

ТЕХНОЛОГИИ BIG DATA В СИСТЕМАХ КОНТРОЛЯ КАЧЕСТВА МЕТАЛЛУРГИЧЕСКОГО ПРОИЗВОДСТВА



Д. Н. Гайнанов
Заведующий кафедрой
«Аналитика больших данных и
методы видеоанализа» Ураль-
ского федерального универси-
тета им. Ельцина Б.Н.



Д. А. Беренов
Аспирант Уральского феде-
рального университета им.
Ельцина Б.Н.

Уральский федеральный университет им. Ельцина Б.Н., Россия
E-mail: damir.gainanov@gmail.com, berenov@dc.ru

Abstract. The paper considers the application of Big Data technologies in quality control systems in metallurgical production. The concept of a technological pyramid is introduced on the basis of which an approach to solving the problem of the optimal assignment of technological routes is developed and it is shown how this approach can be used to reduce the level of manufacturing defects. The proposed methods are approved at the «Severstal» metallurgical plant within the framework of the AS SPC project (automated system for statistical production control).

Введение. В настоящее время аналитика больших данных (Big Data) находит всё большее применение в организации различных производств, в оптимизации технологических процессов. При этом важными особенностями данных, используемых для анализа, являются их объёмы (измеряемые многими терабайтами информации), потоковый характер данных (непрерывная генерация данных) и большое разнообразие используемых параметров (тысячи и десятки тысяч параметров). В приложении к металлургическому производству данный подход предполагает организацию сквозного сбора технологических параметров в процессе движения производимой продукции вдоль технологического маршрута. В результате такого процесса фиксируется история создания каждой единицы продукции с учетом прохождения её через все технологические переделы и с привязкой всех важных технологических параметров, влияющих на качество производимой продукции. В работе развивается подход к анализу данных для такой обширной базы исторических технологических данных с целью повышения эффективности производства и снижения уровня брака в производстве.

Основные определения

Рассмотрим дискретное производство, состоящее из определенного числа технологических переделов, осуществляющихся на соответствующих технологических агрегатах и линиях. Например, в качестве такого производства может служить металлургическое производство.

Пусть $A = \{A_1, \dots, A_n\}$ – совокупность технологических агрегатов, задействованных в производстве. В данной работе важным понятием будет служить понятие единицы продукции (ЕП). Под единицей продукции понимается неделимая часть выходной или входной продукции, получаемой на агрегате или технологической линии. В качестве типичных примеров единиц продукции можно привести сталь, полученную в сталеплавильном агрегате и выпущенную в сталеразливочный ковш; слябы, получаемые после машины непрерывной разливки

стали; горячекатаные рулоны, получаемые как конечный результат работы стана горячей прокатки; холоднокатаные рулоны, получаемые как конечный результат работы стана холодной прокатки.

Определение. Ориентированный граф $G = (A, E)$ с множеством вершин A и множеством дуг $E \subseteq A^2$ будем называть инфраструктурным графом, если $(A_1, A_2) \in E$ тогда и только тогда, когда выходная ЕП агрегата A_1 может служить входной ЕП для агрегата A_2 .

Например, для металлургического производства единица продукции сляб является выходной ЕП для машины непрерывного литья заготовок и одновременно входной ЕП для стана горячей прокатки.

Технологическим маршрутом $P = (A_{i_1}, A_{i_2}, \dots, A_{i_k})$ будем называть любой ориентированный путь в графе G . Множество всех технологических маршрутов $P = \{P_1, \dots, P_k\}$ будем называть технологической базой рассматриваемого производства. Обозначим через $EP = \{ep_1, \dots, ep_n\}$ множество всех возможных единиц продукции рассматриваемого производства. Пусть каждая единица продукции ep_i характеризуется набором параметров $P_i = \{p_{i1}, p_{i2}, \dots, p_{in}\}, i \in \overline{1, K}$. Тогда последовательность

$$AI_i = (A_{i1}, P_{i1}(AI_i), \dots, A_{is}, P_{is}(AI_i)),$$

где $P_{ij}(AI_i)$ – набор значений параметров для ep_{ij} в конкретной реализации технологического маршрута AI_i – будем называть исполненным технологическим маршрутом (ИТМ).

