

Variational Bayesian Modified Adaptive Mamdani Fuzzy Modelling for use in Condition Monitoring

Yu Zhang, Jun Chen, Chris Bingham, Timothy Gordon

School of Engineering
University of Lincoln
Lincoln, U.K.

{yzhang; juchen; cbingham; tgordon}@lincoln.ac.uk

Abstract—The paper proposes a new Adaptive Mamdani Fuzzy Model (AMFM) based system modelling methodology that improves on traditional Mamdani fuzzy rule based system (FRBS) techniques through use of alternative membership functions and a defuzzification mechanism that is ‘differentiable’, allowing a back error propagation (BEP) algorithm to refine the initial fuzzy model. Moreover, a variational Bayesian (VB) method is applied to simplify the results via automatic selection of the number of input rules so that redundant rules can be removed for the initial modelling phase. The efficacy of the proposed VB modified AMFM (VB-AMFM) approach is demonstrated through experimental trials using measurements from a compressor in an industrial gas turbine (IGT).

Keywords— Adaptive Mamdani fuzzy model; back error propagation; variational Bayesian Gaussian mixture model; industrial gas turbine.

I. INTRODUCTION

The adoption of predictive condition monitoring for IGTs has attracted considerable recent attention due to widely recognized benefits of facilitating reduced down-time and assurance of safety [1]. Signal processing based techniques have been applied extensively in condition monitoring for IGTs [2,3]. However, such techniques can only provide warnings after faults have occurred, in which case, model-based system identification approaches have become advantageous by providing early warnings based on comparisons from predictive models.

An adaptive Mamdani fuzzy model (AMFM [4]) has previously been reported in [5] for IGTs. The original Mamdani fuzzy rule based system (FRBS) has historically been a popular type of fuzzy inference systems [6], and relies on the calculation of a relational matrix for each rule, and subsequently, the overall relational matrix. Finally, the output fuzzy set is obtained using the composition rule of inference. However, Mamdani FRBS has an inherent drawback that it is not differentiable with respect to membership function parameters, which prevents the use of back-error-propagation (BEP) techniques to refine the fuzzy models [7]. Therefore, AMFM was proposed to remove this limitation, through the use of alternative membership functions and a defuzzification mechanism.

After building the AMFM, a simplification is needed to improve the interpretability by removing redundancy both in the rules and in the fuzzy sets to achieve optimisation and accuracy. In [6], an immune inspired multi-objective fuzzy modelling has been proposed to achieve this, however, it is very time consuming. Therefore, in this paper, a variational Bayesian Gaussian mixture model (VBGMM) is applied to automatically cluster the initial rules of AMFM in order to simplify the final interpretation. The efficacy of the proposed variational Bayesian method modified AMFM approach (VB-AMFM) is demonstrated using results from experimental trials of a compressor on an IGT.

II. METHODOLOGY DESCRIPTION

The main concept of the methodology is to automatically assign the number of the input rules in AMFM through the application of VBGMM so that redundant rules can be removed, and interpretability of the results can be improved.

A. Adaptive Mamdani Fuzzy Model

Denote X_m and y_m being the inputs and output of the m th data point. AMFM uses Gaussian membership functions for the premises and bell-shape membership functions for the consequents:

$$\mu_{B_i}(y_m) = \frac{1}{1 + \left(\frac{y - c_i^y}{\sigma_i^y} \right)^2}, \quad (1)$$

where μ is the membership degree that y_m maps to B_i , B_i represents the i th bell-shape membership, c_i^y and σ_i^y are the centre and the spread of the i th membership function of the output y , where $i = 1, \dots, k$, and k is the number of rules. The final crisp function is:

$$y^{crisp}(X|\theta) = \frac{\sum_{i=1}^k b_i \cdot \mu_i(X) \cdot \int_y \mu_{B_i}(y) dy}{\sum_{i=1}^k \mu_i(X) \cdot \int_y \mu_{B_i}(y) dy}, \quad (2)$$

where b_i is the centre of area of the membership function $\mu_{B_i}(y)$ and is the peak c_i^y if $\mu_{B_i}(y)$ is symmetric. y^{crisp} is the final defuzzified output of the FRBS.

