Employing Information Hidden in Industrial Process Data

Pavel Ettler[1], Miroslav Kárný[2], Ivan Nagy[2]
[1]COMPUREG Plzeň, s.r.o.
P.O.Box 334, 306 34 Plzeň, Czech Republic
ettler@compureg.cz
[2]Institute of Information Theory and Automation AV ČR
P.O.Box 18, Praha 8, Czech Republic
school@utia.cas.cz, nagy@utia.cas.cz

Abstract

The paper refers to an intermediate stage of the running international project ProDaCTool which aims to develop the decision-support tool for operators of complex industrial processes. It should help to guarantee stable highest quality of product by optimal settings of machinery.

The main idea of the project is to extract valuable information from a huge amounts of process data and to store it in the form of multi-dimensional mixtures of probability density functions. Resulting mixtures will be marked as "good" or "bad" according to selected quality criteria. Then, the system will advise the operator to optimize trajectory of the working point relative to distance from "good" parts of the data space. Solution of the whole task has been based on Bayesian approach.

A cold rolling mill was chosen for experiments and final implementation of the advisory system.

1 Introduction

Present-day industrial processes are controlled by more or less sophisticated control systems which are able to guarantee required performance of plant or machinery according to the settings of many adjustable parameters. In many cases relation between the settings and quality of the product is not transparent and results are too much dependent on skills and experience of operators [8].

Information hidden in process data could potentially support human decisions about parameter settings. An advisory system – when successfully activated – could distinguish specific clusters in multi-dimensional data space and evaluate them according to given quality criteria. High dimensionality, signal noises, huge amount of superfluous or idle data, potential sensor malfunctions, dynamic and stochastic attributes of the process make the task extremely difficult [2, 3].

The idea of the project we refer about consists in transformation of information being mined into mixtures of probability density functions [10] that are well described by a fixed set of well estimated parameters. Bayesian approach [9, 4] has been applied for that purpose.

Mixture estimation is known to be sensitive to initialization. This property is enhanced by unusual dynamic context considered and by employed quasi-Bayes estimation procedure that is adopted because of their inherent adaptability. For this reason, various advanced initialization are under development [5]. They look promising on simulated data. Their heuristic nature, however, does not guarantee success in real cases. This has motivated series of test on real data taken from the target application, namely, a cold-rolling mill.

Carefully selected sufficiently low-dimensional data records were chosen for tests and comparisons with empirical probability density functions constructed on multi-dimensional grid of relative frequencies. Use of these for initialization has been tried at the same time to help to start high-dimensional case by extending low-dimensional marginals.

Extensive experiments have been performed on considerable amounts of rolling mill data afterwards. Comparison of passes considering quality of the product was selected for illustration.

Underlying theory is briefly described in Section 2. Experiments (Sect. 4) have been performed on data (Sect. 3.3) collected on the target process (Sect. 3). They lead to conclusions summarized in Sect. 5.

2 Underlying theory

2.1 Basic scenario

The operator handling the industrial process, see Section 3, deals with a sequence of multivariate data records $d(\mathring{t}) \equiv (d_1, \ldots, d_{\mathring{t}})$ till the time horizon $\mathring{t}$. He takes actions $a(\mathring{t}) \in a^*(\mathring{t})$ and observes innovations $\Delta_a(\mathring{t})$, i.e. $d_t = (\Delta_{a_t}, a_t)$. Quality of his actions are judged according to quality criterion that reflects probability distribution of decisive variables of the process, see Subsection 3.2. Experienced operators reach this goal but the quality reached by inexperienced ones is worse. These modes of process management manifest themselves in multi-modality of the distribution of all data records.

The advisory system estimates this multi-modal distribution and identifies the mode whose marginal distribution on decisive variables is closest to the required one. This is done in off-line regime using vast amount of available data. In on-line advising mode, the operator is guided to keep *all* data nearby this successful mode. Thus, the design of the advisory system consist of the following basic steps:

- multiple mode distribution generated by time-invariant mixture model is fitted to historical data, see Subsection 2.2,
- modes are qualified according to the quality criterion of interest and the best one(s) are selected for use in on-line regime,
- operator is guided in on-line regime to the best high dimensional mode by showing him which variables are out of the currently recommended range.

2.2 Mixture and its estimation

This paper deals with the first and most important item of the above list. Basic facts relevant to it are summarized here.

