

# Intellectual Analysis System of Big Industrial Data for Quality Management of Polymer Films

Tamara Chistyakova<sup>1</sup>(✉), Mikhail Teterin<sup>1</sup>, Alexander Razygraev<sup>1</sup>,  
and Christian Kohlert<sup>2</sup>

<sup>1</sup> State Institute of Technology Saint Petersburg Russia, Saint-Petersburg, Russia  
nov@technolog.edu.ru

<sup>2</sup> Klöckner Pentaplast Europ GmbH & Co. KG, Montabaur, Germany  
c.kohlert@kpfilms.com

**Abstract.** The article is devoted to application of methods and technologies of the intellectual analysis of the large production data for the large-tonnage production of a polymeric film being characterized complex structure, existence of systemic linkages, semi-structured information, power consumption and a large number of the controlled and accumulated industrial information. The novelty consists in application of methods and data mining technologies in production of polymeric films. The article describes the application of neural networks for big data analysis. Application of neural networks allowed to carry out the forecast of quality of production that in turn allowed to reduce flaws. Also, this article describes the application of other data mining algorithms, which enabled to identify the causes of the decreasing quality of products and to find the values of controls in the system, which provide the specified quality and output as recommendations to the operator. Thus, the application of the developed intellectual system for analysis of large industrial data provided to reduce defects and increase productivity. Approbation of the developed system took place at plants of Russia and Germany and confirmed its working efficiency for management and change-over of productions of a polymeric film on various types of product.

**Keywords:** Classification · Neural network · Forecasting · Optimization · Data mining · Polymer films

## 1 Introduction

Modern production of polymer films is characterized by high cost of raw materials, high complexity of the technological process caused by the use of a large number of production stages, the use of additional devices, the monthly change in the product range and a variety of equipment types used [1, 2]. Customers' high demands for production quality cause annual losses of companies in the production of polymeric materials in the order of several billion euros due to the lack of time for analysis of the data and perceptions obtained in real time and the absence of a monitoring system satisfying the needs of the production line operators and of the quality engineers [3]. The most important indicators of the polymer film quality are: the absence of defects on

the surface – black dots, destructive (brown) streaks, inclusions of unmelted polymer and modifier particles, cracks (the bursting of air bubbles) – the color values, the film width and film thickness. The polymer film quality depends on quality of raw material and technological parameters of production [4].

The source information is the data from the extrusion-calender lines (Fig. 1) for the production of polymer films, characterized by: being multi-stage (production line includes the following stages of production: mixing; powder (extruder); moulding (calender); cooling; coiling); being multiproduct; large-tonnage (the equipment recycles 1000 kg per hour); continuity; huge volume of accumulated expertise in the conditions of their production (billion records); a large scope of controlled information (250 sensors); high power consumption [5]; complex structure and the presence of system linkages that describe weakly formalized information systems and complex application objects of study that leads to considerable complication of the rules for the construction of the formalized information-analytical models describing the potential relationships in the data.

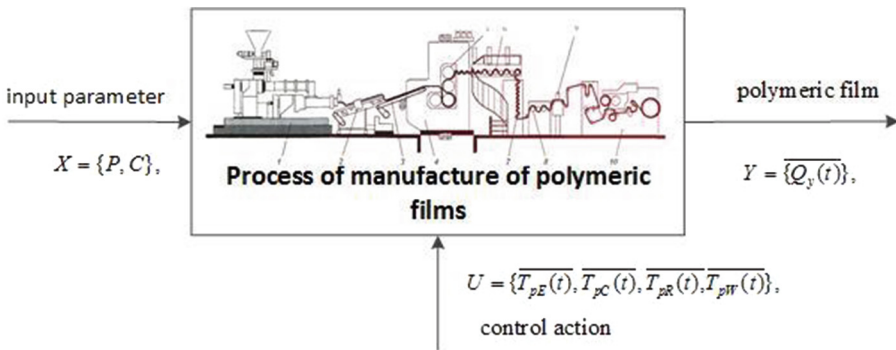


Fig. 1. Description of the production process for polymer films

Under these conditions, traditional approaches to information processing become ineffective. As a result there is an actual problem to develop and apply new multi-method approaches to data analysis. An actual direction of improving the efficiency of analysis of large volumes of semi-structured information, and building of analytic and information models describing the behavior of complex systems is the use of a systematic multi-method technology of intellectual analyze which is called “Data Mining”.

The aim of this project is to create software, which will allow the analysis of production data received from the monitoring system and identify the process parameters that caused the deviations of the actual values of quality thus providing a chance to improve the efficiency of production.