$\theta = \left(b_i, \sigma_i^y, c_i^j, \sigma_i^j \mid i = 1, \dots, k; j = 1, \dots, n \right)$ is the parameter vector where each parameter is linked directly to the identified cluster (rule) centres and spreads. Here, b_i is the output of the i th rule; c_i^j and σ_i^j are the centre and spread of the i th membership function of the j th input, and n is the number of the dimensions of the inputs. And $\int_y \mu_{B_i}(y) dy$, which denotes the area under $\mu_{B_i}(y)$ over the output interval $y: [y_L, y_U]$, is calculated using

$$\int_y \mu_{B_i}(y) dy = \sigma_i^y \left[\arctan\left(\frac{y_U - b_i}{\sigma_i^y}\right) - \arctan\left(\frac{y_L - b_i}{\sigma_i^y}\right) \right] \stackrel{def}{=} g(b_i, \sigma_i^y). \quad (3)$$

A Mamdani FRBS with the pre-specified number of rules is extracted from the numerical data. However, the initial fuzzy model is not ideal since the membership parameters need to be pruned for further accuracy. A constrained BEP algorithm is thus utilized to obtain a vaccine model, through the construction of which, in terms of its predictive performance, many generations of evolutionary search can be saved.

By taking the partial derivatives of (2) with respect to each parameter in θ , the following parameter updating formulas are obtained:

1) Centre of the consequent updating law:

$$b_i(t+1) = b_i(t) - \lambda_1 \cdot \varepsilon_m(t) \cdot \frac{\mu_i(t)(X_m) \cdot \zeta(t)}{\sum_{i=1}^k \mu_i(t)(X_m) \cdot g(b_i(t), \sigma_i^y(t))} + \beta_1 \cdot \Delta b_i(t-1),$$

where

$$\zeta(t) = g(b_i(t), \sigma_i^y(t)) + b_i(t) \cdot g'(b_i(t)) - g'(b_i(t)) \cdot y^{crisp}(X_m | \theta(t))$$

2) Spread of the consequent updating law:

$$\sigma_i^y(t+1) = \sigma_i^y(t) - \lambda_2 \cdot \varepsilon_m(t) \cdot \frac{\mu_i(t)(X_m) \cdot g'(\sigma_i^y(t)) \cdot [b_i(t) - y^{crisp}(X_m | \theta(t))]}{\sum_{i=1}^k \mu_i(t)(X_m) \cdot g(b_i(t), \sigma_i^y(t))} + \beta_2 \cdot \Delta \sigma_i^y(t-1),$$

3) Centre of the premise updating law:

$$c_i^j(t+1) = c_i^j(t) - \lambda_3 \cdot \varepsilon_m(t) \cdot \frac{g(b_i(t), \sigma_i^y(t)) \cdot [b_i(t) - y^{crisp}(X_m | \theta(t))]}{\sum_{i=1}^k \mu_i(t)(X_m) \cdot g(b_i(t), \sigma_i^y(t))} \cdot \mu_i(t)(X_m) \cdot \left[\frac{x_m^j - c_i^j(t)}{(\sigma_i^j(t))^2} \right] + \beta_3 \cdot \Delta c_i^j(t-1),$$

4) Spread of the premise updating law:

$$\sigma_i^j(t+1) = \sigma_i^j(t) - \lambda_4 \cdot \varepsilon_m(t) \cdot \frac{g(b_i(t), \sigma_i^y(t)) \cdot [b_i(t) - y^{crisp}(X_m | \theta(t))]}{\sum_{i=1}^k \mu_i(t)(X_m) \cdot g(b_i(t), \sigma_i^y(t))} \cdot \mu_i(t)(X_m) \cdot \left[\frac{(x_m^j - c_i^j(t))^2}{(\sigma_i^j(t))^3} \right] + \beta_4 \cdot \Delta \sigma_i^j(t-1),$$

where $\varepsilon_m = y^{crisp}(X_m | \theta) - y_m$, $g'(b_i) = g'(b_i | \sigma_i^y)_{b_i}$ and $g'(\sigma_i^y) = g'(b_i | \sigma_i^y)_{\sigma_i^y}$. (4)

Here, $\lambda_1 \sim \lambda_4$ and $\beta_1 \sim \beta_4$ are user-specific parameters and are the step sizes and the gains of momentum terms, respectively. A problem with the BEP updating formulas derived above is that they include no constraints with respect to the updating mechanism of the parameters. Therefore, a constraint handling scheme is added which checks the boundary violation for centers during each iteration step, and the violated solutions are assigned to the boundaries. More details of AMFM can be found in [4, 7].

