Data mining system for predicting quality of polymeric films

T B Chistyakova and M A Teterin

Department of Computer-Aided Design and Control Systems, St. Petersburg State Institute of Technology (Technical University), 26 Moskovsky Avenue, St. Petersburg 190013, Russia

E-mail: mikhail.teterin92@gmail.com

Abstract. A computer data mining system for predicting the quality of polymer films in international, large-scale and multi-assortment industries is described. This article presents a library of statistical and data mining methods that allows, using statistical tests, to test data for distribution normality, predict the quality of polymer film materials for various line configurations and different types of film and includes the following methods: recurrent neural networks, neural network with long short-term memory and convolutional neural network. Analysis of methods of predicting the quality of polymer films was carried out and an algorithm was developed that allows selecting the most appropriate method for predicting the quality of polymer films based on the type of film, line configuration and requirements for film quality. The system includes interfaces that display trends in the process characteristics of the process class. The system was tested using the example of industrial data of the corporation on production of polymer film in plants of Russia and Germany.

1. Introduction

In the modern world, in the production of polymer films, one of the key problems is predicting the quality of the polymer film in real time. The production of polymer films for the food and pharmaceutical industries are multi-assortment systems [1]. They are characterized by the type of line (extrusion, calender), high productivity (more than 1000 kg h\(^{-1}\)), energy intensity, a large amount of controlled information, process efficiency criteria, various defects (each type of line has its own abnormal situations and deviations on the film surface). Multi-variability is due to many formulations, different ranges of thickness (0.025–1.20 mm) and width (up to 6200 mm). Reconfiguration of the line to a new task by the type of polymer film occurs on average once a day. Each unique task, recipe, line type has its own emergency situations in production, permissible values of film quality indicators. For each of these assortments, you need to be able to make real-time decisions. Therefore, to support decision-making by control production personnel, it is necessary to develop a polymer film quality prediction system based on neural network approaches, where for each unique assortment it is necessary to build its own model or one common meta-model analysis of which would provide the required consumer characteristics of polymer films. In case we need to predict data in real time, multiple regression can be used, but to use this approach effectively, it is necessary to verify that the data obeys the law of normal distribution.
Thus, for the production of polymer films, the development of a computer system that would allow checking data for distribution normality and predicting the quality of polymer films using neural network approaches is relevant [2]. The purpose of the work is to create a computer system for predicting the quality of polymer films, which contains a database of line types, a database of recipe data, a database of controlled and calculated characteristics of the polymer film production process, a library of mathematical models for statistical analysis of data and prediction of the quality of polymer films [3].

Figure 1. Informational description of polymer film production as a control object.

2. Formulation prediction tasks. Computer system architecture

Analysis of formulations, configurations and mode parameters of production lines, polymer film quality indicators made it possible to develop an information description of extrusion-calender production (ECP) of polymer film (PF) as a control object, which is presented in figure 1 in the form of a set of vectors of input parameters $X$, control $U$ and disturbing $F$ effects and output parameters $Y$.

Polymer film type $T_{film} = \{R_f, Q_f^o\}$ is manufactured on line having a configuration $C_{line} = \{T_{extrud}, C_{extrud}, \Gamma_{extrud}, C_{calend}, \Gamma_{calend}, n_{take}, \Gamma_{take}, T_{roll}, C_{roll}, \Gamma_{roll}\}$, with specified performance $G_0$. The type of polymer film is determined by its recipe $R_f$ and quality requirements $Q_f^o = \{S_{f0}, \Delta S_{f0}^\text{max}, S_{f0}^\text{max}, \Delta S_{f0}^\text{max}, D_{f0}^\text{max}, n_{black}^\text{max}, n_{air}^\text{max}, n_{water}^\text{max}, n_{gel}^\text{max}, n_{bubble}^\text{max}, n_{def}^\text{max}, S_{f0}, \Delta S_{f0}^\text{max}, S_{f0}^\text{max}, \Delta S_{f0}^\text{max}\}$. 
The main requirements for the quality of the polymer film are: $\delta_{f0}$ – given thickness, mm; $\Delta \delta_{f \text{ max}}$ – maximum allowable thickness deviation from the specified value, mm; $D_{\text{max}}$ – maximum permissible thickness difference, mm; $n_{\text{black max}}, n_{\text{detr max}}, n_{\text{melt max}}, n_{\text{melt max}}, n_{\text{gel max}}, n_{\text{hobl max}}, n_{\text{hol max}}$ – maximum permissible number of surface defects – black dots, destructive strips, inclusions of unmelted polymer, gels, cracks and holes, respectively – on the specified area of the film; $S_{f0}, S_{f\beta}, S_{f\text{fc}}$ – requirements to machinery-direction and transverse-direction shrinkage, %; $L, a, b, \Delta E_{\text{max}}$ – requirements to color coordinates.