The use of this complex allowed to identify the complex structural relations between technological parameters, as well as to increase the quality of the products.

Technical implementation is to create a visualization system that allows us to build trends of technological parameters and quality indicators for the management production staff. This system is related to existing industrial SCADA systems. Information

from the SCADA-system is downloaded to a staging database, where it interacts with the developed intelligent system.

## 2 The Formalized Description of the Process

The formalized description of the polymer film materials production process can be represented in the form (Fig. 1): where  $Y(t)$  – is the vector of output variables, representing quality indicators of the polymer films;  $X(t) = \{P(t) - \text{type of polymer film, } C(t) - \text{component composition of the raw material}\}$ ;  $U(t)$  – vector of control actions, where  $T_{pE}(t)$  – vector of technological parameters of the extruder;  $T_{pC}(t)$  – vector of technological parameters of the calender operation;  $T_{pR}(t)$  – vector of technological parameters of the pulling device;  $T_{pW}(t)$  – vector of technological parameters of the winding machines, where  $t = t_1 \div t_{set}$ .

The technological parameters of the extruder are  $Pr^e(t)$  – the output of the extruder, m/min;  $S$  – extruder screw rotation speed, rev/min;  $T^s(t)$  – the temperature of the screw, °C;  $T^h(t)$  – temperature of the heaters, °C.

The process parameters of four-roll calender include:  $T^W(t)$  – the temperature of rolls, °C;  $Pr^W(t)$  – the velocity of the rolls, m/min;  $To^W(t)$  – torque rolls, N·m;  $P^{concave}(t)$  – pressure concave bending, Pa;  $P^{convex}(t)$  – pressure convex curve, Pa;  $P^{blend}(t)$  – pressure blend curve, Pa;  $P^{offset}(t)$  – pressure offset curve, Pa;  $FL(t)$  – the level of filling of gaps, mm.

Technological parameters of the pulling device:  $T^{tor}(t)$  – the temperature of pulling devices, °C;  $Pr^{tor}(t)$  – the performance of pulling devices, m/min;  $To^{tor}(t)$  – the torque pulling devices, N·m;  $T^{pr}(t)$  – temperature pressure rollers, °C;  $To^{pr}(t)$  – the torque pressure rollers, N·m;  $T^{cr}(t)$  – temperature of the cooling rollers, °C;  $To^{cr}(t)$  – the torque cooling rollers, N·m;  $T^{sr}(t)$  – temperature hardening of rolls, °C;  $To^{sr}(t)$  – the torque hardening of rolls, N·m;  $T^t(t)$  – the temperature of the stretch rollers, °C;

Technological parameters of the pulling device:  $T^{wi}(t)$  – the temperature of the winding machines, °C;  $S^{wi}(t)$  – the speed of winding, m/min;  $S^{wis}(t)$  – winding speed drawing, m/min;  $L(t)$  – the length of the wound roll, m;  $FT(t)$  – the tensioning force on a winding machine, N.

## 3 Problem Definition and Solving

The computer support system allows solving the following tasks:

- the definition of input parameters  $X(t)$  and control actions  $U(t)$ , which led to abnormal situation  $Y_{set1} < Y(t) < Y_{set2}$ , occurred in the period of time from  $t_1$  to  $t_2$  [5];
- the prediction of output parameters  $Y(t)$  for given input parameters  $X(t)$  and control actions  $U(t)$  during a period time from  $t_2$  to  $t_3$ ;
- the definition of control actions  $U(t)$  and input parameters  $X(t)$ , which allow to obtain the best film quality  $Y_{opt}$  [5].

The algorithm of the intellectual analysis for big data (up to 200 million records) allowing to issue recommendations to production staff on prediction of behavior of an object, on clarification of the true reasons, on determination of the best values of the control values to obtain the given quality of production (Fig. 2) is developed for the solution of objectives. At the first stage of managing staff export data from database of industrial parameters and characteristics to support computer system and input the threshold restriction on the film quality parameters [6]. The second stage is the design of a database and export production data into a data mining system database. The experts select the most significant parameters (Fig. 2) and make a request to the data mining system database to form an array of significant information for experts. The array is split into 3 data sets: training data set (70 %), evaluation data set (15 %) and test data set (15 %). Training data set is analyzed using one of the proposed data mining methods, and evaluates the adequacy of the model generated using the evaluation data set. Production quality is predicted using the test data set.

