BAYES FOR ROLLING MILLS: FROM PARAMETER ESTIMATION TO DECISION SUPPORT

Pavel Ettler * **Miroslav Kárný** ** **Tatiana V. Guy** **

* COMPUREG Plzeň, s.r.o.
P.O.Box 334, 306 34 Plzeň, Czech Republic
** Department of Adaptive Systems
Institute of Information Theory and Automation
P.O.Box 18, 182 08 Prague 8, Czech Republic

Abstract: The paper shares experience gained in application of dynamic Bayesian approach to control problems in the field of metal rolling. The contribution introduces basic notions of theory applied and provides the algorithmic as well as application-oriented solutions developed. Specifically, the consistent use of the approach resulted in an advanced decision support system for operators of a reversing cold rolling mill.

Keywords: Industrial control, probabilistic models, system identification, hypotheses, decision support systems.

1. INTRODUCTION

For several decades rolling mill control problems belong to traditional areas of interest for the control community. This paper shows that the consistent Bayesian approach spreading from system identification (Peterka, 1981) to the decision making theory (DeGroot, 1970; Berger, 1985) provides a room for significant improvements even in this well established application domain.

Bayesian concept of probability forms a firm counterpart to the standard statistics. It became rather popular during past decades especially in connection with development of algorithms of decision trees, data mining, etc. The article contributes positively to the disputation between Bayesians and their opponents by providing a long-term practical experience with the utilization of the approach.

Theoretical background and principal algorithmic solutions have been adopted from the academia. Considerable work has been done to implement these solutions on real process and to make them work on-line. Although the application domain of rolling mills is narrowed here to cold reversing ones, the most of conclusions concern the whole domain in general.

Section 2 introduces necessary basics of the Bayesian theory, which are further elaborated. Section 3 concerns numerical and software aspects. Sections 4 and 5 summarize implementation experience while Section 6 outlines open problems. Conclusions 7 provide with overall results.

2. NECESSARY FACTS FROM THE THEORY

Bayes' theorem as a statement about conditional probabilities can be written as:

$$f(a|b,c) = \frac{f(b|a,c)f(a|c)}{\int f(b|a,c)f(a|c)da}, \qquad (1)$$

where a, b, c stand for random variables, $f(a|b)$ denotes probability density function (pdf) of a

conditioned by b and $f(a, b)$ denotes joint probability distribution of a and b.

Complemented by main principles of calculus with pdfs (Peterka, 1981; Hartigan, 1983), this theorem forms a cornerstone for consequent methods of handling of uncertainty.

2.1 System model and its identification

A stochastic system with data $d_t = (u_t, y_t)$, consisting of directly manipulated input u_t and indirectly affected output y_t, observed at discrete time instants $t = 1, 2, \ldots$ is considered. The sequence of *historical data* observed on the system up to time t is denoted by $d(t) = (d_1, d_2, \ldots, d_t)$.

Within the accepted framework, the dependence of the system output y_t on the historical data $d(t-1)$ is generally described via conditional probability density function (pdf) $f(y_t | u_t, d(t-1), \Theta)$, parameterized by an unknown time invariant *parameter* Θ.

Interpretation of Θ as a random variable allows exploiting (1) for the description of uncertainty about Θ after observing data $d(t)$. The corresponding conditional pdf is:

$$f(\Theta | d(t)) = \frac{f(y_t | u_t, d(t-1), \Theta)}{f(y_t | u_t, d(t-1))} f(\Theta | d(t-1)), \quad (2)$$

where $f(y_t | u_t, d(t-1)) =$

$$= \int f(y_t | u_t, d(t-1), \Theta) f(\Theta | d(t-1)) d\Theta. \quad (3)$$

Slow changes of Θ are allowed when adopting adaptation techniques mentioned in Section 3.

