suited for very-large-scale-integrated (VLSI) implementation, making neural networks effective tools for real-time applications.

We consider the use of multi-layer feed-forward networks in this paper, in particular, the two-layer feed-forward neural networks. The two-layer feed-forward neural network is often referred to as a multi-layer perceptron (MLP). It is one of the simplest and most commonly encountered neural networks. A network with two-layers can approximate any continuous mapping to arbitrary accuracy provided the number of hidden units is sufficiently large [10]. In the context of classification problems, two-layer networks can approximate any decision boundary to arbitrary accuracy. They provide universal discriminant functions and are sufficient to model *a posteriori* probabilities of class memberships [11][12].

Figure 1 shows the topology of a two-layer feed-forward network. The network comprises one input layer, one hidden layer and one output layer. Each of the layers contains a different number of units that are connected to every unit in the subsequent layer. Each connection is called a link and is represented by a weight. In the input and hidden layer, biases are introduced and treated as weights with inputs fixed at one. We use i, j and k to denote the unit index for the input, the hidden and the output layer, respectively. Parameter w_{ij} denote the weight connecting the jth input unit and the ith hidden unit, and similarly, w_{kj} the weight connecting the jth hidden unit and the kth output unit. An index 0 denotes a bias unit in the input and hidden layer, the outputs of which are given by 1. In Figure 1, $F_h(\cdot)$ is the nonlinear activation function of the hidden layer. The nonlinearity of the activation functions ensures that the neural network has the generalization capability. In addition, the activation function should be chosen as a differential function so that the gradient-based optimization methods can be used for network parameter estimation. The activation function can take different forms including the logistic sigmoid, tan-sigmoid and softmax function [13]. In this paper, the tan-sigmoid function is used for the hidden layer.

The output layer in Figure 1 has a more general form and allows the use of the more complex softmax activation. In classification problems, the outputs of a network are usually designed to model the *a posteriori* probabilities of class membership. To ensure that the outputs of a network are true probabilities, the output values must lie in the range of $(0, 1)$ and sum up to one. Obviously, the output of a sigmoid function remains between between 0 and 1. However, it does not necessarily sum up to one. The outputs of a softmax activation function satisfy both conditions and can be used to model true probabilities.

The feed-forward networks are supervised networks. They can learn mappings from a suitable data set with targets. The learning depends on the definition of an error function. In the learning process, the error function is minimized with respect to the networks weights and biases. Two error functions will be considered: the sum-of-squares and the cross-entropy error functions. The sum-of-squares error function is simply the sum of the squares of the errors of the outputs of the training data and the target data. The sum-of-squares error function can be derived from maximum likelihood under the assumption that the target data is Gaussian distributed. The outputs of a network trained by minimizing a sum-of-squares error function can be shown to approximate the conditional averages of the target data. In classification problems, when the $1 - of - c$ coding is used, the outputs of a network with the sum-of-squares error function approximate the *a posteriori* probabilities [13]. The cross-entropy error function is a more appropriate error function for classification problems [13]. It is based on the use of a Bernoulli distribution about the target data, which more accurately describes the distribution of the target data.

The learning process of neural networks is to find the optimal weights and biases that minimize the error function. There are different training styles for a feed-forward network including the incremental and batch based methods. In incremental training, the weights and biases are updated each time training data is presented to the network. In batch training, the weights and biases are updated after all the inputs are present. Many numerical optimization techniques can be applied such as the gradient descent based and the quasi-Newton type algorithms and their variants [14].

3.2 Combination of *a posteriori* probability

The neural network classifier has $K = K_r$ output units. Each output corresponds to a radar track. The network is trained using the radar track data, $\underline{\mathbf{x}}_r^{(k)}(n) = [\theta_r^{(k)}(t_n), t_n]$, for $k = 1, 2, \ldots, K_r$ and $n = 1, 2, \ldots, N_r(k)$. Thus, the number of units of the input layer is given by 2. The kth output of the neural classifier is trained to approximate the *a posteriori* probability that the input belongs to the kth class (radar track). Define $\underline{\mathbf{x}}_e^{(i)}(n) = [\theta_e^{(i)}(t_n), t_n]$, for

$i = 1, 2, \ldots, K_e$ and $n = 1, 2, \ldots, N_e(i)$. When $\underline{\mathbf{x}}_e^{(i)}(n)$ is presented to the classifier, the kth output of the network is the *a posteriori* probability that $\underline{\mathbf{x}}_e^{(i)}(n)$ belongs to the kth radar track, $P(c_k \mid \underline{\mathbf{x}}_e^{(i)}(n))$, where c_k denotes the hypothetical event that $\underline{\mathbf{x}}_e^{(i)}(n)$ belongs to the kth radar track.

