2017 21st International Conference on Process Control (PC)
June 6–9, 2017, Štrbské Pleso, Slovakia

Probabilistic Advisory System for Operators Can Help with Diagnostics of Rolling Mills

Ivan Puchr
COMPUREG Plzeň, s.r.o.
Pilsen, Czech Republic
puchr@compureg.cz

Pavel Herout
Department of Computer Science and Engineering
University of West Bohemia, Pilsen, Czech Republic
herout@kiv.zcu.cz

***Abstract*—Advisory system for operators of complex industrial processes has been developed and improved by an international team of scientists and people from industry since 2000. Main purpose of the advisory system is to help operator set up manually adjustable parameters of an industrial process, with the aim to reach required production quality. Industrial process is taken for a stochastic process and input signals of its control system are taken for random variables. Based on Bayesian probability theory, a software toolbox was created for handling mixtures of probability density functions describing behavior of the process. Advisory system was tested and pilot application was installed on rolling mills producing metal strips.**

During the tests, an idea emerged to exploit verified probabilistic approach for complicated diagnostic tasks too. This diagnostics is intended for recognition of process malfunction which cannot be easily revealed by analysis of particular single signals only but analysis in multidimensional data space must be involved instead. Main principle of the advanced diagnostic method consists in finding a representation of process behavior in a short history by a mixture of probability density functions called *historical mixture*. Process behavior in the latest time period is represented by *actual mixture*. Difference between historical and actual mixtures is evaluated by calculation of Kullback-Leibler divergence. Mixtures and divergences are calculated repeatedly in time and a big change in the divergence value can be used as a source of alarm for non-standard process behavior.

***Keywords*—advisory system; Bayesian statistics; mixture of probability density functions; Kullback-Leibler divergence; diagnostics; rolling mills**

I. Introduction

Probabilistic approach to the solution of control and diagnostic problems in the field of complex industrial processes has been elaborated by several cooperating international teams of people from academia and industry since 2000. Development of theoretical background, software realization and initial applications in industry have been supported by several national and international grants. Let us name the first and the last at least: ProDaCTool (Decision Support Tool for Complex Industrial Processes based on Probabilistic Data Clustering, ST-1999-12058) and ProDisMon (Probabilistic distributed industrial system monitor, E!7262). Author 1 of this article participated in all of these projects.

One output of the mentioned research and development projects is a probabilistic advisory system. The advisory system is intended for operators of complex industrial processes. The process is controlled by several local controllers but the operator still has to adjust a set of global parameters manually. The proper setting of these parameters influences the quality of production. The aim of the probabilistic advisory system is to help the operator set the parameters optimally from the production quality point of view, [1].

Probabilistic approach consists in viewing the examined process as a stochastic one, and input and output signals of system controlling the process as random variables. Values of particular signals are acquired in time and archived. Each archived record of values of particular signal channels is taken for a realization of a multi-dimensional random variable. For simplicity, we pretend that the process operates in a particular production mode. With a rolling mill, e. g., we can imagine this production mode as a rolling of metal strip in one direction, with reduction from a particular thickness to another one, with a stable speed, etc. Each signal channel usually varies around its mean in this situation, and thus the realizations of the random variable form a cluster of points in a region of the multi-dimensional data space. Points occur in separate parts of the region with different probabilities.

If we can find a mathematical representation of the probability density function (pdf) corresponding to the occurrences of points in the data space, we can then compare these pdfs and make calculations with them. The simplified main idea of the developed probabilistic advisory system is then as follows:

- From history data recordings, we select the time periods that meet a requested criterion (production quality, e.g.) and then we find a representation of these data in the form of a pdf.
- We find a pdf representing data points acquired during the latest time period of actual production.
- We compare the pdf from history with the pdf from actual production and generate recommendations for the operator on how to adjust the actual working point of the process to get the actual pdf as "near" as possible to that one generated from history data.

- Then, the adjusted working point ensures with high probability that the production will meet the requested criterion, i.e. will be of proper quality.