Joint probability density function (pdf) $f(d(\mathring{t}))$ of all data is the most general description of process managed by the operator. Historical data serve us for fitting a version of this pdf $f(d(\mathring{t})|\Theta)$ parameterized by time invariant parameters $\Theta \in \Theta^*$ to the considered process. Chain rule for pdfs [9] implies that $f(d(\mathring{t})|\Theta) = \prod_{t \in t^*} f(d_t|d(t-1), \Theta)$. Thus, we can specify $f(d(\mathring{t})|\Theta)$ through the "local" models $f(d_t|d(t-1), \Theta)$. Physical nature of the process implies that data record d_t is influenced by a finite history $\phi_{t-1} = (d_{t-1}, \ldots, d_{t-l})$ given by some, usually small, l. The models $f(d_t|\phi_{t-1}, \Theta)$ have to be multi-modal ones from a rich, widely applicable class. Normal dynamic mixtures are practically the most significant class of this type. They have the form

$$f(d_t|\phi_{t-1}, \Theta) = \sum_{c \in c^*} \alpha_c \mathcal{N}_{d_t}(\theta_c \phi_{t-1}, r_c) \text{ where} \quad (1)$$

$\mathcal{N}_d(m, r)$ is normal pdf of d having mean m and covariance r,
$c \in c^* \equiv \{1, \ldots, \mathring{c}\}$ are pointers to various normal pdfs (components) forming the mixture,
α_c is probability with which c-th pdf is chosen,
θ_c is the matrix of the regression coefficients of c-th component,
Θ collects all unknown parameters of the mixture $\theta_c, r_c, \alpha_c,\ c \in c^*$ of the model.

Likelihood function $\prod_{t \in t^*} f(d_t|\phi_{t-1}, \Theta)$ relating the observed data $d(\mathring{t})$ to the unknown parameters Θ is a product of sums and it contains a huge amount of terms. As such, it cannot be evaluated exactly and approximate estimation is needed. Classical EM algorithm [10, 6] or its on-line applicable modification called quasi-Bayes estimation [4] serve to this purpose. Both use the same "trick". The current estimates of parameters are used for estimating the probability with which the data are generated by specific components. These probabilities are used as weights of data for updating of parameter estimates. For normal mixture, the approximate estimation reduces to weighted least squares running in parallel for each component. In maximum likelihood set up, it provides point estimates of Θ. In the adopted Bayesian interpretation, it evolves statistics determining posterior pdfs on parameters. In both cases, selected structure of the model and initial conditions decide on success or failure of the estimation. A novel approach to addressing this problem has been developed recently [5] for constructing prior pdf $f(\Theta)$ to be updated. It is enhanced and judged by experiments performed on real data in Section 4.

3 Industrial process

A fine twenty-high (Sendzimir-type) cold rolling mill was selected for experiments and for the final implementation of the advisory system. The rolling mill processes strips made from copper, nickel and iron alloys with typical nominal input thickness of about 1.5 mm and width of 300 mm. The thickness of the strip is reduced stepwise during particular passes resulting typically in output thickness of about 0.2 mm with $\pm 5\ \mu$m (microns) tolerance range.

3.1 Control system

The rolling mill is governed by a distributed control system with separated regulators of electric drives, roll-positioning system and adaptive AGC (automatic gauge control). The most advanced part of the control system created by the AGC controller, operator's and supervisor's workstations as well as the database server are interconnected by an industrial network [1].

One of the programs running on the database server distinguishes rolling mill states and initiates pass-dependent archiving of process data which are continuously collected from the process sensors and AGC. Data acquisition is triggered by the strip movement after each 0.04 m of the strip length. The system is capable to store every data sample and so full amount of information available for the control is also at disposal for further off-line processing.

3.2 Quality criterion

Variation of the output thickness H_2 is one of the main quality measures for processed metal strips. Statistical coefficient C_p was selected as a criterion for basic quality evaluations:

$$C_p = \frac{tol_{h_2}^{+} + |tol_{h_2}^{-}|}{6\,\sigma_{H_2}}\ , \text{ where} \tag{2}$$

H_2 denotes output thickness, h_2 is its deviation from the nominal value H_{2nom},
$tol_{h_2}^{+}$, $tol_{h_2}^{-}$ are boundaries of tolerance range of h_2,
$\bar{H}_2$, σ_{H_2} are mean and standard deviation of the output thickness H_2 respectively.

The control aim is to maximize C_p.

3.3 Process data

Each data file corresponds to a particular pass while a single coil is processed within 5-10 passes. Typical data file contains up to tens of thousands of samples consisting of 46 signal channels. As the system produces tens of files per day successive processing have to be limited to reasonable extent. Thus the first task consisted in "manual" selection of representative data files covering rolling under various conditions.