B. Variational Bayesian Gaussian Mixture Model

A Gaussian mixture model (GMM) approximates the probability density function of sample data through a linear combination of Gaussian distributions. In this way, the probabilistic distribution of measurements can be expressed as a sum of K Gaussian distributions with their corresponding means μ_k and standard deviations σ_k . A GMM consisting of K mixture components is therefore denoted as

$$p(x | \theta, \mu, \sigma) = \sum_{k=1}^K \theta_k \mathcal{N}(x | \mu_k, \sigma_k), \quad (5)$$

where $\mathcal{N}(x | \mu, \sigma)$ represents a Gaussian distribution density function of a multidimensional variable x , with mean $\mu \equiv \{\mu_k\}$, standard deviation $\sigma \equiv \{\sigma_k\}$, and θ_k are the mixing coefficients.

Let $X \equiv \{x_n\}$ denote a set of N data samples, where each sample consists of the (possibly) multidimensional variable x , and $n = 1, \dots, N$. Assuming the x_n are statistically independent, the probability function of the data set can be expressed as the multiplication of the probabilities of the N data points:

$$p(X|\theta, \mu, \sigma) = \prod_{n=1}^N \left(\sum_{k=1}^K \theta_k \mathcal{N}(x_n | \mu_k, \sigma_k) \right), \quad (6)$$

which is the likelihood function.

To select the necessary number of mixture components, a Variational Bayesian (VB) technique is used [8]. For each data sample x_n , there exists a set of binary latent variables $z_{nk} \in \{0,1\}$ that specify the relationship between the N data points and the K mixture components, effectively indicating which data sample belongs to which mixture component. In the raw GMM technique, the number K is pre-determined, but in VBGMM, K is solved as an unknown through the solutions of $\{z_{nk}\}$.

The joint probability of all variables in the GMM is then given by

$$p(X, Z, \theta, \mu, \sigma) = p(X|Z, \mu, \sigma) p(Z|\theta) p(\theta) p(\mu|\sigma) p(\sigma), \quad (7)$$

where $p(\theta)$, $p(\mu|\sigma)$, and $p(\sigma)$ are Dirichlet, Gaussian and Wishart distributions, respectively.

Marginalizing (7) with respect to Z , μ and σ is intractable. However, through use of a VB framework, a lower bound on $p(X|\theta) = p(X|Z, \mu, \sigma) p(Z|\theta)$ can be found [8].

Denoting $\Psi \equiv \{Z, \mu, \sigma\}$, the marginal likelihood can be written as:

$$p(X|\theta) = \int p(X, \Psi|\theta) d\Psi. \quad (8)$$

$p(\Psi|\theta)$ can be approximated by a variational distribution, i.e. $q(\Psi) \approx p(\Psi|\theta)$. The Kullback-Leibler (KL) divergence of p from q is commonly used to make the evaluation (8) tractable, and it holds that:

$$\ln p(X|\theta) = D_{KL}(q||p) + \mathcal{L}(q), \quad (9)$$

where $\mathcal{L}(q)$ is the lower bound, and $D_{KL}(q||p)$ is the KL divergence. Maximizing the lower bound $\mathcal{L}(q)$ is equivalent to minimizing the KL divergence, which can be accomplished by appropriate choice of the q distributions.

