Monitoring and diagnostics of technological processes with the possibility of dynamic selection of variables

A A Sychugov
Institute of Applied Mathematics and Computer Science of Tula State University (TSU), Tula, Russia
xru2003@list.ru

Abstract. The article describes a method for monitoring and diagnosing the state of technological processes with the possibility of dynamic selection of variables. As a basis, the algorithm known from machine learning algorithm of single-class classification OneClassSVM was used, which showed a high percentage of accuracy in determining the normal state of the technological process. A model of a technological process as an infinite data stream with changing properties with a large number of parameters is proposed. A probabilistic model of the state space is used for possible dynamic selection of variables. An experiment is described in which the effectiveness of the proposed method was studied, and its results are presented. Its validity is shown, and it is noted that it can be used in modern systems for monitoring the state of technological processes to identify malfunctions and threats to information security.

Systems for monitoring and diagnostics of technological processes are used primarily in industrial enterprises that consume natural resources and have an impact on the environment. The introduction of such systems increases the effectiveness of environmental protection activities and contributes to the maintenance of safety at the enterprise itself [1, 2].

To effectively perform their functions, monitoring systems should use methods for analyzing the state of technological processes, which will allow detecting abnormal situations at the early stages of their development.

In [3, 4] a method for analyzing the state of technological processes based on the representation of the technological process as a continuous data stream with a large number of variables is proposed.

Set of states \( \Omega = \{\omega_1, \omega_2, \ldots, \omega_m\} \) of technological process as a dynamic system is divided into regular \( \Omega^+ \subset \Omega \) and abnormal ones \( \Omega^- \subset \Omega \). There is a problem of detecting abnormal states of the technological process as the initial stage of the appearance of a non-standard state.

The technological process is represented as a set of operations \( A = \{a_1, a_2, \ldots, a_n\} \), each of them changes in some way a finite number of geometric, physical, mechanical and other properties of the material object processed by this technological process.

Each operation \( a_i \in A \) of technological process is described by a certain set of variables \( x_i^{a_i} \),

\[
\forall a_i \in A: X^{a_i} = \left\{ x_1^{a_i}, x_2^{a_i}, \ldots, x_P^{a_i} \right\},
\]

which is divided into controlled \( \hat{X}^{a_i} = \left\{ \hat{x}_1^{a_i}, \hat{x}_2^{a_i}, \ldots, \hat{x}_P^{a_i} \right\} \), used for the operation of a control system or condition analysis, and uncontrolled

\[
\forall a_i \in A: X^{a_i} = \left\{ x_1^{a_i}, x_2^{a_i}, \ldots, x_P^{a_i} \right\},
\]
\[ \hat{X} = \{ \hat{x}_{a_1}, \hat{x}_{a_2}, \ldots, \hat{x}_{a_p} \}, \]
that is, they are defined during the design of the technological process and do not change during its operation.

Controlled variables are values obtained from geographically distributed sensors used in the process.

The following statements are true.
1. The sets of controlled and uncontrolled variables do not intersect.
2. Controlled variables change over time, i.e. they are actually time series:
   \[ \forall \hat{x}_{a_i} : \hat{x}_{a_i}(t) = f(\hat{x}_{a_i}(t-1), t), \]
   where \( f \) – a time series function that describes a change in a variable \( \hat{x}_{a_i} \).
3. All controlled variables change in some ranges.
4. Uncontrolled variables do not change over time, or their change is due to a structural change in the technological process.

The following features are typical for modern production facilities.
1. The real-time operation mode is characterized by rapid changes in the values of a large number of controlled variables.
2. The values of the controlled variables \( A_{a_j} X(\hat{x}_{a_i}) \) at some point in time \( t = \tau \) determine the state of the process at this time \( \omega_{\tau} \), in other words it is actually a point in a multidimensional space \( \Xi \), described \( AX \).
3. Not all elements of the set \( AX \) have the same significance for determining \( \omega_{\tau} \).

The task of diagnosing the technological process in this case is to determine the fact that the current state is regular, which corresponds to the task of single-class classification [5, 6, 7, 8, 9, 10] for the solution of which it is proposed to use the OneClassSVM algorithm [11, 12, 13, 14].