For $\{\underline{\mathbf{x}}_e^{(i)}(n),\ n = 1, 2, \ldots, N_e\}$, we have a set of *a posteriori* probabilities. The function of the probability combiner is to combine all the local *a posteriori* probabilities to provide the global *a posteriori* probability that the ESM track belongs to each individual radar track. Let $\mathbf{X}_e^{(i)} = \{\underline{\mathbf{x}}_e^{(i)}(n);\ n = 1, 2, \ldots, N_e(i)\}$. Using the Bayesian theorem, we can obtain the global *a posteriori* probability $P(c_k \mid \mathbf{X}_e^{(i)})$ as

$$P(c_k \mid \mathbf{X}_e^{(i)}) = \frac{p\{\mathbf{X}_e^{(i)} \mid c_k\} P(c_k)}{p(\mathbf{X}_e^{(i)})}, \tag{3}$$

where $P(c_k)$ denotes the *a priori* probability. Since no underlying correlation among the bearing measurements is assumed, the measurements can be considered class conditionally independent. Then, (3) can be written as

$$P(c_k \mid \mathbf{X}_e^{(i)}) = \frac{1}{\alpha} \prod_n p\{c_k \mid \underline{\mathbf{x}}_e^{(i)}(n)\} P(c_k), \tag{4}$$

where α is a normalization factor.

We can also compute the global *a posteriori* probability in a recursive manner. Let $\mathbf{X}_e^{(i)}(n) = \{\underline{\mathbf{x}}_e^{(i)}(t_1), \underline{\mathbf{x}}_e^{(i)}(t_2), \ldots, \underline{\mathbf{x}}_e^{(i)}(t_n)\}$. According to the Bayesian theorem, the *a posteriori* probability $P\{c_k \mid \mathbf{X}_e^{(k)}(n+1)\}$ can be written as

$$\begin{aligned} p\{c_k \mid \mathbf{X}_e^{(i)}(n+1)\} \\ = \frac{p\{\underline{\mathbf{x}}_e^{(i)}(n+1) \mid c_k\} p\{c_k \mid \mathbf{X}_e^{(i)}(n)\}}{p\{\mathbf{x}_e^{(i)}(n+1) \mid \mathbf{X}_e^{(i)}(n)\}}, \end{aligned} \tag{5}$$

where $\{\mathbf{x}_e^{(i)}(n)\}$ are assumed to be class conditionally independent. For a network with the sum-of-squares error function, since there is no guarantee that the outputs satisfy the sum of one constraint, $p\{c_k \mid \mathbf{X}_e^{(i)}(n+1)\}$ may need to be normalized each time when it is updated.

The *a priori* probability $P(c_k)$ can be approximated by the average of the kth output of the network over all the training data [12]

$$P(c_k) \approx \frac{1}{N_a} \sum_n^{N_r(k)} P\{c_k \mid \underline{\mathbf{x}}_r^{(k)}(t_n)\}, \tag{6}$$

where $N_a = \sum_k N_r(k)$ is the total number of training data. The *a priori* probability can also be simply chosen to be the fraction of each radar track measurements. If the *a priori* probabilities for the ESM measurements are different from those of the radar tracks, the *a posteriori* probability can be corrected by multiplying a correction factor, which is the ratio of the *a priori* probability corresponding to the radar tracks to that of the ESM tracks.

When the global *a posteriori* probabilities that an ESM track is associated with each individual radar track are estimated, the association logic assigns the ESM track to the radar track that has the maximum *a posteriori* probability. This classification is known as the minimum error rate classification [15]. Let $P_{max}^{(i)}$ denote the maximum *a posteriori* probability. The probability of misclassification is given by

$$e(\mathbf{X}_e^{(i)}) = 1 - P_{max}^{(i)}. \tag{7}$$

If the largest *a posteriori* probability is relatively low, the association becomes doubtful, and we should reject any association. The decision to reject an association can be done comparing the largest *a posteriori* probability with a preset threshold. If the largest *a posteriori* probability is below the threshold, the association should be rejected. It is known that the highest probability of misclassification occurs when all the *a posteriori* are equal. The maximum possible probability of misclassification is given by [16]

$$e_{max} = 1 - \frac{1}{K}. \tag{8}$$

The rejection threshold should be selected between $1 - e_{max}$ and 1. The value of the threshold decides the size of the rejection class and the total misclassification probability. Decreasing the threshold will increase the probability of misclassification while increasing it will increase the size of the reject class, which in turn increases the complexity of any system that is used to handle the reject class.