II. Mixtures of Gaussian Functions

During the development of the advisory system, mixtures of Gaussian functions were selected for the probability density functions representing the occurrences of data points. Fig. 1 presents an example of a simple mixture. The mixture consists of two bivariate Gaussian functions for two random variables X_1, X_2 defined according to [2] as follows:

$$f_{X_1,X_2}(x_1,x_2) =$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)} \qquad (1)$$

where $\sigma_1, \sigma_2 \in R$ are standard deviations, $\mu_1, \mu_2 \in R$ are means, $\rho \in R$ is correlation coefficient and
$\sigma_1 > 0$ *and* $\sigma_2 > 0$ *and* $|\rho| < 1$.

Mixture is defined as weighted sum of particular Gaussian functions $f_1(x_1, x_2)$ and $f_2(x_1, x_2)$.

Fig. 2 illustrates how a set of data points in a two-dimensional data space (left figure) is represented by a histogram first (middle figure), and then by a mixture of Gaussian functions (right figure). In the right figure, small circles represent the means of particular Gaussian functions whereas ellipses denote their shape.

In this simple two-dimensional example, it could be easy to recognize the position and shape of particular data clusters directly from the left figure. However, we must realize that the mixture of pdfs enables to approximate a big number of data points by a small set of parameters of Gaussian functions and that we can calculate with them and that we are not limited to one or two dimensions only.

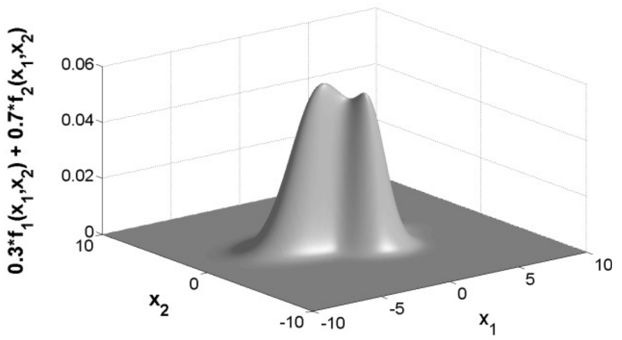

Fig. 1. Example of a mixture defined as weighted sum (weights 0.3 and 0.7) of two bivariate Gaussian functions with $\mu_{11} = 1, \mu_{12} = -1, \sigma_{11} = 1, \sigma_{12} = 1, \rho_1 = 0, \mu_{21} = -1, \mu_{22} = 1, \sigma_{21} = 2, \sigma_{22} = 1$ and $\rho_2 = 0$

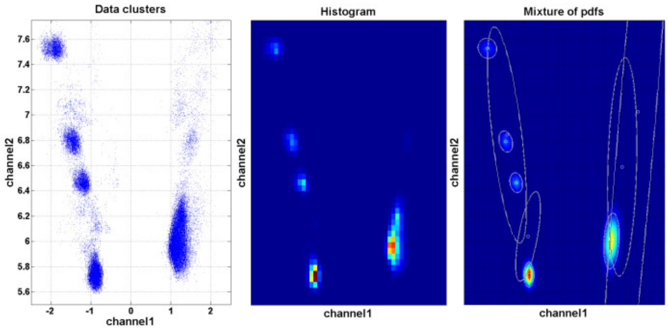

Fig. 2. Representation of data clusters by histogram and by mixture of Gaussian functions.

It is a task of estimation algorithms to find the representation of the data clusters by means of the mixtures of Gaussian functions. These algorithms were developed and programmed in the form of toolbox functions by the key participant in the above mentioned research projects, the Department of Adaptive Systems, Institute of Information Theory and Automation, Czech Academy of Sciences.

III. Bayesian Statistics

Besides the mixtures of Gaussian functions, Bayesian statistics is another key theoretical background of the developed advisory system. One of the most important notions of the Bayesian statistics, that the advisory system is based on, is the prior knowledge. In classical statistics, there is nothing to be said about the investigated process before an experiment starts. On the contrary, in the Bayesian statistics, we can take advantage of a prior information.

One possibility is that prior information can come from previous observations, i.e. historical recordings of individual channels in our case. Prior information can be given by expert knowledge too. The expert knowledge can limit the ranges of particular signal values according to physical limitation, e.g.

From the control theory point of view, the default black-box model of the process in the form of a mixture of probability density functions can be changed into a grey-box model by utilization of prior information derived from an expert knowledge of the process, [3].