The variational posterior distribution $q(\Psi)$ can be considered to factorize over the subsets $\{\Psi_i\} = \{Z, \mu, \sigma\}$, i.e.

$$q(Z, \mu, \sigma) = q_Z(Z) q_\mu(\mu) q_\sigma(\sigma), \quad (10)$$

The best distribution for each factor can be found by a free-form minimization over q_i :

$$q_i^*(\Psi_i) = \frac{\exp(\mathbb{E}_{j \neq i} [\ln p(X, \Psi)])}{\int \exp(\mathbb{E}_{j \neq i} [\ln p(X, \Psi)]) d\Psi_i}, \quad (11)$$

where $\mathbb{E}_{j \neq i}[\cdot]$ represents the expectation of the distributions $q_j(\Psi_j)$ for all $j \neq i$.

Since the variational factors are mutually coupled, they can be solved using the expectation step (E-step):

1) choose a set of initialization parameters, which are normally selected to be real, small and positive to give broad prior distributions of the variables (i.e. initially neglecting the original distributions),

2) calculate the posterior distributions in (10) through use of (11), and

3) iteratively update until the differences are smaller than some pre-selected tolerance.

After calculating the variational factors, the lower bound $\mathcal{L}(q)$ can be evaluated. An EM (expectation-maximization) optimization procedure is again adopted for the maximization of the lower bound $\mathcal{L}(q)$ (which minimizes the marginal log-likelihood $\ln p(X|\theta)$). The maximization step (M-step) for the mixing coefficients θ can be obtained, and the optimised distributions q_i are obtained through iterative updating the variational factors through the E-step until reaching the convergence criterion. In the interests of brevity, for a more in-depth discussion of the properties and application of the VB framework, the reader is directed to [8].

III. CASE STUDY

System identification via VB-AMFM is applied to a compressor on an IGT system, as shown in Fig. 1. A set of daily data (1440 points for 1440 minutes) is collected from fuel demand (input, kW) and compressor outlet pressure (output, bar) measurement readings.

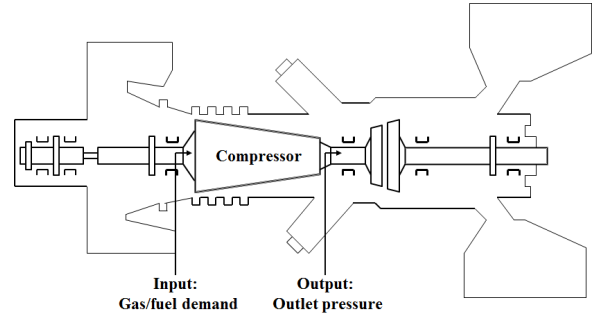


Fig. 1. Input and output sensor measurements for a compressor.

A. AMFM Results

Here, the initial Mamdani FRBS is extracted by G3Kmeans clustering algorithm [9]. The input data are further split randomly into 2 sets, with 25% for checking the performance of the model resulting from using the remaining 75% data. The predictive performances of the initial Mamdani FRBS, in terms of the root-mean-square error (RMSE) of the predicted and the actual outputs, are shown in Fig. 2, where 20 rules are selected for measurement distributions. Each input rule (distribution of the input signals) corresponds to an output rule (distribution the

output signals fall in). AMFM’s membership functions are shown in Fig. 3 showing the 20 rules for the input fuel demand and the output compressor outlet pressure. It can be seen that many input and output rules are replicated and can be removed for simplicity.

