Big Data Analysis in Film Production

T. B. Chistyakova, F. Kleinert and M. A. Teterin

Abstract The article analyzes the current trends of digitalization for large innovative industrial production, which are international, large-capacity, distributed in different geographical locations and having several production lines at each plant. Such trends of digitalization as predictive analytics and 6 sigma methodology, which includes Ishikawa diagram and DMAIC (definition, measure, analysis, improvement, control) cycle, are considered. The novelty of the work lies in the application of methods and technologies of intellectual analysis of large industrial data for production of polymeric films and in the application of mathematical models that allow online calculation of uncontrolled consumer characteristics of products (thickness, color of polymeric films) and integrate them into one single system of data mining. Developed software solution includes visualization unit, forecast unit, statistical data analysis unit. Software solution allows us: determine the types of films with the best yield; check the production data for normalcy; calculate process capability index; calculate key performance indicators. Application and testing of the big data analysis system on the example of large industrial Corporation Kloeckner Pentaplast proved its efficiency.

Keywords Polymer films · Predictive analytics · Big data · Machine learning · Statistic analysis

---

T. B. Chistyakova (✉) · M. A. Teterin
Saint Petersburg State Institute of Technology, Moskovsky Prospect 26, 190013 Saint Peterburg, Russia
e-mail: nov@technolog.edu.ru

F. Kleinert
Klöckner Pentaplast Europe GmbH & Co. KG, Montabaur, Germany

A. G. Kravets et al. (eds.), *Cyber-Physical Systems: Advances in Design & Modelling*, Studies in Systems, Decision and Control 259,
https://doi.org/10.1007/978-3-030-32579-4_18

1 Introduction

Productions of polymer films are modern, international, the innovation and are characterized by multiassortment, large capacity; continuity; a large number of data sources (about 35 plants with several production lines on each of them); a large number of different data storages (BDE, Business Warehouse, SAP, OCS) which need to be integrated among themselves; a large number of administrative and production personnel (6300 employees making decisions); in the large volume of the accumulated expert industrial information (one billion records); large number of controlled information (>250 sensors); the total production of equal 1000000 tons of polymeric products a year [1, 2]. The analysis of industrial productions showed that they are characterized: imbalance of classes for the reason that defects on production arise seldom; a complexity and existence of system communications which describe poorly formalized information systems and difficult application entities of management that leads to considerable complication of rules of creation of the formalized information and analytical models describing patterns in data [3–5].

Polymer films found broad application in different areas: in food industry, in medicine and in electronics. The most important consumer characteristics of polymer films are: color; width; film thickness; shrinkage value; existence of black points; existence of destructive (brown) bands; existence of inclusions of unfused polymer; existence of the modifier, cracks (the burst air bubbles) [3, 6]. On process, there are indicators, which can lead to an abnormal emergency situation on an object—reduction of material in a gap between rolls, break of a film [7]. Depending on a scope, different requirements to consumer characteristics of a film are imposed to polymeric film materials (film thickness in the range from 25 to 1200 μm, film width in the range from 100 to 2500 mm).

Digital transformation of the industrial industry consists of 4 stages:

1. Introduction of automated control systems for internal processes of the enterprise (SAP system).
2. Creation of services for cloud computing.
3. Introduction of systems for the analysis of trends and forecasting of data.
4. Introduction of systems for training in precedents (machine learning).

At the current stage digital transformation is at the 4th stage and is planned to complete digitalization by 2020. Introduction of systems for training in precedents consists of 4 under stages: creation of storages of data; to create the module of visualization of data; to increase productivity by means of innovative methods of machine learning and to connect business processes with artificial intelligence.

Task of control for production personnel is search of values of the operating influences and factors of production which provide implementation of requirements to consumer characteristics of a polymeric film.

The purpose of work is development of a program complex which with use of methods of machine learning, methods of the statistical analysis will allow to find set of the operating influences and factors which provide performance of preset values of consumer characteristics of a polymeric film and to visualize set of characteristics of industrial production in a form, ergonomic for administrative and production personnel.

For the solution of objectives the methodology six sigma allowing to improve the existing processes by methods is used: DMAIC (Define, Measure, Analyze, Improve, Control), Isikava's chart, calculation of key performance indicators of process and calculation of indexes of reproducibility of process.