The prior information is very useful in the above mentioned estimation algorithms. Let us say we have an amount of data available coming from the latest time period of production process on a rolling mill. The estimation algorithm tries to find a representation of this data in the form of a mixture of Gaussian functions. Prior information coming from expert knowledge can be in the form of limitations, e.g. that rolling force is positive and less than 5 MN, input and output strip tensions are positive and less than 160 kN, etc. Prior information coming from previous observations can be simply derived from data recordings from previous time periods. All this prior information helps the estimation algorithm reach more precise results in a shorter time. And that is the criterion which is of great importance, because we have to calculate the mixtures in multidimensional data space in real time while the process produces the metal strip in the speed of 5 to 10 meters per second, typically.

IV. Kullback-Leibler Divergence

Above in this paper, we learnt how the mixtures of Gaussian functions can represent the behavior of a process in a particular time period. The principle of the advisory system consists mainly in comparison of process data taken in different time periods and represented with the help of mixtures. The question is how to compare and measure the differences between mixtures in a simple and reliable way. During the development of the advisory system, Kullback-Leibler divergence was selected for this purpose, [4]. The Kullback-Leibler divergence enables us to measure the amount of information that is gained or lost if a probability distribution is replaced by another one. In the case of our advisory system, this criterion is used to measure the divergence between the historical mixture derived from historical data and the actual mixture representing data from the latest time period.

We will see later in this paper that Kullback-Leibler divergence is used as the key measure for distinguishing between a proper behavior of a process and its malfunction.

Let us note that Kullback-Leibler divergence is not a simple distance measure related to Euclidean distance between the means $[\mu_{11}, \mu_{12}]$ and $[\mu_{21}, \mu_{22}]$ of the two Gaussian functions in Fig. 1., e.g. This divergence takes into account the shapes of particular Gaussian functions given by σ and ρ parameters too.

V. Diagnostics

The principles described in previous chapters were verified and tested in industrial application of the advisory system. Good results encouraged the idea to extend the advisory system by some diagnostic functions too.

In industrial applications, diagnostic systems can usually easily recognize typical signal malfunctions like broken wire of an analogue signal, a signal modulated by electrical noise, a signal value exceeding limits given by a physical principle, etc. All these faults can be detected while taking into account the particular single signal only. On the other hand, there exists a wide set of more complicated faults that are hard to recognize with the help of a standard approach. A typical group of such faults are situations where the values of a single signal look completely trouble-free, but in relation to actual values of other signals or in relation to the latest history, the signal must be considered faulty.

We will demonstrate such an example with the help of a set of main signals on a rolling mill. The symbols displayed in the diagram of a typical rolling mill in Fig. 3. have the following meaning:

- T_1, T_2 ... tension in strip on input and output sides.
- v_1, v_2 ... strip speed measured by deflection rolls on input and output sides.
- H_1, H_2 ... strip thickness measured by thickness gauges on input and output sides.
- G ... uncompensated rolling gap.
- F ... rolling force.

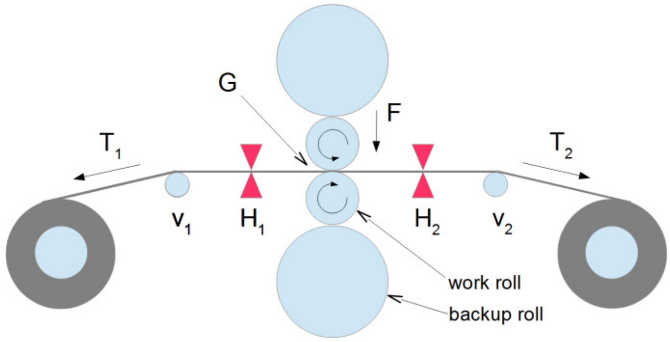

Fig. 3. Diagram of a rolling mill with symbols of main process variables.

The quality of production is evaluated mainly by the output thickness of the strip (H_2). All the other process variables influence H_2 substantially. If the production process is in a stable state, i.e. rolling speed is more or less constant, all process variables vary slightly around their means. Each substantial change of a variable causes a change in at least some other variables. Nevertheless, the relation between variables cannot be easily described. One of the reasons is that the production process is also influenced by conditions that are not measured, e.g. temperature of work rolls.