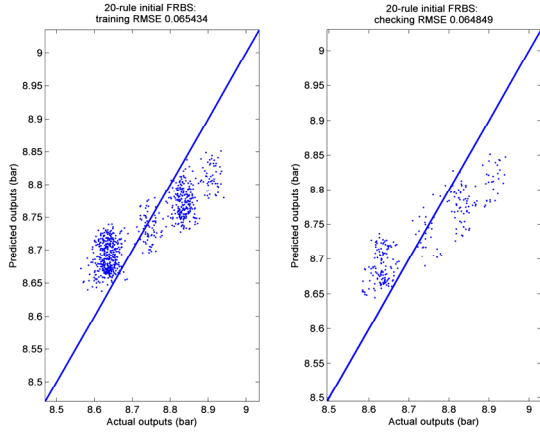
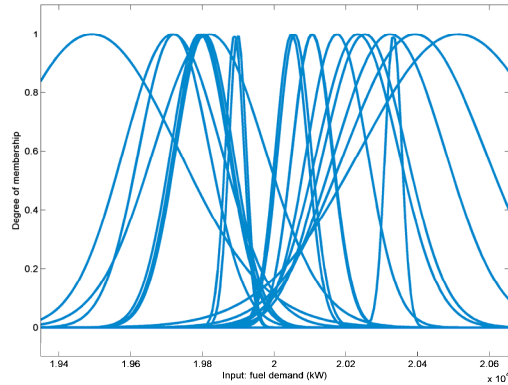
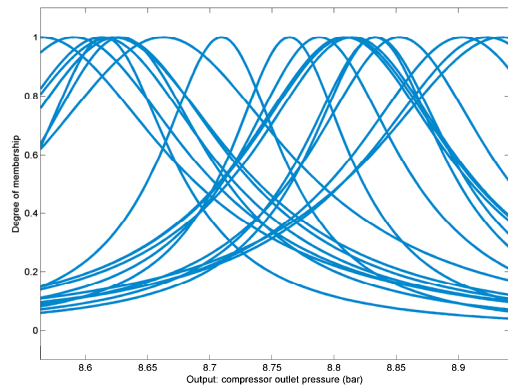


Fig. 2. Predictive performance of the initial Mamdani FRBS (20-rules)



(a)

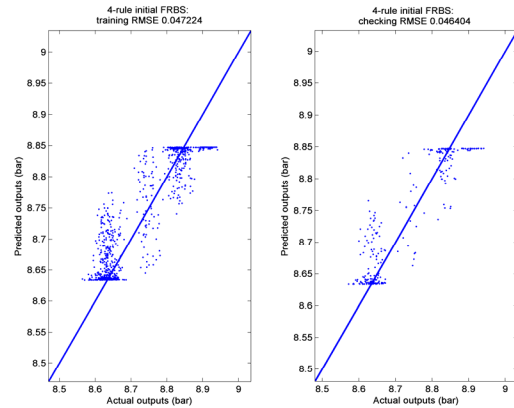


(b)

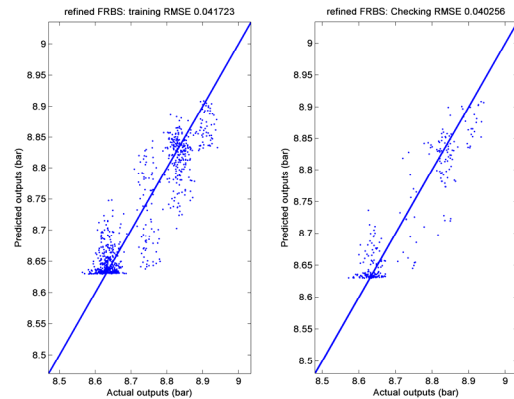
Fig. 3. Associated membership functions of AMFM for (a) Input: fuel demand of IGT and (b) Output: compressor outlet pressure.

B. VB-AMFM Results

For the same test data, the initial Mamdani FRBS is now extracted using the proposed VBGMM clustering algorithm, with number of the initial rules automatically calculated; 4 in this case. By applying to the measurements as before, the results are shown in Fig. 4(a). It can be seen that the ‘optimised’ 4-rule initial FRBS has benefited from improved performance compared to the results in Fig. 2, by presenting lower ‘training’ and ‘checking’ RMSEs. Moreover, the refined fuzzy model developed through use of the constrained BEP algorithm, as shown in Fig. 4(b), provides further improved performance. For completeness, the membership functions for VB-AMFM are given in Fig. 5. The knowledge gained from the distributions and combinations of the membership functions has demonstrated the advantageous properties of the VB-AMFM method by expressing clearer semantic meanings of its consequents, i.e. Fig. 5(b) can be considered as a simplified version for Fig. 3(b).