2.2 Selection among set of possibilities

Applying (1) to the discrete set $\{H_i\}_{i=1}^n$ of hypotheses H_i about some properties of the inspected system, the Bayes formula gets the form:

$$f(H_i | d(t)) = \frac{f(d(t) | H_i) f(H_i)}{\sum_{j=1}^n f(d(t) | H_j) f(H_j)}. \quad (4)$$

Similarly to (2), the probability of i-th hypothesis after observing data $d(t)$ becomes:

$$f(H_i | d(t)) = \frac{f_i(y_t | u_t, d(t-1))}{f(y_t | u_t, d(t-1))} f(H_i | d(t-1)), \quad (5)$$

where $f_i(y_t | u_t, d(t-1))$ is the conditional pdf of the system output defined by hypothesis H_i.

The technique is used for estimation of the model structure (model order) and for selection of process model among pre-specified set of models.

2.3 Approximation by probabilistic mixtures

Joint pdf $f(d(t) | \Theta)$ of historical data is used for selection of the approximate description of the closed control loop. The chain rule for pdfs allows to express $f(d(t) | \Theta)$ through the models $f(d_t | d(t-1), \Theta)$. Physical nature of the process implies that current data record d_t is influenced only by a recent part of the whole process history, $\phi_{t-1} = (d_{t-1}, \ldots, d_{t-l})$ given by some finite l. This motivates an approximation of the joint distribution of data by:

$$f(d(t) | \Theta) = \prod_{t \in t^*} f(d_t | \phi_{t-1}, \Theta). \quad (6)$$

Considering process data describe multiple modes, the particular models $f(d_t | \phi_{t-1}, \Theta)$ are chosen to be multi-modal. As a probabilistic multi-modal model can be approximated by a finite mixture of uni-modal models, so-called components (Kárný, at al., 2003), the following approximation by Gaussian dynamic mixtures is accepted:

$$f(d_t | \phi_{t-1}, \Theta) = \sum_{c \in c^*} \alpha_c \mathcal{N}_{d_t}(\theta_c \phi'_{t-1}, r_c), \quad (7)$$

where $\mathcal{N}_d(m, r)$ is Gaussian pdf of d having mean m and covariance r; $'$ denotes transposition; $c \in c^*$ are pointers to particular Gaussian pdfs (components) forming the mixture; α_c is probability with which c-th component is chosen; θ_c is the matrix of unknown coefficients determining expected value $\theta'_c \phi_{t-1}$ of the c-th component. Θ collects all unknown parameters of the mixture model, i.e. $\Theta = \{\theta_c, r_c, \alpha_c, \ c \in c^*\}$.

The mixture model (7) describes flexibly a wide range of behaviors of closed control loop. This flexibility is paid by impossibility to apply the exact Bayesian estimation. However, a simple, recursively applicable, quasi-Bayesian estimation (Kárný, at al., 2003) copes well with exponentially increasing complexity of the posterior pdf. Thus, a good estimate of the mixture parameter Θ can be obtained. Consequently, the design of an advisory mixture, discussed below, can take Θ as known one and omit it formally from conditions of pdfs.

2.4 Advisory mixtures

Process operator can be taken as a part of the closed-loop and his actions are expressed by historical data (at least implicitly). His activities (changing set points and modes, etc.) within the process closed-loop can be viewed as actions of a special kind of controller. Advisory system (Kárný, at al., 2004) tries to improve closed-loop behaviour by providing advices to the operator who can either accept or reject them. Advising,

followed by the operator, can be interpreted as replacing the current operator strategy by an optimized one, which should make the closed-loop behavior closer to the desired one.

Generally, advising supposes the following steps:

(1) Recognition of the part of the closed-loop mixture model describing the operator's strategy,
(2) Mapping of operation (control) aims on the optimality criterion,
(3) Computation of the optimized recommended strategy,
(4) Presentation of the optimization results to the operator in a suitable form.