If the entire volume of data is used for diagnostics, the time and other overhead costs can be significant and reduce the effectiveness of the classification method used to almost zero.

For this reason, the task arises either to reduce the number of variables processed [16] without losing the quality of classification, or to use the existing set more effectively by "increasing" or "reducing" the influence of individual variables in the final learning criteria.

Finding an informative subset of variables in a large volume of data stream is a non-trivial task. First, the data stream is infinite, so any off-line algorithm for selecting variables is not applicable, and second, one needs to get a subset of variables in real time. In general, it can be noted that all existing algorithms for selecting variables, both using data packets [17] and single objects [18, 19], cannot process large data streams efficiently due to time and memory constraints.

Currently, there are three main approaches to selecting variables: filtering, wrapping, and embedded methods [20, 21, 22, 23, 24, 25, 26].

In [15], a Bayesian tool for selecting variables in the problem of pattern recognition on data streams is proposed. Here we propose a mathematical model for selecting variables based on the results obtained in [15] for the case of detecting the normal states of a technological process using a single-class classification.

Let the space \( \Xi \), described by a set of controlled parameters \( \tilde{X}^A \) [4], is probabilistic, and the probability that the current state of the system belongs to a set of regular ones is described using a logistic function [27] of the form:

\[ \varphi(\tilde{X}^A | y, W, \rho) = \frac{1}{1 + \exp\left( -\frac{2}{\sigma^2} y(W, \tilde{X}^A - \rho) \right)} \quad (1) \]
Let additionally directing vector $W=(w_1,\ldots,w_n)\in\Xi$ of the separating hyperplane $\left\langle W,\bar{X}^A \right\rangle - \rho$ [4] be considered as a vector of independent normally distributed random variables with zero mathematical expectation and with different variances $R=(\eta_1,\ldots,\eta_n)$

$$\psi(w_i | \eta_i) = \frac{1}{\sqrt{2\pi\eta_i}} \exp\left(-\frac{1}{2}\eta_i w_i^2\right)$$  \hspace{1cm} (2)

With respect to the displacement $\rho$ of the hyperplane $\left\langle W,\bar{X}^A \right\rangle - \rho$ [4], no a priori information is assumed, hence

$$\Psi(W,\rho | R) = \Psi(W | R) = \prod_{i=1}^{n}(1/\eta_i)^{-1/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \frac{w_i^2}{\eta_i}\right)$$  \hspace{1cm} (3)

Given that the variances $R=(\eta_1,\ldots,\eta_n)$ are positive definite and there are no a priori assumptions about the type of density distribution of variances, it is fair to assume that there are independent a priori gamma distributions of quantities that are the reverse of the variances

$$\gamma(1/\eta_i | \alpha, \beta) = (1/\eta_i)^{\alpha-1} \exp[(-\beta(1/\eta_i))]$$  \hspace{1cm} (4)

where $\alpha > 0, \beta > 0$ - parameters of gamma distribution.

At the same time, the mathematical expectations and variances $\forall \eta_i \in R$ are the same and are determined by:

$$M(1/\eta_i) = \alpha / \beta$$
$$D(1/\eta_i) = \alpha / \beta^2$$
$$\sigma(1/\eta_i) = \sqrt{\alpha / \beta}$$  \hspace{1cm} (5)

To determine the influence of parameter values $\alpha$ and $\beta$ on the type of a priori gamma distribution of inverse variances, the component of the guiding vector $W$, in addition to the expectation and standard deviation (5), it is also necessary to consider the ratio of the standard deviation to the mathematical expectation. $(\sigma / M | \alpha, \beta) = 1 / \sqrt{\alpha}$ [28].

If $(\sigma / M | \alpha, \beta) \to 0$, then the training criterion uses all the features of the objects.

If $(\sigma / M | \alpha, \beta) \to 1$, the learning criterion shows a pronounced tendency to over-selection in the selection of features, suppressing most of them, even useful ones.