We consider an approach to construct optimization algorithms for non-smooth functions that are given by its values (no analytical formula) and their derivatives are not known [1–4]. In this approach the original optimization problem is replaced with a randomized problem, allowing the use of Monte-Carlo methods for calculating integrals. It will be shown, that the value of the gradient can be obtained using values of target function only, in the framework of Monte-Carlo methods for calculating integrals. Furthermore, this value should not need to be precise, because it is recalculated at each iteration step.

Instrumentation of key stages of the production line includes extruder (stage $s = 1$ – extrudate preparation), calender (stage $s = 2$ – extrudate molding in polymer film), detachable-draw and cooling rolls (stage $s = 3$ – removal and cooling of polymer film).

Extruder has type $T_{\text{extrud}}$, defined by the configuration $C_{\text{extrud}}$ and geometrical parameters $\Gamma_{\text{extrud}}$ of screws, and the nature of their movement (single-screw extruder, reciprocating extruder, co-rotating and counter-rotating twin-screw extruders).

Calender is configured $C_{\text{calend}}$, defined by a number $n_{\text{calend}}$ and scheme of a relative positioning (L-shaped, $\Gamma$-shaped, etc.) rolls, and geometrical parameters $\Gamma_{\text{calend}}$.

Control actions at the calendering stage $U_l(t)$ are: $V_l(t), T_{l}(t), i = 1, \ldots, n_{\text{calend}}$ – circumferential velocities (ms$^{-1}$) and outside surface temperatures (°C) of calender rolls; $\tau_{l}^{\text{left}}(t), \tau_{l}^{\text{right}}(t)$ – operation time of the left and right electric motors moving (lifting/lowering) the external sizing roll, which provides for adjustment of the thickness of the polymer film, s; $x(t), r(t)$ – spatial skew (horizontal axis displacement) of the internal calibration roll (m) and the counter-bending force applied to the external calibration roll (N). Corrective actions $x(t)$ and $r(t)$ are used to ensure the uniformity of the polymer film [4].

During the stages of calendering, removal and cooling of the polymer film, emergency situations arise related to defects in the thickness of the polymer film (for example, thickness differences), surface defects of the polymer film (for example, black dots, yellow-brown destructive strips, inclusions of unalloyed polymer, holes).

The computer system being developed (figure 2) includes the following components: information support (database of production design characteristics of the process, database of equipment parameters, database of material parameters, database of production data, and knowledge base of emergency situations); data visualization subsystem; database and knowledge editing module; data mining subsystem, subsystem for evaluation of quality indicators by mathematical models [5, 6].

The data mining subsystem contains a library of statistical and mining methods for checking data for normality of distribution and predicting the quality of polymer films. The library of statistical data analysis methods contains the Pearson criteria and the Kolmogorov–Smirnov criterion.
Figure 2. Computer data mining system functional diagram for polymer film quality prediction.

The prediction methods library includes the following mathematical models and methods: classical
implementation of the neural network, recurrent neural network, long-term short-term memory neural network (LSTM), convolutional neural network and neural network, which combines convolutional and LSTM approach [7–14].

Performance testing of the computer system is carried out according to large industrial data of pharmaceutical and food packaging polymer films of a given formulation based on rigid polyvinyl chloride, assembled on a production line implementing a blow extrusion method. Each data set collected for the day of production contained about 500 thousand measured values of 210 process parameters (control effects, consumer characteristics of the polymer film).

Figure 3. Histogram of distribution of black dots and holes.

3. Computer system test results
As an example, the work of a computer system on handling emergencies related to the occurrence of black dots and holes is considered. In figure 3, it can be seen that data on black dots and holes are not subject to the law of normal distribution, so multiple linear regression cannot be used and quality indicators can be predicted using only neural networks.