In some situations, the standard relation between process variables changes suddenly. The reason can be malfunction of a sensor, hardware failure in signal transmission, unmeasurable change of technological conditions, etc. From the diagnostics point of view, we would like to detect the situation and warn the operator. Then, the operator can stop the production of defective strip and avoid financial loss. Following algorithm was developed for the detection of these problematic situations:

- All process variables are recorded during the production process.
- Production time is divided into relatively short periods (in order of seconds).
- Data in each time period are approximated by a mixture of Gaussian functions.
- Kullback-Leibler divergence between the mixture representing the latest time period and the mixture from previous time period is calculated repeatedly.
- Big change in the sequence of Kullback-Leibler divergence values indicates a non-standard situation.

The algorithm was tested with data coming from control systems of rolling mills. For demonstration, we will describe three typical examples of problematic situations that can be revealed by the algorithm:

A. Malfunction of Thickness Gauge Caused by Mechanical Particles

Contact thickness gauges can often suffer from malfunction caused by mechanical particles clung to the surface of strip. Head of the gauge looses contact to the strip and measured thickness does not correspond to the real thickness.

B. Malfunction of Thickness Gauge Caused by Deformation of Strip

Situation similar to *A*. can happen if the strip measured by contact gauge is too tough. In that case, the strip has concave cross profile in the position of thickness gauge. As a result of this, the lower transducer of the gauge looses contact with the strip.

C. Temporary Loss of Input Tension in Strip

Strip comes to the rolling mill usually in the form of a coil. If the spires are reeled too loosely, the input tension T_1 can sometimes decrease suddenly for a short time. The spires are tighten at that moment. As a result of this, the output thickness changes slightly up and down. But a more serious problem is, that the tighten spires scratch the surface of the strip. If this situation is not detected, scratched surface can be revealed as late as at the customer and can result in a complaint. Let us mention that tension measurement is usually not precise enough so that this situation can be recognized with the help of tension signal only.

An advantage of the algorithm consists in the fact that all of the mentioned sample situations can be detected without a special tuning of algorithm parameters. Of course, it is necessary to set a reasonable width of the time period for the mixture recalculation. It is also necessary to find a condition representing the "big" change in divergence values, but this is easy, because the change is in the order of hundreds of percent in above mentioned situations.

The above mentioned *C.* situation is used for demonstration of the algorithm in Fig. 4. The algorithm is simplified for this purpose. The mixtures are calculated for two signals only, H_1 and H_2, so we are able to display the mixtures in two dimensions. Other process variables are taken for constant. Eight time periods with 200 samples each are shown. First chart is the course of input thickness. Second chart displays output thickness with sudden change near the sample number 1050. Eight small charts show mixtures calculated for particular time periods. Mixture representing the sixth period is apparently different. The last chart displays values of Kullback-Leibler divergence. Big change is visible at the end of the sixth time period where the divergence from mixture 5 to mixture 6 was calculated.