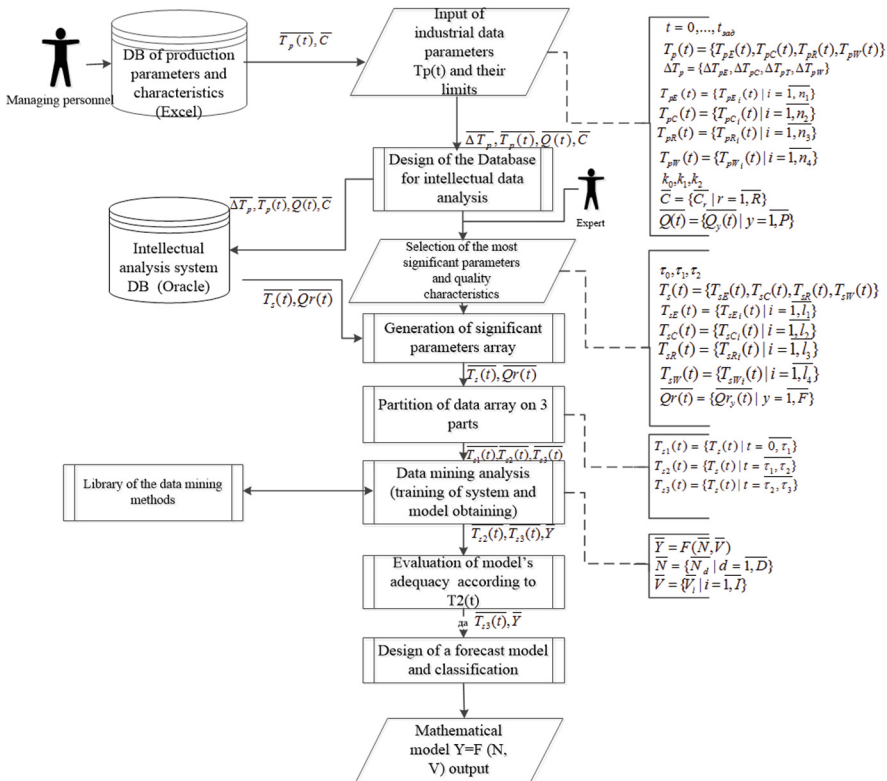


Fig. 2. The data mining algorithm for industrial production of the polymer film

The software solution (Fig. 3) is developed for realization of this algorithm and includes the following components: infoware (database of equipment parameters, production data, and knowledge base of emergency situations); mathware

(mathematical model based on neural network, decision trees, support vector machines, and naive Bayesian classifier) and optimization algorithms [6].

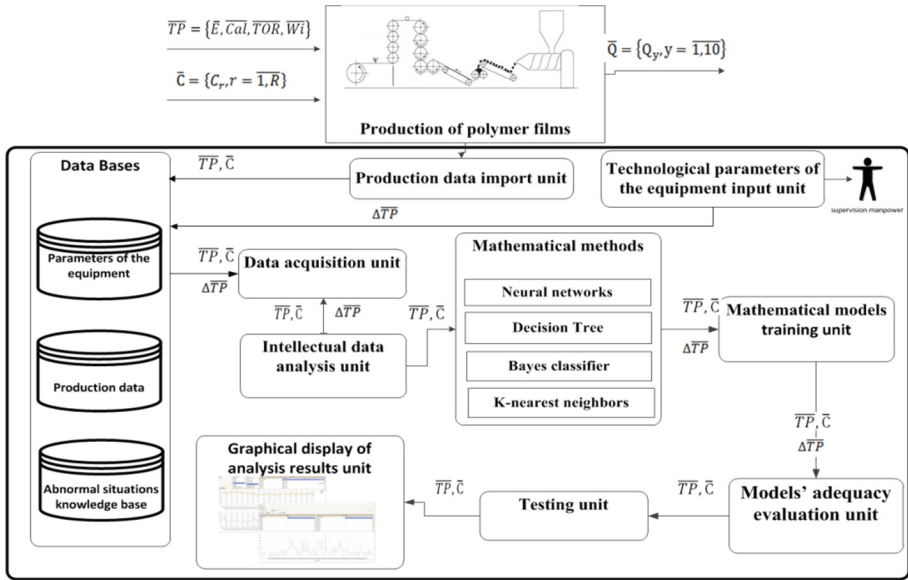


Fig. 3. The functional structure of a software solution

Using the wide range of algorithms is caused by the fact that each method solves the problem with the given accuracy and it has disadvantages. The decision trees allow to identify cause of the defect on the film surface, have a high degree of interpretability and high speed training.