4 Experiments

Matlab software was chosen as a common environment for experiments. Data files had to be converted into Matlab format, combed and re-structured to be usable for each project partner. Files were assorted according to differences in quality afterwards.

4.1 Mixture estimation

After basic experiments with simulated data real records were taken into consideration. High dimensionality of the problem makes evaluation of results rather difficult. To allow clear principal depiction of results in this paper four significant data channels were selected from original data matrix:

1. Strip tension on the input side;
2. Strip tension on the output side;
3. Ratio of electric currents of the input and output coiler drives;
4. Electric current of the main mill drive.

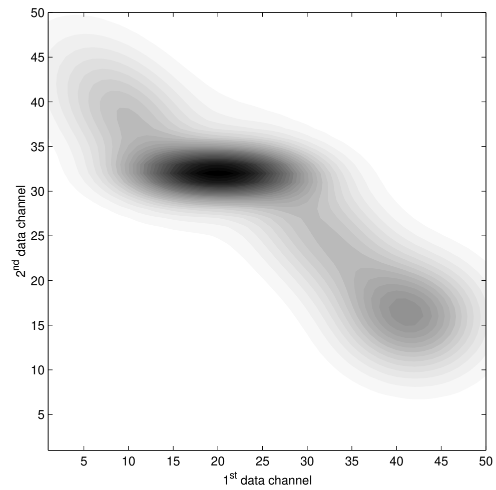

Figure 1: Filled contour plot of distribution of probability density for the 1st and 2nd data channels

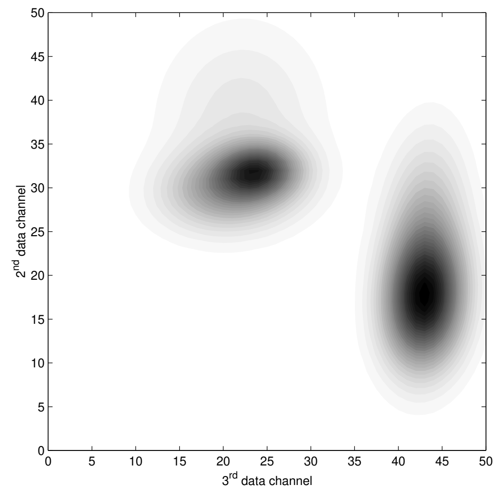

Figure 2: Filled contour plot of distribution of probability density for the 3rd and 2nd data channels

The first 2-D projection of the resulted 4-D subspace of pdfs is shown on Fig. 1. Overlapped data

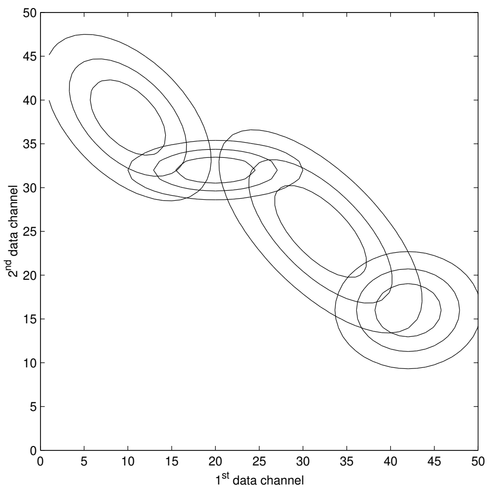

Figure 3: Contour plot of components of estimated mixtures for 1st and 2nd data channels.

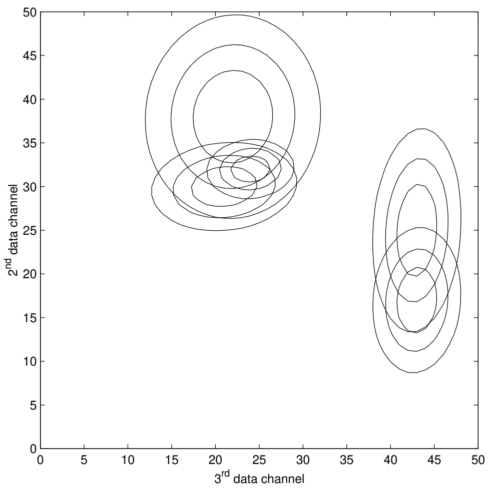

Figure 4: Contour plot of components of estimated mixtures for 3rd and 2nd data channels.

clusters were transformed into mixtures with congruent overlapping. This makes the assignment of quality value to particular mixtures indefinite. Apparently constriction to two dimensions – even if properly selected according to importance of signals – cannot provide sufficient information. Investigation in the third dimension led to usable result as shown on Fig. 2. Here two separable shapes can be recognized when the 3rd channel was taken into consideration. Both 2-D projections are shown succinctly on contour plots on Figs. 3, 4. Here composition of mixtures is clearly depicted.