В результате производственной деятельности рассматриваемого производства будет сгенерировано множество исполненных технологических маршрутов на текущий момент времени t :

$$P_{\text{ИТМ}}(t) = \{AI_i \mid i = [1, q(t)]\}.$$

В процессе производственной деятельности происходит процесс непрерывного накопления ИТМ. Заметим, что накапливаемая информация имеет все признаки больших данных, а именно:

- 1 имеет значительные объемы, измеряемые многими терабайтами информации;
- 2 накопление информации происходит в потоковом режиме с большой скоростью;
- 3 накапливаемая информация характеризуется большим разнообразием и содержит значения нескольких

Определение. Технологической пирамидой $\text{Pir}(G, v)$ в графе G с корневой вершиной v будем называть подграф графа G , порожденный множеством вершин $\langle v \cup G(v) \cup G^2(v) \cup \dots \cup G^k(v) \rangle_G$ такой, что любой ориентированный путь P графа G , начинающийся в вершине v , целиком лежит в этом подграфе.

Здесь через $G^k(v)$ обозначено множество всех вершин v' графа G таких, что существует простой путь из вершины v в вершину v' длины $k-1$. Терминальной вершиной графа (подграфа) называется вершина, из которой не выходит ни одной дуги, лежащей в этом графе (подграфе). Пусть V' – множество терминальных вершин подграфа

$\langle v \cup G(v) \cup G^2(v) \cup \dots \cup G^k(v) \rangle_G$. Для каждой терминальной вершины $v_i \in V'$ существует некоторая единица продукции ep_i , являющаяся выходной ЕП для этой вершины, причем таких ep_i может быть несколько в зависимости от вида ИТМ, в результате которого была получена данная единица продукции.

Исполненный технологический маршрут $AI_i = (A_{i1}, P_{i1}, \dots, A_{is}, P_{is})$ будем называть продуктовым, если единица продукции на выходе терминальной вершины A_{is} ИТМ AI_i – обозначим эту вершину $\text{term}(AI_i)$ – является одним из видов конечного продукта, поставляемого на рынок. Любую вершину $v' \in (v \cup G(v) \cup G^2(v) \cup \dots \cup G^k(v))$ будем называть развилкой (рис. 1), если $|G(v')| > 1$.

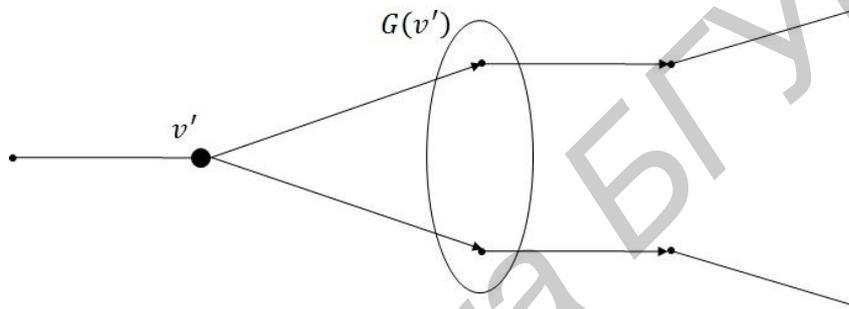


Рис. 1. Вершина–развилка в инфраструктурном графе

Контроль качества продукции на основе технологий Big Data. Предлагаемый подход может быть эффективно применен при построении системы контроля качества продукции для сложного металлургического производства. Ниже рассмотрим задачу прогнозирования брака ЕП при выполнении заданного технологического маршрута. Очевидно, что в этом случае даже такое простое решение, как прекращение дальнейшего исполнения технологического маршрута, которое может привести к производственному браку выглядит целесообразным и, как минимум, даёт экономию на затратах на продолжении работ на последующих агрегатах, которые могут достигать существенных значений.

Для каждой ЕП, после верификации ее качества, можно определить множество наборов параметров корневой вершины, при которых получалось бы то или иное качество ep . Это будут классы в обучающей выборке. Соответственно, нам необходимо определить решающее правила в вершинах–развилках так, чтобы на обучающей выборке они правильно детерминировали вершины. В этом случае для каждого применяемого технологического маршрута $P = (A_1, A_2, \dots, A_k)$ формируется множество ИТМ на текущий момент времени t :

$$P_{\text{ИТМ}}(t) = \{AI_i \mid i = \overline{1, q(t)}\}.$$

Для каждого ИТМ из $P_{\text{ИТМ}}(t)$ определены параметры конечной ЕП, выходящей после завершения выполнения этого ИТМ. На основе этих параметров может быть определено качество полученной единицы продукции. В простейшем случае это могут быть два класса: годные или бракованные единицы продукции.