One can control the degree of selectivity of feature selection by varying the parameter values of $\alpha$ and $\beta$ in the a priori distribution of variances. Let these parameters be set together according to the rule [28]:

$$\alpha = 1 + 1/(2\mu)$$
$$\beta = 1/(2\mu)$$  \hspace{1cm} (6)

In this case, there is a parametric family of gamma distributions defined by only one parameter $\mu \geq 0$, so that $M(1/\eta_i) = 1 + 2\mu$ and $D(1/\eta_i) = 2\mu(1 + 2\mu)$:

$$\gamma(1/\eta_i | \mu) = (1/\eta_i)^{1/(2\mu)} \exp[(-1/(2\mu)(1/\eta_i))]$$  \hspace{1cm} (7)

The proposed probabilistic model of the process state space $\Xi$ described by a set of controlled variables [3, 4] allows us to formulate a learning criterion for constructing a dividing hyperplane in the problem of analyzing process States in terms of a generalized linear approach to restoring dependencies [26], in the following form
\[
\left( W, \tilde{X}_i^A, R | \mu \right) = \arg \min \left\{ \sum_{i=1}^{n} \left[ r_i \left( w_i^2 + \frac{1}{\mu} \right) \left( 1 + 1 \right) \text{ln} \beta_i \right] + \sum_{i=1}^{n} \left( \left( \tilde{X}_i^A - \rho \right) \right) \right\}
\]

This criterion has the ability to automatically increase the absolute values of some components \( w_i \in W \) of the guiding vector \( W \) by reducing the coefficients \( r_i > 0 \) for them in the regularizing penalty \( \sum_{i=1}^{N} r_i w_i^2 \rightarrow \text{min} \), and "press to zero" other components, increasing their coefficients.

**Conclusion**

Using the single-class classification method OneClass SVM allows one to determine the normal state of the technological process with high accuracy (more than 90%). The use of the training criteria proposed in this work for selecting variables for training the classifier increases the accuracy by an average of 3%. This result was obtained during an experiment based on data obtained from Essity production lines [29]. During the experiment, the model was trained on the same data in the first case without dynamic selection, and in the second case using dynamic selection of variables.

In general, the proposed method can be applied in systems of monitoring and diagnostics of technological processes in order to identify abnormal conditions.

The research was carried out with the financial support of the RFBR in the framework of the scientific project no. 19-07-01107 \ 19.

**References**

[1] Maydanovich O. V. Intelligent technologies for automated monitoring of complex technical objects // Proceedings of SPIIRAS. Saint Petersburg Institute of Informatics and automation of the Russian Academy of Sciences, 2013, no. 6, pp. 201-216

[2] Plotnikova A. G. Analysis of technological process parameters using control maps // "Modern innovations: fundamental and applied research" Collection of scientific papers on the materials of the VIII International scientific and practical conference. 2018 Publisher: "Problems of science. 2018. pp. 16-18

[3] Aleksey A. Sychugov, Yulija V. Frantsuzova, Aleksander P. Anchishkin A method for analyzing the condition of technological process // Asia Life Sciences Supplement 19(1): 241-500, 2019 The Asian International Journal of Sciences, 241-251 ISSN: 01173375

[4] Sychugov, A. A., Chernov D. V. Anchishkin, A. P. Method of analysis of production processes // Radio Engineering. Moscow: Radio Engineering, 2019. Volume 83, no. 8(11). Pp. 59-66. DOI: 10.18127/j00338486-201908(11)-07

[5] Rohit B., Agarwal S. Stream Data Mining: Platforms, Algorithms, Performance Evaluators and Research Trends // International Journal of Database Theory and Application. 2016. T. 9. No 9. C. 201-218

[6] Fong S., Wong R., Vasilakos A. Accelerated PSO Swarm Search Feature Selection for Data Stream Mining Big Data // IEEE Transactions on Services Computing. 2015. C. 1-1

[7] Yusuf B., Reddy P. Mining Data Streams using Option Trees // International Journal of Computer Network and Information Security. 2012. T. 4. No 8. C. 49-54.

[8] Moya, M., Koch, M., and Hostetler, L. (1993). One-class classifier networks for target recognition applications. In Proceedings world congress on neural networks, pages 797–801, Portland, OR. International Neural Network Society, INNS.