2 The Formalized Description of Polymeric Film Production Process Materials

The formalized description of polymeric film production process materials (see Fig. 1) can be submitted as $P(t) = \{X(t), U(t), Y_1(\mathrm{t}), Y_2(t)$, where: $X(t)$—vector of input variables; $U(t)$—vector of the operating influences; $Y_1(t)$—consumer characteristics which are controlled automatically; $Y_2(t)$—consumer characteristics which are calculated by mathematical models$\}$; $X(t) = \{Pt(t)$—type of a polymeric film, $C(t)$—component composition of raw materials, $Cf(t)$—equipment configuration; $Xf(t)$—vector of criteria for comparison of samples of films$\}$; U(t) = $\{UpE(t)$—vector of the operating impacts of work of an extruder; $UpC(t)$—vector of the operating impacts of work of a calender; $UpR(t)$—vector of the operating influences of the pulling device$\}$, where $t = t_1 – t_{adjusted}$.

The vector of input variables includes: $P(t) = \{n(t)$—flow index of the polymer; μ—viscosity of polymer, Pa s; $\alpha(t)$—relative change of the size of a film, calculated

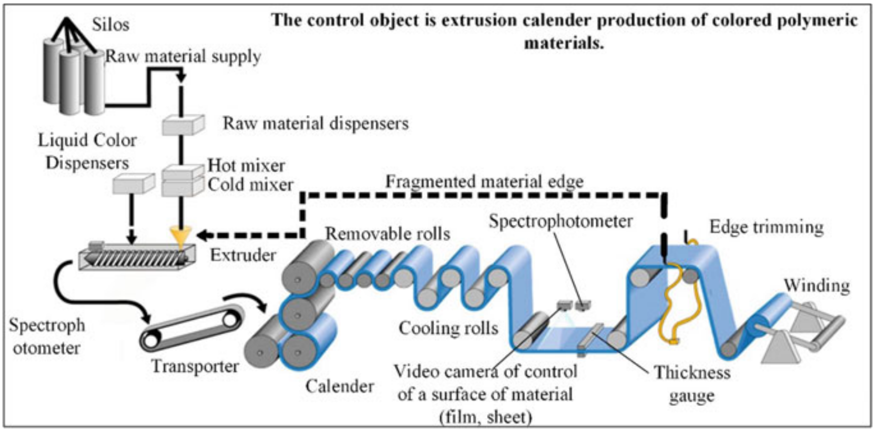

Fig. 1 Description of polymeric film production process

as an equation root; $G_0(t)$—module of elasticity of material} $C(t) = \{M(t)$—composition weight}; $Cf(t)$ {Ln—line number} $Xf(t) = \{Tqe(t)$—requirements for composite uniformity of the extrudate; $Tqk(t)$—requirements to dimensional quality of a film and to quality of a surface of a film (lack of defects: black points, destructive strips, inclusions of unfused polymer etc.)}.

Vector of the operating impacts of work of an extruder consists of $S(t)$—speed of rotation of the screw, turns/min; $V(t)$—Rotating speed of a spiral in a boot funnel, turns/min; $Ts(t)$—screw temperature, °C; $Th(t)$—temperature of heaters, °C, $Tb(t)$—case temperature.

The vector of the operating impacts of work of a four-roll calender includes: $TW(t)$—temperature of rolls, °C; $ToW(t)$—torque of rolls, N m, $Pbend(t)$—bend pressure, Pa; $Poffset(t)$—shift pressure, Pa.

Vector of the operating influences of the pulling device: $Ttor$—temperature of the pulling devices, °C; $Totor(t)$—torque of the pulling devices, N m; $Tpr(t)$—temperature of clamping rollers, °C; $Topr(t)$—torque of clamping rollers, N m; $Tcr(t)$—temperature of the cooling rollers, °C; $Tocr(t)$—torque of the cooling rollers, N m; $Tsr(t)$—temperature of the tempering rolls, °C; $Tosr(t)$—torque of the tempering rolls, N m; $Tt(t)$—temperature of tension rollers, °C.

$Y_1(t) = \{Blp(t)$—quantity of black points on the set surface area; $Hel(t)$—gelik on 10 m^2; $Air(t)$—burst air bubbles on the set surface area; $Dest(t)$—destructive, brown strips on the set surface area; $Inc(t)$—inclusions of unfused polymer, on the set surface area; $Fib(t)$—fibers, on the set surface area; $Th_1(t)$—film thickness, mkm}. $Y_2(t) = \{Shr(t)$—shrinkage size; $Lc(t)$, $ac(t)$, $bc(t)$—rated color coordinates of a finished product; $Th_2(t)$—film thickness, mkm}.