4.2 Verification

Verification of results of mixture estimation for low-dimensional case can be easily performed. For this purpose densities of original data points were evaluated. Data space was divided by an orthogonal grid and relative frequencies of data occurrences were evaluated for every small subspace. Fig. 5 depicts

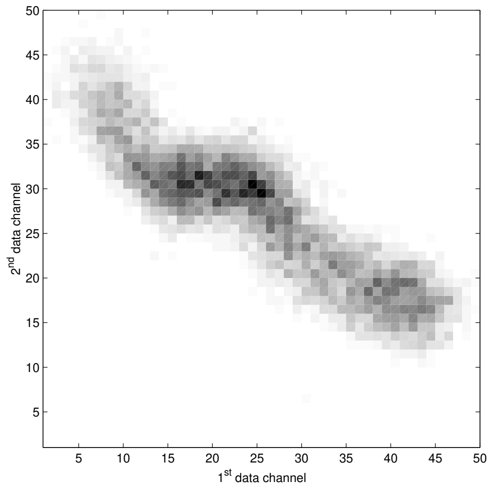

Figure 5: Pseudocolor plot of distribution of relative frequencies in the 50×50 class grid for the 1st and 2nd data channels

projection of relative frequencies for 1st and 2nd data channels evaluated for 50×50 grid where the darkest gray corresponds to the highest relative frequency of occurrences.

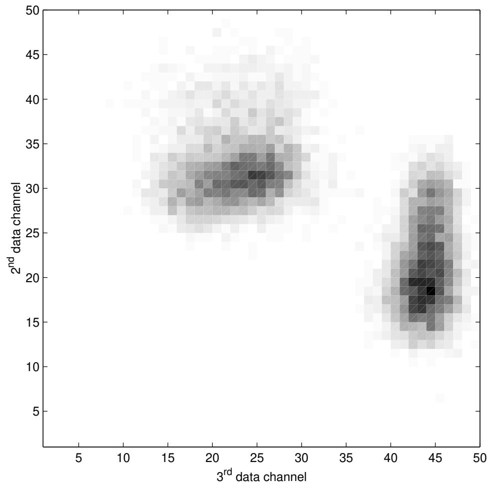

Figure 6: Pseudocolor plot of distribution of relative frequencies in the 50×50 class grid for the 3rd and 2nd data channels

Similarly Fig. 6 corresponds to projection for the

3^{rd} and 2^{nd} data channels. Simple comparison of figures 1, 5 and 2, 6 indicate that estimation was successful in this case. Although this verification method is not suitable for high dimensions it constitutes beliefs in the used approach.

4.3 Considering quality criterion

Extensive experiments were performed with rolling mill data to check out the method. A particular case was selected for illustration in this paper:

The quality criterion C_p (Eq. 2) was evaluated for data files arranged according to the pass numbers. Although C_p should continuously increase for increasing pass number, some cases were found for higher passes where the increment of C_p for successive passes was lower then expected. The effect was too slight to be normally observed by operators as for higher passes the output thickness deviations are safely within the tolerance limits. Moreover in the situation where dispersion of the output thickness has become very small thanks to proper actions of AGC during preceding passes, other influences such as imprecision in coordination of electrical drives arise relatively.

An example of such situation is depicted on Fig. 7. Comparison of mixtures of pdfs is shown in

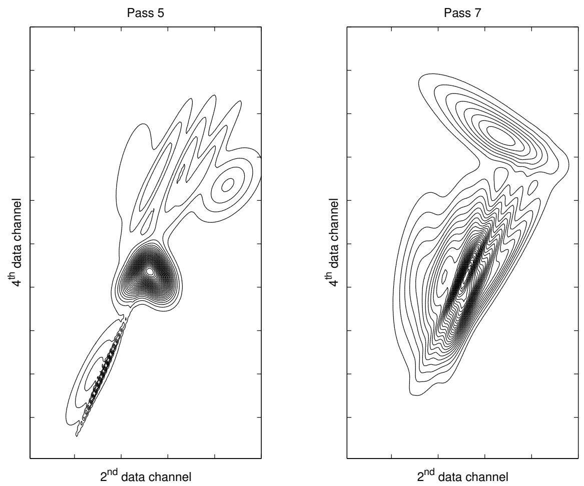

Figure 7: Comparison of 2-D mixtures corresponding to the 5^{th} and 7^{th} passes.

two dimensions (strip tension on the output side vs. electric current of the main mill drive) for the sake of clarity. Almost 10^6 samples for each pass were taken into consideration while odd and even passes had to be handled separately for technological reasons. The 5^{th} and 7^{th} passes are investigated here because for the 7^{th} pass the increment of C_p was noticeably low. It was traced that the "hill" in the most upper part of the right plot is shady in this respect. Therefore in this particular case operators would be advised to avoid settings which produce data in this area.