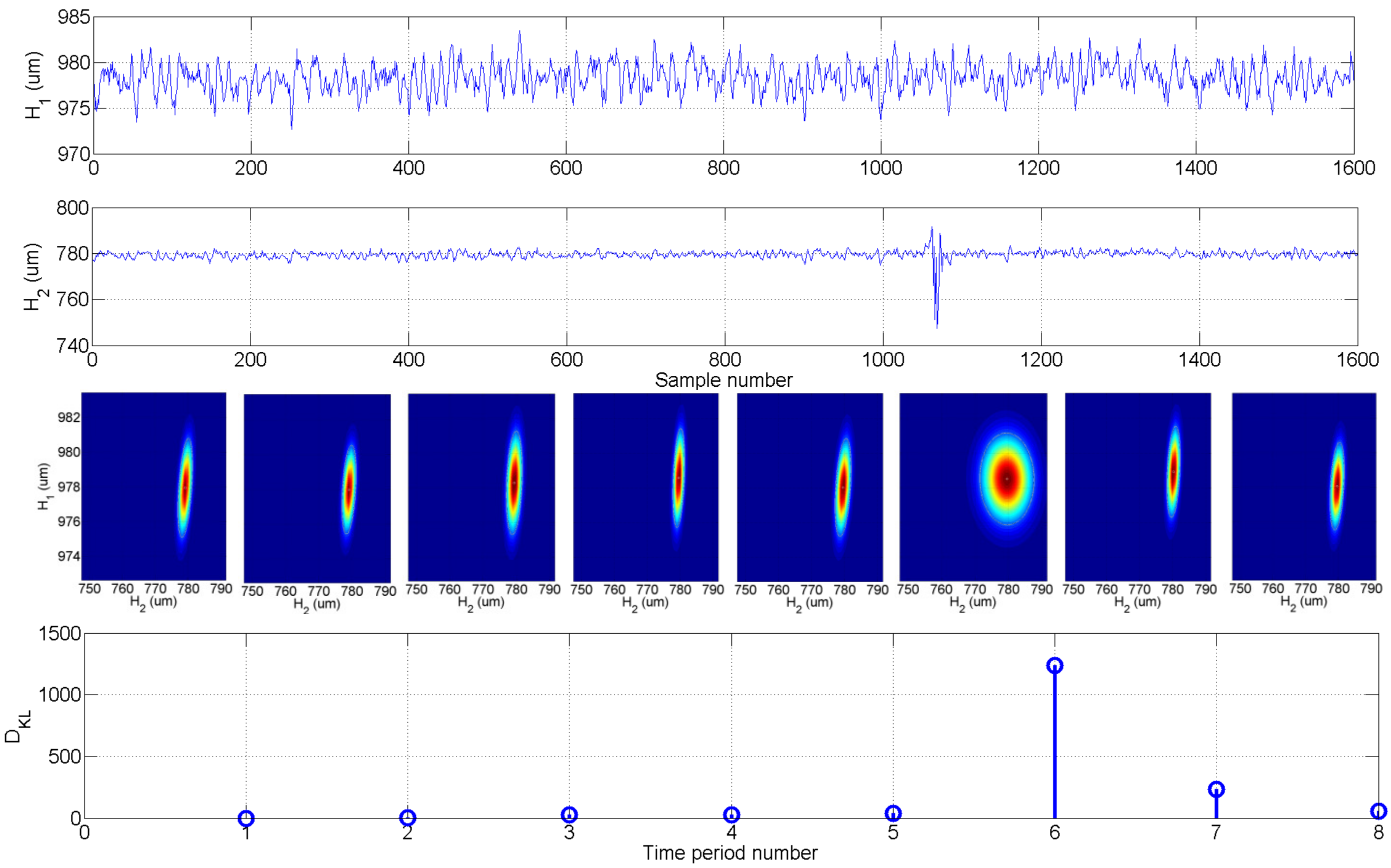

Fig. 4. Big change in Kullback-Leibler divergence (D_{KL} in the bottom chart) indicates problematic situation in the process of metal strip rolling.

VI. Conclusion

Probabilistic approach to description of behavior of complex industrial processes was used in the development of advisory system for the support of operators. Verified methods were tested for diagnostic purposes too. Results of the algorithm using Kullback-Leibler divergence between two mixtures of Gaussian functions are promising, but validation in real applications is necessary.

Acknowledgment

This work was supported by Ministry of Education, Youth, and Sport of the Czech Republic—University specific research—1311.

References

[1] I. Puchr, P. Ettler, “Continuous decision making for specific tasks related to metal processing industry”, in Proceedings of the ROADEF 2011, 12e congrès annuel de la Société française de Recherche Opérationnelle et d’Aide à la Décision, Saint-Étienne, FR, 2011.

[2] A. Gut, An intermediate course in probability, Dordrecht: Springer, 2009, p. 125.

[3] M. Kárný, J. Böhm, T. V. Guy, L. Jirsa, I. Nagy, P. Nedoma and L. Tesař, "Optimized Bayesian dynamic advising", Berlin: Springer, 2006, p. 20.

[4] S. Kullback, R. Leibler, "On information and sufficiency", Annals of mathematical statistics, vol. 22, pp. 79-87, 1951.