The program complex allows solving the following problems:

1. Using control maps of Shukhart, indexes of a reproducibility of process, and the rule of 3 sigma of normal distribution of a random variable to find and the operating influences, which provide the set requirements to consumer characteristics $Y_{1,adj1}$ $Y_{1i}(t)$ $Y_{1,adj2}$, where $t = t_1 - t_2$ [8, 9].
2. Forecasting of output parameters $Y_{1,i}(t)$ at the set input parameters $X(t)$ and the operating influences $U(t)$ in the period of time from t_2 to t_3.
3. To issue recommendations about the operating influences in case of a deviation from the set requirements to consumer characteristics $Y_{1,adj1} \le Y_{1i}(t) \le Y_{1,\mathrm{adj2}}$.
4. Display of multidimensional data $X(t)$, $U(t) \in R^N$ on the plane $z(t) \in R^2$ in the period of time from t_1 to t_2 [10].

3 Functional Structure of a Program Complex for Quality Control of Polymeric Materials

The program complex (see Fig. 2) includes the following components: information support (database of production design characteristics of process, database of parameters of the equipment, database of parameters of material, base of production

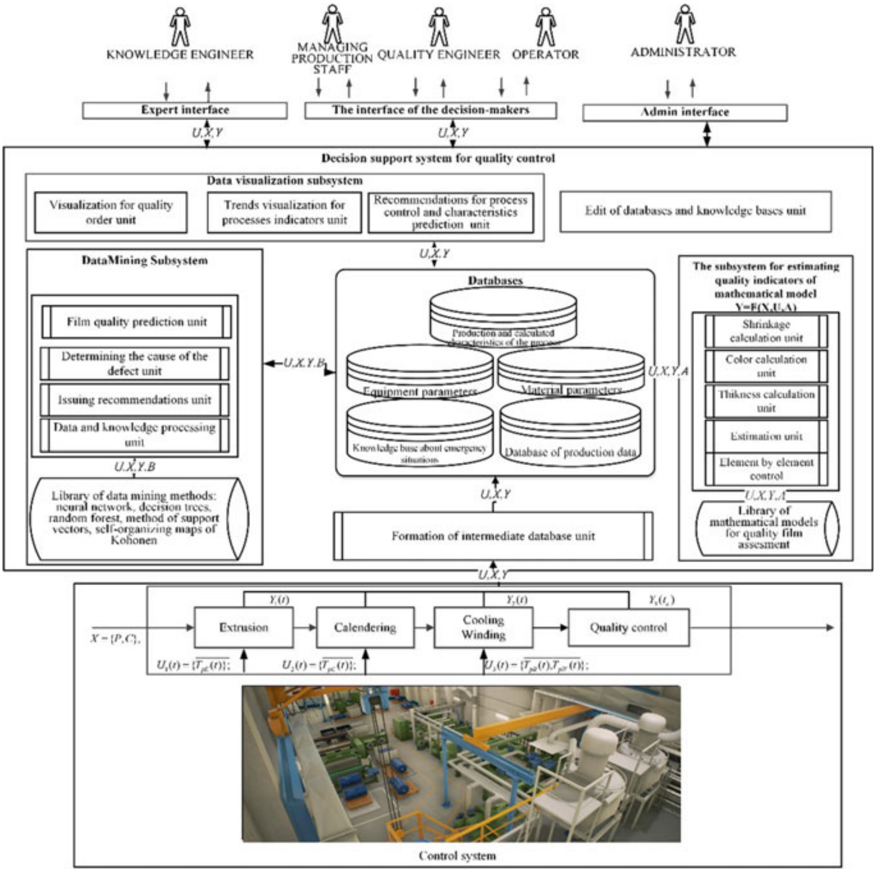

Fig. 2 Functional structure of a program complex for quality control of polymeric materials

data, and knowledge base of emergency situations); subsystem of data visualization [10–15]; module of editing databases and knowledge; subsystem of control and forecasting of quality of polymer films; subsystem of forming of estimated figures of merit of a polymer film.

Testing of operability of the software product happened according to industrial data of the plants of Russia and Germany: for a month of production which contained 500 thousand various data on 250 process parameters on shrinkage (rated size) and black points.

The module of verification of industrial data on normal distribution (see Fig. 3) was tested [16–19]. In case data submit to normal distribution, then for the analysis of data linear algorithms (linear regression, a method of basic vectors, etc.) were used. If the hypothesis of compliance of distribution of probability of result of measurements to the normal law is rejected, then either nonlinear algorithms are used or robust methods of processing of industrial data are applied.