The neural network was used for solving the prediction task. Rumelhart multilayer perceptron was used ideally for predicting film quality in the production. The input of neural network receives input parameters  $X(t)$  and control actions  $U(t)$  in normalized form, and the at output we get the predicted output parameter values  $Y_{pred}(t)$ . The neural network was trained by backpropagation. This iterative gradient method minimizes the error of the neural network and obtains the goal output value [7–9] (Fig. 4).

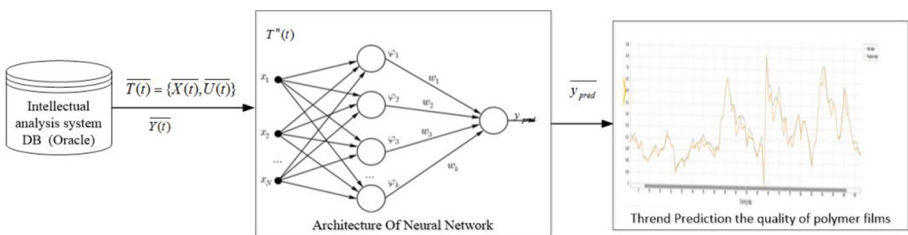


Fig. 4. The scheme of the forecasting unit

The linear regression model provides the search for the relationship between the input and the output variable based on the regression equation. The regression model learns quickly and can be easily implemented, but in case of lack of data, model would be inadequate, and in case of the large amount of data, the model can be overfitting [10]. Regression model significance is evaluated using Fischer’s test.

There also was implemented a classification algorithm based on Random Forest search technology. The algorithm implements subsequent procedure for building machine learning algorithm composition and generates the ensemble of classifiers using random sampling with replacement, that means selection of several random subsets while sampling. This algorithm is used when high accuracy of classification is required. The algorithm tends to overfitting, especially on noisy data and requires large amounts of memory for storage of mathematical models [11–13].

Testing the software solution performance was conducted on data obtained from industrial plants in Russia and Germany Corporation “Klößner Pentaplast”. The data contain 200 million rows for 250 different data processing parameters. Testing considered one of the major defects – black dots. Regression model, algorithm of Random Forests, a neural network (Rumelhart multilayer perceptron), a decision tree and knowledge bases about abnormal situations were used.

As a test example the case in which the neural network predicted abnormal situation is demonstrated. Neural network training occurred the extrusion-calender data for line № 14 of the corporation “Klößner Pentaplast”. Data is read from the database in a sequential chronological order. The input layer of the neural network receives 7 parameters, the number of neurons in the hidden layer is 15. The forecast trend is shown on Fig. 5. The trend for the black dots number in case of abnormal situation exceeding the permissible value of black dots quantity is shown [14].

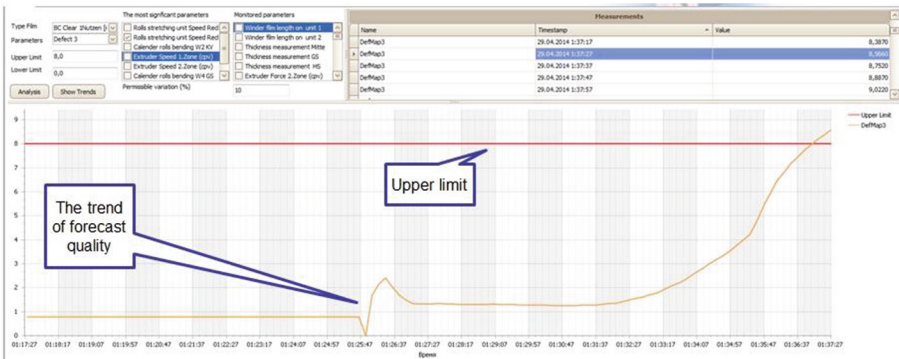


Fig. 5. Prediction trend for the number of black dots

To find the causes of the predicted abnormal situation the regression model and Random Forests algorithm were used. An array of the most important data such as the speed of the calender rolls, the speed of the tempering rolls, the speed of removable rolls, the roll bending of calender rolls, the screw speed in the extruder, the temperature of the screw in the extruder, the temperature of the mixture in a hot mixer including

1000 measurements (2 h) were formed. Testing was conducted for the film brand “BC Clear 1Nutzen” with thickness 100  $\mu\text{m}$ .

Based on the obtained statistical quality and process parameters data simple regression models were built. Models’ significance is estimated using Fischer’s test dispersion method. Estimation showed that the highest correlation is observed between the number of black dots and the following technological parameters: velocity of the 3<sup>rd</sup> calender roll, velocity of the removable roll, velocity of the tempering rolls.