The results of the forecasting system are shown in table 1 and figures 4 and 5. For the “black dots” defect, the recurrent neural network gave the value of the AUC prediction index (area under the curve) [7] equal to 0.675. A neural network with brief long-term memory gave an AUC of 0.943. The convolutional neural network produced an AUC of 0.886. Integration of the brief long-term neural network and

| Model      | AUC “Black dots” | AUC “Holes” |
|------------|------------------|-------------|
| RNN        | 0.675            | 0.576       |
| LSTM       | 0.943            | 0.742       |
| CNN        | 0.886            | 0.768       |
| LSTM + CNN | 0.800            | 0.728       |

Table 1. Results of the forecasting system for polymer films production.
convolutional network yielded an AUC of 0.800. For an emergency situation involving holes, a recurring neural network gave an AUC of 0.576, a neural network with long short-term memory gave a result of 0.742, a convolutional neural network gave a result of 0.768, and the integration of a short long-term neural network and a convolutional network gave an AUC of 0.728.

The peculiarity of the emergency situation connected with the appearance of black dots on the film is that this is a frequent emergency situation in the production of polymer films and we have a certain number of black dots constantly, while holes are a rarer defect. Thus, if a defect in production occurs constantly, then it is better to use a model with long short-term memory, and if an abnormal situation rarely occurs, then it is better to use a convolutional network in the tasks of predicting the quality of the polymer film. Unfortunately, integration of neural network with long short-term memory and convolutional network did not bring any improvements in improving model accuracy. This is probably because this type of neural network is retrained and more data is needed for this type of modes.

![Figure 4. Results of neural network approaches for abnormal situation related to black dots on film surface.](image)

Figure 5. Results of neural network approaches for emergency situation related to holes on film surface.

4. Conclusions
The data mining system architecture for polymer film quality prediction is presented, which includes a library of data mining methods, which allows testing the data mining system on various line configurations, recipes and defects characterized by different repetition rates. The results of the testing confirmed the operability of the computer prediction system (library of data mining methods) for processing film defects such as black dots and holes. The Statistical and Mining System has enabled data to be checked for distribution normality and a suitable mining method for each type of defect. If the accuracy of the model is important and the emergency situation is constant, then it is better to use a neural network with long short-term memory. If the abnormal situation associated with this type of defect rarely occurs, then you need to use a convolutional network.

References
[1] Kohlert M and König A 2012 Advanced polymeric film production data analysis and process optimization by clustering and classification methods Frontiers in Artificial Intelligence and Applications 243 1953–61
[2] Kohlert M and Chistyakova T 2015 Advanced process data analysis and on-line evaluation for computer-aided monitoring in polymer film industry Bull. of Saint Petersburg State Institute of Technology (Technical University) 29 83–88

[3] Kohlert M and König A 2015 High dimensional, heterogeneous multi-sensor data analysis approach for process yield optimization in polymer film industry Neural Computing and Applications Materials 26 581–88

[4] Chistyakova T, Teterin M, Razygraev A and Kohlert C 2016 Intellectual analysis system of big industrial data for quality management of polymeric Lecture Notes in Computer Science 9812 234–42

[5] Kohlert M and Hissmann O 2017 Applied industry 4.0 in the polymer film industry Proc. the 16th TAPPI European PLACE Conf. pp 183–90

[6] König A and Gratz A 2005 Advanced methods for the analysis of semiconductor manufacturing process data Advanced Techniques in Knowledge Discovery and Data Mining (London: Springer) pp 27–74

[7] Hastie T, Tibshirani R and Friedman J 2009 The Elements of Statistical Learning: Data Mining, Inference, and Prediction (New York: Springer)

[8] Heravi S, Osborne D and Birchenhall C 2004 Linear versus neural network forecasts for European industrial production series Int. J. of Forecasting 20 435–46

[9] Hochreiter S and Schmidhuber J 1997 Long short-term memory Neural Computation 9 1735–80

[10] Glorot X and Bengio Y 2010 Understanding the difficulty of training deep feedforward neural networks Proc. the Thirteenth Int. Conf. on Artificial Intelligence and Statistics pp 249–56

[11] Pandhare V, Singh J and Lee J 2019 Convolutional neural network based rolling-element bearing fault diagnosis for naturally occurring and progressing defects using time-frequency domain features Proc. Prognostics and System Health Management Conf. pp 320–26

[12] Jiang J, He H, Yan J and Xie P 2019 Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox IEEE Trans. Ind. Electron. 66 3196–207

[13] Ince T, Kiranyaz S, Eren L, Askar M and Gabouj M 2016 Real-time motor fault detection by 1-D convolutional neural networks IEEE Trans. Ind. Electron. 63 7067–75

[14] Davis J 2006 The relationship between precision-recall and ROC curves Proc. the 23rd Int. Conf. on Machine Learning pp 233–40