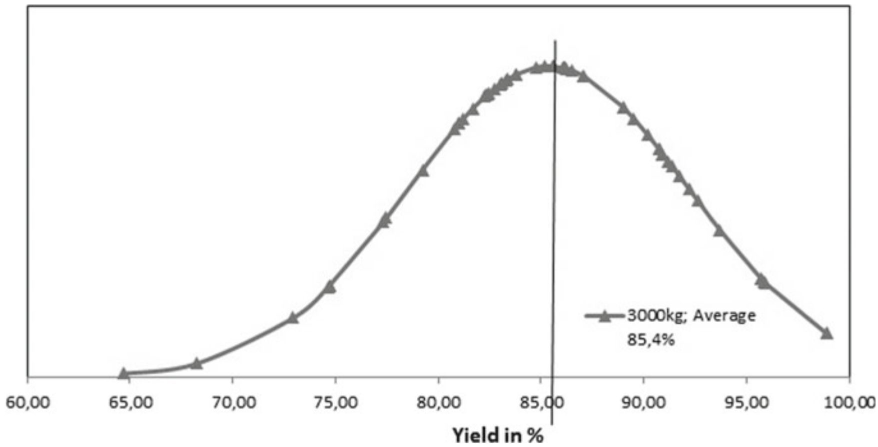

Fig. 3 Function of density of probability for an yield in %

In addition, indexes of reproducibility of process (see Fig. 4) are calculated and by means of a multiple linear regression (see Fig. 5) managing influences which had the greatest impact on consumer properties are defined[20–25]. According to a multiple linear regression the most significant technological parameters for black points are rotating speed of the screw in an extruder, material feed speed in a funnel, high temperature of mix that completely matches expert data.

Testing proved operability of a system of the analysis of big data on production of polymeric films.

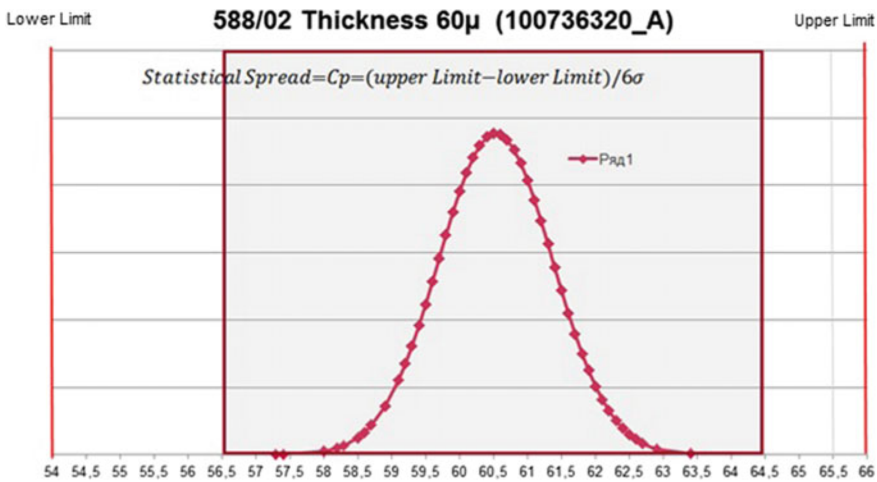

Fig. 4 Interface of work of a subsystem of calculation of indexes of reproducibility of process

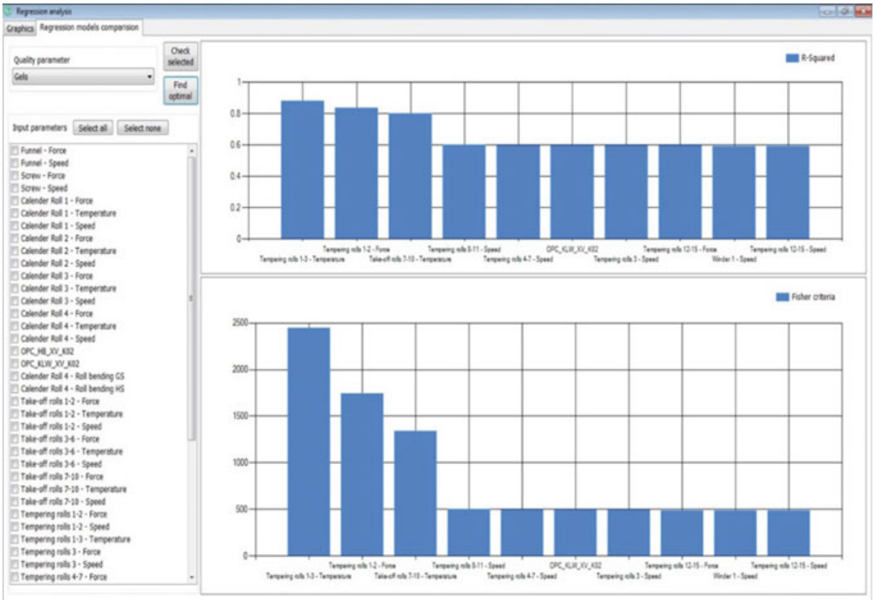

Fig. 5 Search of the most significant process parameters for use of multiple linear regression

4 Conclusion

The program complex, algorithms, which were developed, allow predicting the best supplier of polymeric materials, to visualize data in an ergonomic form, to predict consumer qualities of a polymeric film and to issue recommendations about elimination of emergencies. Use of the developed software product allows increasing the professional level of administrative and production personnel of production.