The detailed description of the approach and the essence of the solution are given by (Kárný, *at al.*, 2004) and (Kárný, *at al.*, 2003), respectively.

ad (1) Generally, the operator can influence closed-loop behaviour in two ways, by: i) changing the frequency with which the process behavior stays at different operation modes described by mixture components (7); ii) direct selecting values of some entries within the data record d_t (typically set points). Formally, in the first case, the operator modifies the estimated probabilities $\{\alpha_{c_t}\}$ of random pointers c_t to active components to another distribution $\left\{\,^{\lfloor I}\alpha_{c_t}\right\}$:

$$f(c_t, d_t|d(t-1)) = \alpha_{c_t}\mathcal{N}_{d_t}(\theta_{c_t}\phi'_{t-1}, r_c) \rightarrow$$
$$^{\lfloor I}f(c_t, d_t|d(t-1)) = {}^{\lfloor I}\alpha_{c_t}\mathcal{N}_{d_t}(\theta_{c_t}\phi'_{t-1}, r_c). \quad (8)$$

In the second case, the data, influenced by the operator, play role of system input u_t from the operator's perspective. Using the same perspective, the remaining entries of d_t form system output y_t. Thus, the second way can be formally written as replacement:

$$\mathcal{N}_{y_t}(\,^{\lfloor y}\theta_c[u_t, \phi_{t-1}]', \,^{\lfloor y}r_c)\mathcal{N}_{u_t}(\,^{\lfloor u}\theta_c\phi'_{t-1}, \,^{\lfloor u}r_c) \rightarrow$$
$$\mathcal{N}_{y_t}(\,^{\lfloor y}\theta_c[u_t, \phi_{t-1}]', \,^{\lfloor y}r_c)\mathcal{N}_{u_t}(\,^{\lfloor Iu}\theta_c\phi'_{t-1}, \,^{\lfloor Iu}r_c). (9)$$

The estimated parameters θ_c, r_c define the parameters $^{\lfloor y}\theta_c, {}^{\lfloor y}r_c, {}^{\lfloor u}\theta_c, {}^{\lfloor u}r_c$ via the chain rule applied to multivariate Gaussian distribution.

Thus, within the adopted terminology, any changing the operator's strategy means modifications of the type (8) and (9) that can be performed either sequentially or simultaneously.

ad (2) Considering closed-loop behaviour described in terms of distributions (7), it makes natural to have operation (control) aims expressed in terms of distribution (namely *target distribution*) as well. The design of the advisory system is then formulated as minimization of the Kullback-Leibler divergence (Kullback and Leibler, 1951) of distribution, describing closed-loop behaviour,

from the target distribution. The minimization is performed over the optional $^{\lfloor I}$-elements in (8) and/or (9). This *fully probabilistic design* was developed in (Kárný, *at al.*, 2003; Kárný, *at al.*, 2004). For Gaussian mixtures and Gaussian target distributions, it is approximately solvable. The resulting algorithms operates on kernels of quadratic forms in a way similar to discrete Riccati equation extended to mixtures and quadratic criteria (Böhm and Kárný, 2001).

ad (3, 4) Inserting of the obtained minimizing distribution into the closed-loop description gives *advisory mixture*, which supposes the fully cooperating operator. Graphical presentation of the marginal distributions of optional variables should stimulate operator's actions. Understanding the probabilistic nature of advising is not required for the operator, as he gets rather simple instruction: *Try to reach the highest hill of those displayed.*

In complex situations, it is reasonable to measure divergence of the current closed-loop description from the target permanently and alert the operator only when some threshold is crossed. Also it is necessary to select the most critical variables to be modified under the given situation. It can be done using a slight extension of the methodology sketched above (Kárný, *at al.*, 2004).

3. MAKING IT FUNCTIONAL

The accumulated experience showed that adequacy of the underlying models, reliability of the resulting algorithms and implementation ease are the dominant aspects determining usability of "academic" solutions. These aspects are touched here.

Formulas outlined in the previous sections produce usable solutions for Gaussian pdfs and their mixtures (7). Locally linear models in (7) are sufficient, if changes of their parameters, caused by aging and by shifts of working range, can be tracked. This adaptation is reached by considering standard forgetting of obsolete information. The stabilized forgetting technique (Kulhavý and Zarrop, 1993) is used as the most reliable one.

Quasi-Bayesian estimation with Gaussian mixtures reduces to manipulations with positive definite, extended information matrices. They are recursively updated similarly as in the well established LG area (Linear model, Gaussian probability distributions).

Numerical problems arising from number representation and limited precision in real computers have been overcome by updating the lower-triangular-diagonal decompositions of these information matrices (Bierman, 1977).