On the next step the array, obtained by experts, is divided into 3 parts. Training, evaluation and testing used data mining method – Random Forest. Random Forest showed that the most important parameters are almost all the speeds of the production line rolls. Thus, the results of Random Forests analysis and Fisher’s test match (Table 1).

**Table 1.** The cause of the defect extracted from the knowledge base

Cause	Action	Probability
Insufficient amount of stabilizer	To increase the stabilizer not more than valid 1.1 %	55 %
The irregular shape of the melt	The melt should leak out. To adjust roll bending. To follow the profile	30 %
Mixture temperature is too high	The temperature of the mixture in a hot mixer or the residence time in the hot mixer should be reduced	6 %
Configuration of OCS is wrong	OCS must be configured	6 %
Dots look matte black	To clean the knife head and the body of kneader	2 %
The black dots on the edges	Clean cheeks and a locking ring	1 %

Testing proved the efficiency of developed algorithms for intellectual analysis of industrial data in the production of polymer films, the algorithm of data reduction, the algorithm for finding the reasons of black dots occurrence and the whole software solution.

## 4 Conclusion

Thus, the intelligent system for quality control of polymer films, including a database of technological parameters, a library of mathematical methods (neural network, decision trees, Random Forests, k-nearest neighbors), the interfaces of visualization of technological and production quality parameters, allows to perform neural network forecast quality of polymer films, to identify the true causes of flaws, to predict, to find

combinations of variables, which is basis for resource- and energy-saving management, as it leads to a rational use of energy and reduce of flaws.

The basis of intelligent system is a neural network that adapts to the type of the polymer film using the input parameters to be analyzed, control actions and the values of quality indicators. Neural network with 7 neurons in the input layer, 15 neurons in the hidden layer and 1 neuron on the output layer is trained by back propagation method. Neural network has proven its predictive capability and the accuracy of classification reached 93 %, which meets the requirements of the company where the testing a program complex was done.

**Acknowledgments.** This research is partially supported by the Russian Foundation for assistance to small innovative enterprises in science and technology (Projects No. 5263GU1/2014) and by the German Academic Exchange Service (DAAD) (Grant of Leonard Euler).

## References

1. Kohlert, M., König, A.: Advanced polymeric film production data analysis and process optimization by clustering and classification methods. *Front. Artif. Intell. Appl.* **243**, 1953–1961 (2012)
2. Kohlert, M., Chistyakova, T.: Advanced process data analysis and on-line evaluation for computer-aided monitoring in polymer film industry. *J. Saint Petersburg State Inst. Technol. (Tech. Univ.)* **29**, 83–88 (2015)
3. Optical Control Systems GmbH. <http://www.ocsgmbh.com>
4. Kohlert, M., König, A.: High dimensional, heterogeneous multi-sensor data analysis approach for process yield optimization in polymer film industry. *Neural Comput. Appl. Mater.* **26**(3), 581–588 (2015)
5. Chistyakova, T., Polosin, A., Razygraev, A., Kohlert, K.: Software system to control the color of thin rigid polymeric materials. *J. Autom. Ind.* **7**, 12–18 (2012)
6. Kohlert, M., Teterin, M., König, K., Chistyakova, T.: Computer support in production polymeric films on data mining. *Math. Methods Technics Technol.* **3**, 20–25 (2012)
7. Verikas, A., Bacauskiene, M.: Feature selection with neural networks. *Pattern Recogn. Lett.* **23**(11), 1323–1335 (2002)
8. Haykin, S.S.: *Neural Networks and Learning Machines*. Pearson Education, Upper Saddle River (2009)
9. Segal, E., Pe'er, D., Regev, A., Koller, D., Friedman, N.: Learning module networks. *J. Mach. Learn. Res.* **6**, 557–588 (2005)
10. Seber, G., Lee, A.: *Linear Regression Analysis*. Wiley, New York (2003)
11. Biau, G., Devroye, L., Lugosi, G.: Consistency of random forests and other averaging classifiers. *J. Mach. Learn. Res.* **9**, 2015–2033 (2008)
12. Breiman, L.: Bagging predictors. *Mach. Learn.* **24**, 123–140 (1996)
13. Breiman, L.: Random forests. *Mach. Learn.* **45**, 5–32 (2001)
14. König, A., Gratz, A.: Advanced methods for the analysis of semiconductor manufacturing process data. In: Pal, N.R., Jain, L. (eds.) *Advanced Techniques in Knowledge Discovery and Data Mining*, pp. 27–74. Springer, London (2005)