It is necessary to emphasize that the project aspires to cope high-dimensional cases where interrelations are difficult to reveal.

4.4 Mixtures vs. actual working point

Resulting system is assumed to include a graphical interface to operators which, in principle, will project the actual working point of the process into the estimated mixtures as shown on Fig. 8 for 2-D case. Besides, the system will generate simple advices to the user to move the working point to the appropriate location along a suitable trajectory.

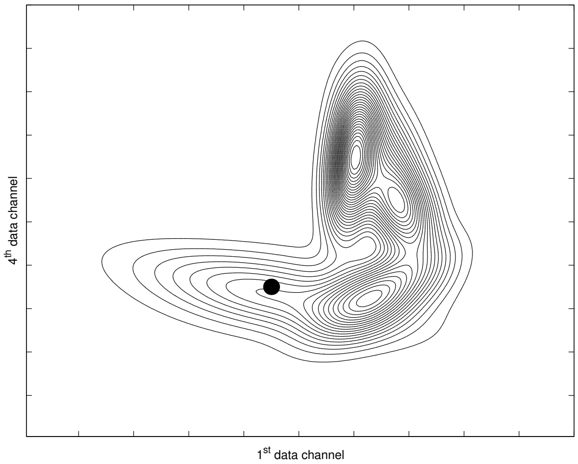

Figure 8: Projection of the actual working point into the 2-D mixture of pdfs.

5 Conclusions

The project aims to transform valuable information hidden in huge amount of process data into mixtures of probability density functions. The information should then be used to support decisions of process operators concerning optimal settings of complex machinery. The project is in the stage of development of new algorithms and off-line processing of data from a cold rolling mill. The project should be completed in 2002 by a full scale industrial application.

6 Acknowledgements

This work has been supported by the grant IST-1999-12058 ProDaCTool of the EC and by the national grant GACR 102/99/1564.

References

[1] P. Ettler, F. Jirkovský: Digital controllers for ŠKODA rolling mills. In: Advanced Methods in Adaptive Control for Industrial Applications, Ed.: Lecture Notes in Control and Information Sciences – 158, Springer – Verlag, 1991, pp. 31–35.

[2] P. Ettler, J. Štika: A Trend towards Mining of Information from Industrial Process Data. In: Preprints of the 3rd European IEEE Workshop CMP'98, Prague, Czech Republic, September 1998, pp. 243–246.

[3] P. Ettler, M. Valečková, M. Kárný, I. Puchr: Towards a Knowledge-Based Control of a Complex Industrial Process. In: Proceedings of the American Control Conference 2000, Chicago, USA, June 2000.

[4] M. Kárný, J. Kadlec, E. L. Sutanto: Quasi-Bayes Estimation Applied to Normal Mixture. In: Preprints of the 3rd European IEEE Workshop CMP'98, Prague, Czech Republic, September 1998, pp. 77–82.

[5] M. Kárný, P. Nedoma, I. Nagy, M. Valečková: Initial description of multi-model dynamic models, In: Preprints of ICANNGA00, 2001, submitted.

[6] I. Nagy, M. Kárný, P. Nedoma: Factorised EM algorithm in mixture estimation, In: Preprints of ICANNGA00, 2001, submitted.

[7] M. Rudemo: Empirical choice of histograms and kernel density estimators. In: Scand. J. Statist., 9, 1982, pp. 65-78.

[8] E. L. Sutanto: Intelligent Reasoning for Complex Processes using Multivariable Cluster Analysis. In: Preprints of the 3rd European IEEE Workshop CMP'98, Prague, Czech Republic, September 1998, pp. 69–76.

[9] V. Peterka: Bayesian system identification, In: P. Eykhoff: "Trends and Progress in System Identification", 1981, 239–304, Pergamon Press, Oxford

[10] D. M. Titterington, A. F. M. Smith, U. E. Makov: Statistical Analysis of Finite Mixtures, John Wiley & Sons, Chichester, 1985, ISBN 0 471 90763 4