The fully probabilistic design in Gaussian-mixture setting is rather close to the well known LQG control design (LG model with quadratic design criterion) and the similar numerical problems as in the learning are faced. Again, the use of factorized kernels of quadratic forms, making the essence of the design (Kárný, at al., 2003), is the "must" guaranteeing the reliable solution.

Implementation ease is strongly influenced by the transfer of algorithmic solutions. At the beginning of cooperation, the academic algorithms had to be re-coded completely. This way is adequate for very particular cases only. More complex solutions have to exploit convergence of the software platforms involved and has to be supported by well specified programming rules as well as specific software tools allowing the transfer with minimum re-coding (Nedoma and Andrýsek, 2003).

During two decades of continuous development particular modules of the presented approach have been implemented on several hardware platforms (former Intel's Multibus system, Škoda proprietary system and the industrial PC) and under various operating systems ranging from former Intel's iRMX to RT Linux for real-time applications and several versions of MS Windows and Linux for the others.

4. VERIFIED SOLUTIONS

4.1 Structure estimation

Selection of the proper process model proved to be basic task when the idea to utilize the approach for the adaptive thickness control arose in the middle of 1980s. Model had to be simple enough to allow real time estimation on computers industrially available. It strongly limited the number of parameters estimated on-line. At the same time, the model had to be descriptive enough: the transport delay, which is characteristic of the thickness measuring system does not allow fast changes of parameter estimates. Therefore, a substantial effort had been devoted to proper choice of the most significant parameters.

Utilized computer-aided model selector coming from academia was based on testing of hypotheses (2.2). It had to be boosted by signal reliability analysis and by experience from the field.

Several settled model versions are available nowadays differing according to the type and parameters of a given rolling mill. All proven models have an absolute term modelling non-zero offset. Their data vectors include some of the key process variables or their combinations: input and output thickness deviations, tensions and speeds, screw-down or hydraulic actuator position, rolling forces and the speed ratio.

Classification techniques has been used occasionally to contribute to other solutions such as mill stretch compensation and control of the roll-positioning system. In the latter case, the approach helped to reject the model based control in favor of a modified PID control.

4.2 Parameter estimation

Considering a rolling mill, parameter estimation serves mainly to model-based thickness controller. After structure estimation of the process model for the particular mill, initial values of parameter estimates and their admissible ranges had to be stated. Flat prior pdfs with the mean in the middle of the range proved to be good initials as the process is excited enough to cause fast convergence of the estimation. Cautious controller is used, so that both initial moments (mean and variance) of parameters are being used for its evaluation. Several "details" have made the solution work properly:

- The coefficients of the adaptive controller start to be recalculated after stabilization of the estimates. This guarantees smooth start.
- The estimated model offset is not used for control design. The output of a robust feedback controller is being used instead.
- Exceeding of given limits for parameter estimates causes reset of the estimation.
- The estimation is being switched off in cases when its outputs could be deteriorated, e.g. for welded strip sections.

Recursive estimation algorithm is being triggered by the strip movement, for each about 50 mm of the strip length. The exact value of the triggering distance depends on the particular mill.

Starting with the first application in 1986 (Ettler, 1986), the Bayes-based thickness controller has been installed standardly for various types of 20-high and four-high reversing cold rolling mills with very good long term performance.

5. NOVEL SOLUTION

A qualitative advance on utilization of the theory has been made by implementation of a decision support system (DSS) for operators of a fine rolling mill. The advisory system outlined in Sections 2.3 and 2.4 which forms kernel of the DSS was brought to practical application in the framework of the project ProDaCTool and it was commissioned in 2002 (Quinn, et al., 2003).

5.1 Mixture estimation

Off-line mixture estimation of huge amounts of process data describing the good rolling was accomplished. Resulting 30 mixture sets approximately describe process behavior for various working conditions and material types. Static mixtures were employed for the given case. Fig. 1 shows an example of 10-dimensional mixture and its composition formed by Gaussian components, both projected into two-dimensional space.
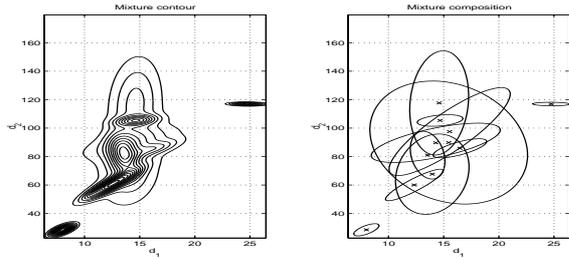
Fig. 1. Composition of a static mixture formed by Gaussian components – two dimensional projection for data channels d_1 and d_2.