[9] Ritter, G. and Gallegos, M. (1997). Outliers in statistical pattern recognition and an application to automatic chromosome classification. Pattern Recognition Letters, 18:525–539.

[10] Roberts, S., Tarassenko, L., Pardey, J., and Siegwart, D. (1994). A validation index for artificial neural networks. In Proceedings of Int. Conference on Neural Networks and Expert Systems in Medicine and Healthcare, pages 23–30.
[11] Koch, M., Moya, M., Hostetler, L., and Fogler, R. (1995). Cueing, feature discovery and one-class learning for synthetic aperture radar automatic target recognition. Neural Networks, 8(7/8):1081–1102.

[12] Scholkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C.: Estimating the support of a high-dimensional distribution. Neural Computation 13, 1443–1471 (2001)

[13] Albert D. Shieh, David F. Kamm: Ensembles of One Class Support Vector Machines. MCS 2009, LNCS 5519, pp. 181–190, 2009. Springer-Verlag Berlin Heidelberg 2009

[14] Tax, D.M.J., Duin, R.P.W.: Support vector domain description. Pattern Recognition Letters 20, 1191–1199 (1999)

[15] Pavel A. Turkov, Olga Krasotkina, Vadim Mottl, Sychugov A.A. Feature Selection for Handling Concept Drift in the Data Stream Classification // Machine Learning and Data Mining in Pattern Recognition - 12th International Conference, MLDM 2016, New York, NY, USA, July 16-21, 2016, Proceedings. Lecture Notes in Computer Science 9729, Springer 2016, ISBN 978-3-319-41919-0, p. 614-629

[16] Zhou X., Li S., Chang C., Wu J., Liu K. Information-value-based feature selection algorithm for anomaly detection over data streams. Tehnicki vjesnik, 2014, 21(2), pp. 223-232.

[17] Cai D., Zhang C., He X. Unsupervised Feature Selection for Multi-cluster Data. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining KDD '10, ACM New York, NY, USA, 2010, pp. 333-342.

[18] Maung C., Schweitzer H. Pass-efficient Unsupervised Feature Selection. Advances in Neural Information Processing Systems. Neural information processing systems foundation, 27th Annual Conference on Neural Information Processing Systems, NIPS 2013, pp. 1628-1636.

[19] Yang H., Lyu M.R., King I. Efficient Online Learning for Multitask Feature Selection. ACM Transactions on Knowledge Discovery from Data (TKDD), July 2013, Vol. 7, Issue 2, Article 6.

[20] Turkov P. A., Krasotkina O. V., Mottl V. V., Sychugov A. A. Feature selection in the classification problem with concept displacement for data flows // News Of TSU. Technical science. Issue 11. Part 1 Tula: TSU Publishing House, 2016. Pp. 81-99

[21] Sauerbrei W. The use of resampling methods to simplify regression models in medical statistics. Journal of the Royal Statistical Society, Series C (Applied Statistics), 1999, Vol. 48, Issue 3, pp. 313-329.

[22] Sauerbrei W., Schumacher M. A bootstrap resampling procedure for model building: Application to the cox regression model. Statistics in Medicine, December 1992, 11(16), pp. 2093-109.

[23] Fan J., Samworth R., Wu Y. Ultrahigh Dimensional Feature Selection: Beyond The Linear Model. Journal of Machine Learning Research, 2009, Volume 10, 2013-2038.

[24] Zou H. Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, 2005, Vol. 67, pp. 301-320.

[25] Zou H. The adaptive lasso and its oracle properties. Journal of the American Statistical Association, 2006, Vol. 101, Issue 476, pp. 1418-1429.

[26] Zou H. Li R. One-step sparse estimates in nonconcave penalized likelihood models (with discussion). Annals of Statistics, 2008, 36, pp. 1509-1566.

[27] http://www.machinelearning.ru

[28] Mottl V. V. Restoring dependencies based on empirical data: a generalized linear approach // Lecture notes. MIPT.

[29] http://soatulgu.ru/rffi_19-07-01107/index.html#