The software product underwent successful approbation on production of polymeric film materials of the plants of Russia and Germany. The program complex has adaptive flexible architecture thanks to which expansion of its functionality due to development and connection of additional program modules is supported.

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. St. 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., Araztaganova, A.M., Kohlert, C.: Computer system for thermal shrinkage polymer films obtaining processes control. Herald Kazan Technol. Univ. **19**(17), 101–105 (2016)
6. Chistyakova, T., Teterin, M., Razugraev, A., Kohlert, C.: Intellectual analysis system of big industrial data for quality management of polymeric films. In: Lecture Notes in Computer Science—T.9812, pp. 234–242 (2016)
7. König, A., Gratz, A.: Advanced methods for the analysis of semiconductor manufacturing process data. In: Advanced Techniques in Knowledge Discovery and Data Mining, pp. 27–74 (2005)
8. Tadesse, D.G., Carpenter, M.: A method for selecting the relevant dimensions for high-dimensional classification in singular vector spaces. Advances in Data Analysis and Classification, Springer; German Classification Society - Gesellschaft für Klassifikation (GfKl); Japanese Classification Society (JCS); Classification and Data Analysis Group of the Italian Statistical Society (CLADAG); International Federation of Classification Societies (IFCS), vol. 13(2), pp. 405–426, June (2019)
9. Altman, D.G., Martin B.J.: Statistics notes: the normal distribution. BMJ, **310**, 298 (1995)
10. Cox, Michael AA and Cox, Trevor F. Multidimensional scaling. In: Handbook of Data Visualization, pp. 315–347. Springer (2008)
11. Friedman, J., Hastie, T., Tibshirani, R.: The Elements of Statistical Learning (Springer series in statistics New York) (2001)
12. Korhonen, P.J., Silvennoinen, K., Wallenius, J.: O¨orni A Can a linear value function explain choices? An experimental study. Eur. J. Oper. Res. **219**(2), 360–367 (2012)
13. Belloni, A., Chernozhukov, V., et al.: Least squares after model selection in high-dimensional sparse models. Bernoulli **19**(2), 521–547 (2013)
14. B¨uhlmann, P., et al.: Statistical significance in high-dimensional linear models. Bernoulli **19**(4), 1212–1242 (2013)
15. B¨uhlmann, P., Van De Geer, S.: Statistics for high-dimensional data: methods, theory and applications. Springer Science & Business Media (2011)
16. Candes, E., Tao, T.: The Dantzig selector: statistical estimation when p is much larger than n. Ann. Stat. 2313–2351 (2007)
17. Reid, S., Tibshirani, R., Friedman, J.: A study of error variance estimation in lasso regression. Statistica Sinica **26**, 35–67 (2016)
18. Sala-i-Martin, X.X.: I just ran two million regressions. Am. Econ. Rev. 178–183 (1997)
19. Pavlyshenko, B.: Machine Learning, Linear and Bayesian Models for Logistic Regression in Failure Detection Problems (2016)
20. Tibshirani, R.: Regression shrinkage and selection via the Lasso. J. Roy. Stat. Soc. Ser. B, (1996)
21. Yang, D., Usynin, A., Hines, J.W.: Anomaly-based intrusion detection for SCADA systems. In: 5th International Topical Meeting on Nuclear Plant Instrumentation Controls, and Human Machine Interface Technology. pp. 797–803 (2006)
22. Duncan, A.J.: Quality control and industrial statistics. R.D. Irwin, Homewood, IL (1974)
23. Montgomery, D.C., Runger, G.C.: Introduction to Statistical Quality Control, 5th edn. John Wiley, New York (2008)
24. Hussein, Maher, Allawi Al-Morshedi, Abbas, Shomran, Haideer.: The Comparison Between Shewhart Control Chart, Cusum and EWMA. https://doi.org/10.13140/rg.2.2.34006.34880 (2013)
25. Salah, H., Ahmed, M., Saleh, K., Sherif, A.: Effect of sample size on the performance of Shewhart control charts. Inter. J. Adv. Manuf. Technolo. https://doi.org/10.1007/s00170-016-9412-8 (2016)