4 Computer Simulations

Computer simulations are used to study the performance of the proposed track association algorithm. An ESM sensor and a radar, both stationary, are simulated. A common coordinate system is selected where the location of the ESM sensor is the origin, and its x and y axes point the east and north, respectively. The radar is assumed to be located at (u, v), where $u = 40km$ and $v = 10km$.

The target kinematics are modeled using a standard near-constant-velocity two-state model [17], where it is assumed that the x and y coordinates are decoupled and noises into different coordinates are mutually independent. Seven targets are simulated. The initial positions and velocities are listed in Table 1, where the position and velocity parameters are in the units of km and km/s, respectively. The standard deviations of the velocities are selected to be 0.06 times of the mean velocity for each target in one sampling interval. An observation duration of 250 seconds is simulated. Figure 2 and 3 show the trajectories and bearings of the targets in the system coordinates.

The radar is assumed to have a nominal sampling interval of 2 seconds. The radar and ESM sensor are each simulated with detection probabilities of 0.95 and 0.80, respectively. The actual times that the measurements are received are simulated to be uniformly distributed around their nominal times with a twenty percent perturbation of the nominal intervals. The radar and ESM sensor are not necessarily

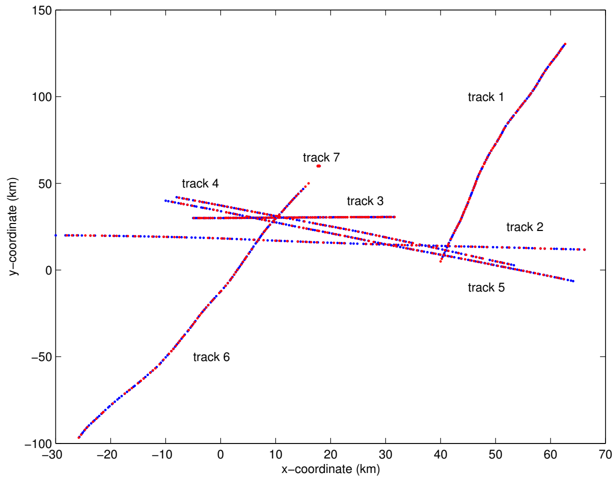

Fig. 2: Target trajectories in the system coordinates.

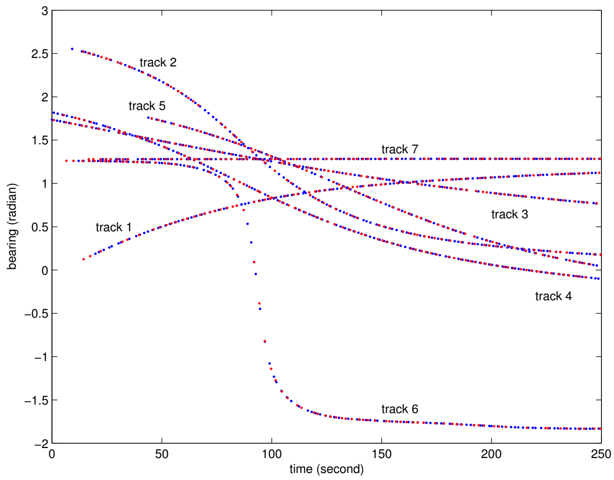

Fig. 3: Bearing tracks in the system coordinates.

Table 1: Initial target locations and velocities

	$x(0)$	$y(0)$	$\dot{x}(0)$	$\dot{y}(0)$
T1	10.0000	5.0000	0.0720	0.4948
T2	-30.0000	20.0000	0.4000	0.0000
T3	-5.0000	30.0000	0.1500	0.0000
T4	-10.0000	40.0000	0.2830	-0.1698
T5	-8.0000	42.0000	0.2830	-0.1698
T6	16.0000	50.0000	-0.1829	-0.5715
T7	18.0000	60.0000	-0.0005	0.0000

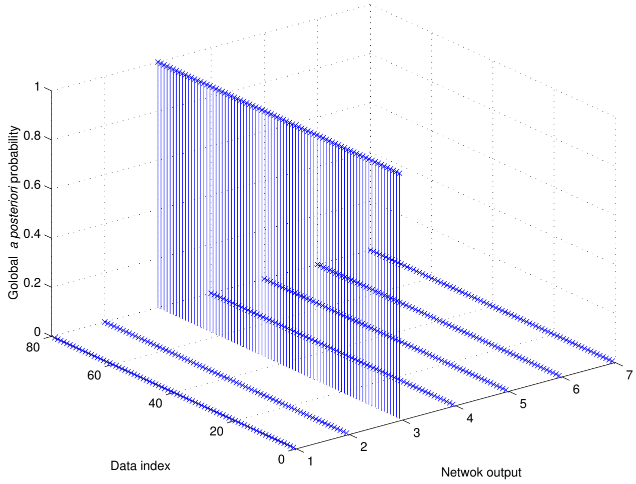

Fig. 4: Local *a posteriori* probabilities when an ESM track is fed the association network.