Estimated mixtures are employed within the on-line DSS according to working conditions and material being processed. Learning ability was introduced into the system by an independent on-line run of quasi-Bayesian algorithm, which modifies the estimates by incoming actual data. When the coil of the same material is processed later on, the updated mixture estimate, often connected to the specific coil pass, is used for advising and also as an initial estimate in a new independent learning, see Fig. 2.
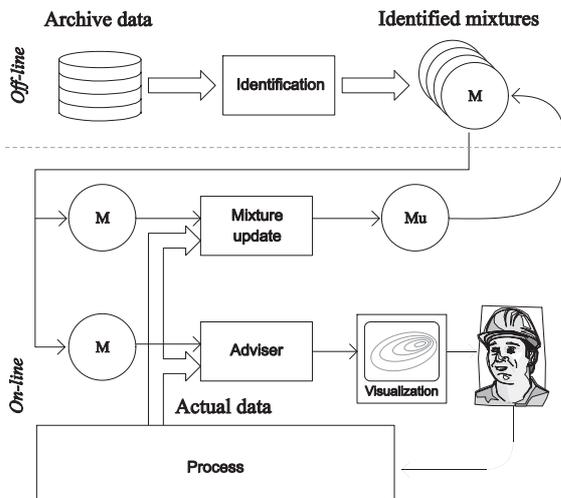
Fig. 2. Diagram of the DSS. Symbol M means an identified mixture. M_u is the updated mixture which will be used for the next relevant pass.

5.2 Decision support

For a given rolling mill, the aim of the implementation was to stabilize the highest possible product quality and to allow better utilization of possible rolling speed range. Concerning product quality it had been more difficult to tune strip tensions for higher speeds therefore lower speeds had been preferred.

Ten key data channels were selected for processing while three of them (rolling speed, input and output tensions) allowed direct influencing of the process. Target pdf specifying the optimality criterion was derived from the main quality marker – deviation of the output strip thickness (Ettler P., *at al.*, 2000).

Depending on particular working conditions, an appropriate estimate from the set of the off-line results is loaded at the beginning of operation. Incoming data are then being used for calculation of advices. Fig. 3 shows an example of the 2-dimensional projection of the advisory mixture originated from the off-line estimate and conditioned by current data.
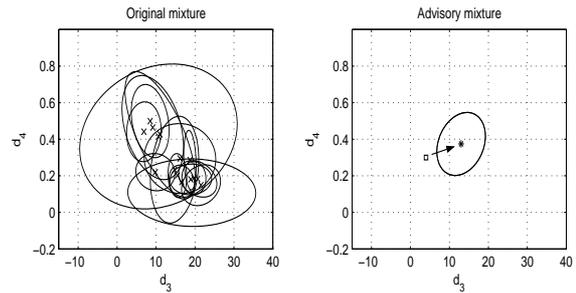
Fig. 3. An advisory mixture made from the original mixture using actual data – two dimensional projection for data channels d_3 and d_4. Symbol □ stands for the actual working point which should be moved to the location of ∗.

Generation of advices is based on evaluation of the distance between the optimal and actual working points within the multidimensional space. Recommendations are provided for the operator through a traffic light-like indicator together with a simple text such as: "Increase input tension to x", where x means the recommended value. Smoothly animated graphical representation of the situation is provided at the same time.

For the assessment of the project achievements, production data from the period of 12 months before installation were compared with the period of 3 months of the full operation. The system helped to find the optimal combinations of adjusted values. Results indicated that the quality markers were stabilized and improved by about 7% in average and the rolling speed being chosen

by operators increased by 17 – 30% depending on particular working conditions.