Заметим, что для любой вершины из маршрута P , являющейся вершиной–развилкой в

технологическом графе $G = (A, E)$ может быть составлена обучающая выборка, состоящая из двух классов: годные и бракованные единицы продукции.

$$\begin{cases} AI_1 = (\text{Ind}_1, i_1, A_1, \text{Pr}_{11}, A_2, \text{Pr}_{12}, \dots, A_k, \text{Pr}_{1n}), \\ AI_2 = (\text{Ind}_2, i_2, A_1, \text{Pr}_{21}, A_2, \text{Pr}_{22}, \dots, A_k, \text{Pr}_{2n}), \\ \dots\dots\dots \\ AI_k = (\text{Ind}_k, i_k, A_1, \text{Pr}_{k1}, A_2, \text{Pr}_{k2}, \dots, A_k, \text{Pr}_{kn}), \end{cases}$$

где (A_1, A_2, \dots, A_k) – последовательность вершин технологического графа $G = (A, E)$, предшествующих вершине–развилке в маршруте P , и Ind_i – индикатор, принимающий значение 0, если единица конечной продукции при данном ИТМ получилась бракованной, и значение 1, если единица конечной продукции при данном ИТМ получилась годной. Таким образом, в этом случае получается классическая задача обучения с учителем, в которой обучающая выборка разбита на два класса.

Пусть n – совокупное множество параметров для части ИТМ маршрута P , предшествующей рассматриваемой вершине–развилке v' . Тогда обучающая выборка в данной задаче может быть записана как два множества B_0 и B_1 n -мерных векторов:

$$\begin{aligned} (B_0) & \left\{ \begin{array}{l} (a_{11}, \dots, a_{1n}), \\ \dots \\ (a_{m1}, \dots, a_{mn}), \end{array} \right. \\ (B_1) & \left\{ \begin{array}{l} (a_{(m+1)1}, \dots, a_{(m+1)n}), \\ \dots \\ (a_{l1}, \dots, a_{ln}), \end{array} \right. \end{aligned}$$

где совокупность B_0 – представляет бракованные единицы продукции, и B_1 – годные единицы продукции. Тогда задача распознавания образов в геометрической постановке будет сводиться к решению следующей системы линейных неравенств:

$$\begin{cases} a_{11} \cdot x_1 + \dots + a_{1n} \cdot x_n > 0, \\ \dots \\ a_{m1} \cdot x_1 + \dots + a_{mn} \cdot x_n > 0, \\ a_{(m+1)1} \cdot x_1 + \dots + a_{(m+1)n} \cdot x_n < 0, \\ \dots \\ a_{l1} \cdot x_1 + \dots + a_{ln} \cdot x_n < 0. \end{cases} \quad (1)$$

Если система совместна и $\bar{x} = (x_1, \dots, x_n)$ – её решение, то практическое использование этого решения состоит в том, что для вновь рассматриваемого технологического маршрута при достижении вершины–развилки v' будут иметься конкретные значения параметров (a_{j1}, \dots, a_{jn}) , реализовавшиеся при данном маршруте. В этом случае можно прогнозировать

результат исполнения намеченного технологического маршрута P : если $a_{j_1} \cdot x_1 + \dots + a_{j_n} \cdot x_n > 0$, то прогнозируется брак конечной единицы продукции, если же $a_{j_1} \cdot x_1 + \dots + a_{j_n} \cdot x_n < 0$, то прогнозируется получение годной единицы продукции при дальнейшей реализации рассматриваемого технологического маршрута P . Если же система несовместна, то для решения задачи распознавания в геометрической постановке могут быть применены методы распознавания образов, разработанные в работах [1–7].

Для заданного технологического маршрута P указанная задача может решаться для каждой вершины–развилки графа $G = (A, E)$, входящей в маршрут P . Таким образом, при достижении каждой вершины–развилки мы можем вычислить прогноз по тому, будет ли получено годное изделие или бракованное изделие при реализации данного технологического маршрута. Предположим, что рассмотрены все используемые технологические маршруты P и для каждой вершины–развилки технологического графа $G = (A, E)$ решены задачи прогнозирования результата дальнейшего исполнения технологического маршрута. Тогда стратегия повышения эффективности производства на основе рассматриваемой в данной работе оптимизации технологических процессов на основе технологий Big Data состоит в следующем.