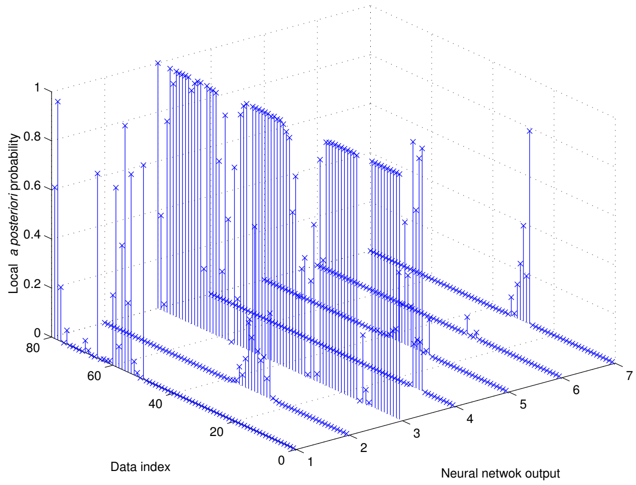

Fig. 5: Global *a posteriori* probabilities when an ESM track is fed the association network.

synchronized in time. The measurements are simulated to contain additive *i.i.d* Gaussian noises.

Figure 4 and 5 show the local and global *a posteriori* probabilities, respectively. The neural network uses the entropy error function and has 25 hidden units. After training, we feed the 4th ESM bearing track to the network and record both the neural network and probability combiner outputs. It can be seen that the local probabilities, or the neural network outputs, scatter over different radar track and do not give any clear solutions to the association problem. On the other hand, the global *a posteriori* probabilities, after a few inputs, point to the correct track association with close to 1 *a posteriori* probability.

Figure 6 shows the variations of rate of correct association versus the number of hidden units. The standard deviations for the radar and ESM bearing measurements are selected to be 0.8^o and 1.5^o, respectively. The standard deviation for the radar range measurements is given by 0.2km. For each number of hidden units, 100 tests are repeated and the number of correct association for each ESM bearing track is counted. For an ESM track, a correct association means that it produces largest *a posteriori* probability in

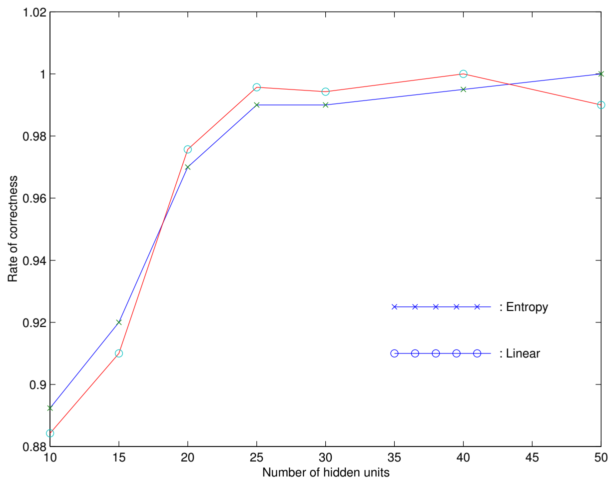

Fig. 6: Rate of correct association versus number of hidden units.

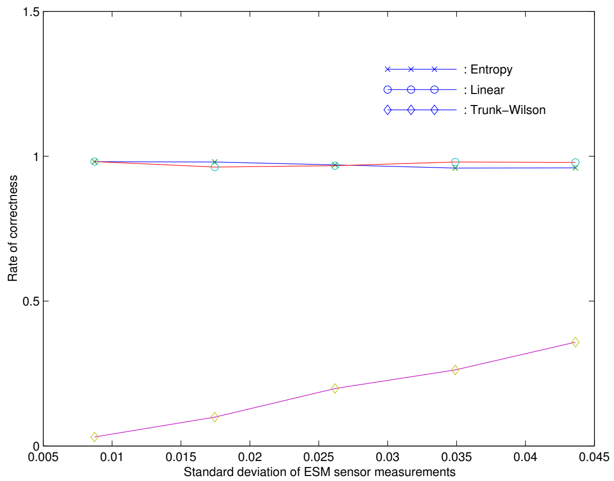

Fig. 7: Rate of correct association versus standard deviation of the ESM bearing measurement.

the network output corresponding to the correct radar track, and at the same time, larger than a preset threshold. The threshold is set to be 0.5. In Figures 6, the averaged rates of correct association is plotted for algorithms using both the sum-of-squares and entropy error functions. It can be seen that the algorithms show similar performance. Their performance improves as the number of hidden units increases. It can be observed that when the number of hidden units is greater than 20, the rates of correct association approach to 1. In this particular case, as the number of hidden units increases, both networks generalize well and over-fitting has not been not observed.