(a)



(b)

Fig. 4. Predictive performance of (a) the initial VB modified Mamdani FRBS (4-rules) and (b) refined VB-AMFM after BEP

IV. CONCLUSION

This paper has proposed a VB-AMFM methodology, where VBGMM is applied to remove redundant rules in the inputs in order to simplify the whole fuzzy model. The interpretation of

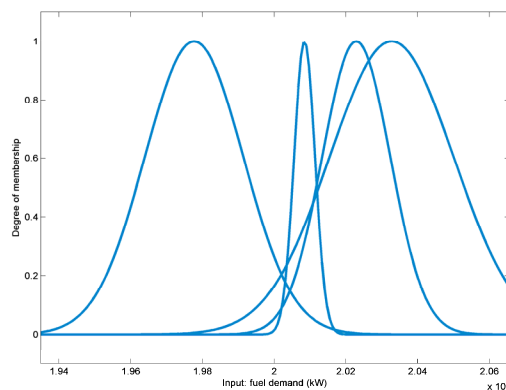
the results is shown to be clearer, while the accuracy has remained. The efficacy of the proposed method is further demonstrated through a trial on IGT compressor.

ACKNOWLEDGMENT

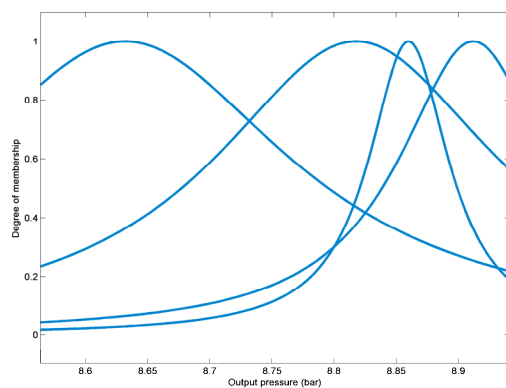
The authors would like to thank Siemens Industrial Turbomachinery, Lincoln, U.K., for providing research support and access to real-time data to support the research outcomes.

REFERENCES

- [1] Y. Zhang, C. Bingham, Z. Yang, B.W.K. Ling, M. Gallimore. "Machine fault detection by signal denoising—with application to industrial gas turbines." *Measurement*, vol. 58, pp. 230-240, 2014.
- [2] Y. Zhang, C.M. Bingham, M. Gallimore. Applied Sensor Fault Detection, Identification and Data Reconstruction. *Advances in Military Technology*, vol. 8, no. 2, pp.13-26, 2013.
- [3] Y. Zhang, C.M. Bingham, M. Gallimore. Fault detection and diagnosis based on extensions of PCA. *Advances in Military Technology*, vol. 8, no. 2, pp. 27-41, 2013.
- [4] J. Chen, M. Mahfouf. "Improving transparency in approximate fuzzy modeling using multi-objective immune-inspired optimisation." *International Journal of Computational Intelligence Systems*, vol. 5, no. 2, pp. 322-342, 2012.
- [5] Y. Zhang, J. Chen, C. Bingham, M. Mahfouf. "A new adaptive Mamdani-type fuzzy modeling strategy for industrial gas turbines." *Proc. of IEEE Int. Conf. on Fuzzy Systems*, China, July, 2014.
- [6] E.H. Mamdani, S. Assilian. "An experiment in linguistic synthesis with a fuzzy logic controller." *International Journal of Man-machine Studies*, vol. 7, pp. 1-13, 1975.
- [7] J. Chen, "Biologically inspired optimisation algorithms for transparent knowledge extraction allied to engineering materials processing", Ph.D. dissertation, University of Sheffield, U.K., 2009.
- [8] A. Corduneanu, C.M. Bishop. "Variational Bayesian model selection for mixture distributions." *Artificial Intelligent Statistics*, pp. 27–34. 2001.
- [9] J. Chen, M. Mahfouf, C. Bingham, Y. Zhang, Z. Yang, M. Gallimore. "An evolutionary based clustering algorithm applied to data compression for industrial systems". *Advances in Intelligent Data Analysis XI, Lecture Notes in Computer Science*, vol. 7619, pp.103-113, 2012.



(a)



(b)

Fig. 5. Associated membership functions of VB-AMFM for (a) Input: fuel demand and (b) Output: compressor outlet pressure.