Пусть подлежит реализации технологический маршрут $P = (A_{i_1}, A_{i_2}, \dots, A_{i_k})$, в котором содержится несколько вершин–развилок $(A_{j_1}, A_{j_2}, \dots, A_{j_m})$ технологического графа $G = (A, E)$. Запускаем процесс производства до первой вершины–развилки A_{j_1} включительно:

$$(A_{i_1}, A_{i_2}, \dots, A_{j_1}).$$

По завершении технологической операции на агрегате A_{j_1} фиксируется реализовавшийся набор параметров $(a_1, a_2, \dots, a_{n_1})$. Для данного маршрута P , вершины–развилки $v' = A_{j_1}$ нами ранее было вычислено решающее правило

$$R(a_1, a_2, \dots, a_{n_1}) \in \{0, 1\},$$

которое дает прогноз по конечной продукции при дальнейшем продолжении маршрута P .

Если $R(a_1, a_2, \dots, a_{n_1}) = 1$, то мы продолжаем дальнейшую обработку продукции согласно технологическому маршруту до достижения следующей вершины–развилки A_{j_2} . Если $R(a_1, a_2, \dots, a_{n_1}) = 0$, то дальнейшее следование по технологическому маршруту нецелесообразно, поскольку в результате его исполнения ожидается получение брака. В этом случае необходимо проверить, существуют ли другие технологические маршруты P_i , для которых уже исполненный на текущий момент времени подмаршрут маршрута P , включающий текущую вершину–развилку, также является подмаршрутом маршрута P_i . Если таких маршрутов нет, то дальнейшую обработку полученной единицы продукции следует прекратить и отправить данную единицу продукции на повторное использование в качестве исходного сырья. Если же

такие маршруты существуют, то необходимо для каждого такого маршрута P_i взять соответствующее правило R_{P_i} и выбрать те из них, для которых

$$R_{P_i}(a_1, a_2, \dots, a_{n_i}) = 1.$$

Далее, среди таких маршрутов следует отобрать наиболее эффективные и востребованные, возможно, с учетом текущих планов производства.

Заключение. В работе рассмотрены вопросы построения системы контроля качества в сложном многопередельном производстве, например, металлургическом с помощью методов аналитики больших данных на основе накапливаемых исторических технологических данных. Получены следующие основные результаты:

Введено понятие технологической пирамиды для инфраструктурного графа производства, играющее ключевую роль в предлагаемых методах оптимизации технологических параметров;

Приведена постановка задачи контроля качества продукции на основе технологических пирамид для системы назначения и выполнения технологических маршрутов;

Предлагается методология решения задачи назначения технологических маршрутов на основе методов распознавания образов с целью снижения брака производства;

Предлагается применение разработанной методологии оптимизации задачи технологических процессов к задаче прогнозирования и снижения брака в производстве.

Литература

- [1]. Гайнанов Д. Н. Комбинаторная геометрия и графы в анализе несовместных систем и распознавании образов. М.: Наука, 2014.
- [2]. Gainanov Damir. N. Graphs for Pattern Recognition. Infeasible Systems of Linear Inequalities. DeGruyter, 2016.
- [3]. Гайнанов Д. Н. О комбинаторных свойствах несовместных систем линейных неравенств и выпуклых многогранников // Математические заметки. 1985. Т. 38. № 3, С. 463–474.
- [4]. Мазуров Вл. Д. Метод комитетов в задачах оптимизации и классификации. М.: Наука, 1990.
- [5]. Мазуров Вл. Д., Хачай М. Ю. Комитеты систем линейных неравенств // АиТ. 2004. № 2. С. 43–54.
- [6]. Хачай М. Ю. Об оценке числа членов минимального комитета системы линейных неравенств // ЖВМиМФ. 1997. Т. 37. № 11. С. 1399–1404.
- [7]. Gainanov D. N. Alternative Covers and Independence Systems in Pattern Recognition // Pattern Recognition and Image Analysis. 1992. V. 2. No. 2. P. 147–160.
- [8]. Gainanov D. N., Matveev A. O. Lattice Diagonals and Geometric Pattern Recognition Problems // Pattern Recognition and Image Analysis. 1991. V. 1. No. 3. P. 277–282.