Figures 7 shows the variations of rate of correct association versus the standard deviation for the ESM bearing measurements. The number of hidden units is selected to be 20. For the Trunk-Wilson algorithm, we use the one-step fixed-lag smoothing approach proposed by Helmick *et al.* [18] to time-translate the radar measurements to the times of the ESM sensor measurements. For simplicity, a single-model Kalman filter is used for the constant velocity target model. The upper threshold is calculated using computer simulation, in which the mean difference is set to equal to the standard deviation of the ESM bearing measurement, and the false alarm probability is set to 0.01. Since the threshold is computed for only a limited number of points, a polynomial of order 5 is used to interpolate the results. Figure 7 shows the rates of correct association for networks using the sum-of-squares and entropy functions, and the Trunk-Wilson algorithm. The neural network based association algorithms perform similarly in that they all have rates of correct association that are close to 1. The Trunk-Wilson algorithm, however, has low rates of correct association for all standard deviation values. It is interesting to note that the performance of the Trunk-Wilson algorithm improves as the standard deviation of the ESM sensor measurements increases. This can be explained as follows. Since the radar bearing measurements in the system coordinates are not Gaussian due to the nonlinear local radar to system coordinate conversion, the difference between the radar and ESM measurements does not follow the Gaussian distribution for small ESM bearing measurement noise. The Gaussian assumption that the Trunk-Wilson algorithm is based becomes invalid. However, as the standard deviation of the ESM measurements increases, the ESM sensor measurements become dominant and the difference between the sensor measurements can be better approximated as Gaussian distributed.

References

[1] D. L. Hall, *Mathematical techniques in multisensor data fusion*, Artech House, Norwood, MA, 1992

[2] Y. Bar-Shalom, ed., *Multitarget-multisensor tracking : Advanced application*, Artech House, 1990

[3] R. G. Wiley, *Electronic Intelligence: The Analysis of Radar Signals*, 2nd ed., Artech House, Norwood MA, 1993

[4] G. V. Trunk and J. D. Wilson, "Association of DF bearing measurements with radar tracks", *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-18, no. 4, pp. 438-447, July 1987

[5] J. F. Smith, "A fuzzy logic multisensor association algorithm", *Proc. SPIE*, vol. 3068, pp. 76-87, April 1997

[6] G. H. Wang, S. Y. Mao, Y. He and Y. J. Liu, "Triple-threshold radar-to-ESM correlation algorithm when each radar track is specified by different number of measurements", *IEE Proc.- Radar, Sonar Navig.*, vol. 147, no. 4, pp. 177-187, August 2000

[7] A. Farina and R. Miglioli, "Association of active and passive tracks for airborne sensors", *Signal Processing*, vol. 69, pp. 209-217, 1998

[8] K. Becker, "A general approach to TMA observability from angle and frequency measurements", *IEEE Trans. Aerospace Electron. System*,vol. 32, no. 1, pp. 487-494, January 1996

[9] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Macmillan College Publishing Company, 1994

[10] K. Hornik, M. Stinchcombe and H. White, "Multilayer feedforward networks are universal approximations", *Neural Networks*, 2 (5), pp. 551-560, 1990

[11] E. K. Blum and L. K. Li, "Approximation theory and feedforward networks", *Neural Networks*, 4(4), pp. 511-515, 1991

[12] M. D. Richard and R. P. Lippmann, "Neural network classifiers estimate Bayesian *a posteriori* probabilities", *Neural Computation*, vol. 1, pp. 425-464, 1989

[13] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995

[14] J. E. Dennisand R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, Englewood Cliffs, NJ : Prentice-Hall, 1983

[15] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley: New York, 1973

[16] P. Devijer and J. Kittler, *Pattern Recognition: A Statistical Approach*, Prentice-Hall International, 1982

[17] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*, Academic Press, 1998

[18] R. E. Helmick, J. E. Conte, S. A. Hoffman and W. D. Blair, "One-step fixed-lag IMM smoothing for alignment of asynchronous sensors", *Proc. 1994 SPIE Aerospace Sensing Symposium - Signal and Data Processing of Small Targets*, vol. 2235, pp. 507-518, Orlando, FL, April 1994