6. CHALLENGING PROBLEMS

A potential utilization of hypothesis testing approach, Section 2.2, arises for the elimination of roll eccentricity. In this case, typical terms in the model equation are represented by goniometric functions. As the ranges of possible disturbance frequencies, made by the eccentricity, are relatively narrow and deducible from diameters of rolls, the set to be classified is endurably small. Under these prerequisites, the compensator, based on Bayesian model selection, should be able to compete with a typical fast-Fourier-transform approach to this problem (Rath, 2000). It was verified by a case study but it was not implemented yet.

Fully probabilistic description of the process is intended to improve prediction of the outgoing strip thickness within the rolling gap to revitalize one of the classical thickness control principle.

Concerning the decision support area, the running research is devoted to dynamic aspects, multiple criteria and distributed solutions. Moreover, much is to be done to make application of the theory really straightforward.

7. CONCLUSIONS

The paper scans results of a long term cooperation of academic and application teams related to various applications of Bayesian paradigm to reversing rolling mills. Specifically:

On-line estimation of process model parameters is being applied for two decades for the thickness control. The adaptive controller turned out to cope well with diversity of working conditions and types of material being processed.

Testing of hypotheses is being utilized off-line to the proper model selection for particular control problems.

Novel utilization of the approach is represented by the decision support tool for process operators. The on-line advisory system was applied for the 20-high cold rolling mill with satisfactory results.

The results are important on their own but they also emphasize importance of the chain theory→algorithm→software→implementation.

Acknowledgements

REFERENCES

Berger, J.O. (1985) *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York.

Bierman, G.J. (1977). *Factorization Methods for Discrete Sequential Estimation.* Academic Press, New York.

Böhm, J. and M. Kárný (2001). Quadratic adaptive control of Gaussian mixtures. In J. L. Martins de Carvalho, editor, *Proceedings of the European Control Conference ECC'01*, Porto, Faculdade de Engenharia da Universidade.

DeGroot, M.H. (1970). *Optimal Statistical Decisions.* McGraw-Hill Company, New York.

Ettler, P. (1986). An Adaptive Controller for Škoda twenty-roll cold rolling mills. In: *Proceedings of 2nd IFAC Workshop on Adaptive Systems in Constrol and Signal Processing*, Lund Institute of Technology, pp. 277–280.

Ettler, P., M. Valečková, M. Kárný, I. Puchr (2000). Towards a knowledge-based control of a complex industrial process. In: *Proceedings of the 2000 American Control Conference*, Chicago, AACC/Omnipress, pp. 2063–2066.

Hartigan, J.A. (1983). *Bayes Theory.* Springer series of statistics, New York, Springer – Verlag.

Kárný, M., J. Böhm, T. V. Guy, P. Nedoma (2003). Mixture-based adaptive probabilistic control. In: *International Journal of adaptive control and signal processing.* John Wiley & Sons, Inc., 17 [2], pp. 119-132.

Kárný, M., J. Böhm, T.V. Guy, L. Jirsa, I. Nagy, P. Nedoma, and L. Tesař (2004). *Optimized Bayesian Dynamic Advising: Theory and Algorithms.* Springer, London. to appear.

Kulhavý, R. and M. B. Zarrop (1993). On general concept of forgetting. *International Journal of Control*, 58(4):905–924.

Kullback, S. and R. Leibler (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–87.

Nedoma, P. and J. Andrýsek (2003). *Mixtools. Application Program Interface. User's Guide.* Technical Report 2088. ÚTIA AV ČR. Praha.

Peterka, V. (1981). Bayesian approach to system identification. In: *Trends and Progress in System Identification* (P.Eykhoff, ed.). IFAC Series for Graduates, Research Workers and Practicing Engineers 1, Pergamon Press, pp. 239–303.

Rath, G. (2000). *Model Based Thickness Control of the Cold Strip Rolling Process.* Doctoral Thesis, University of Leoben.

Quinn, A., P. Ettler, L. Jirsa, I. Nagy and P. Nedoma (2003). Probabilistic advisory systems for data-intensive applications. In: *International Journal of adaptive control and signal processing.* John Wiley & Sons, 17 [2], pp. 133-148