

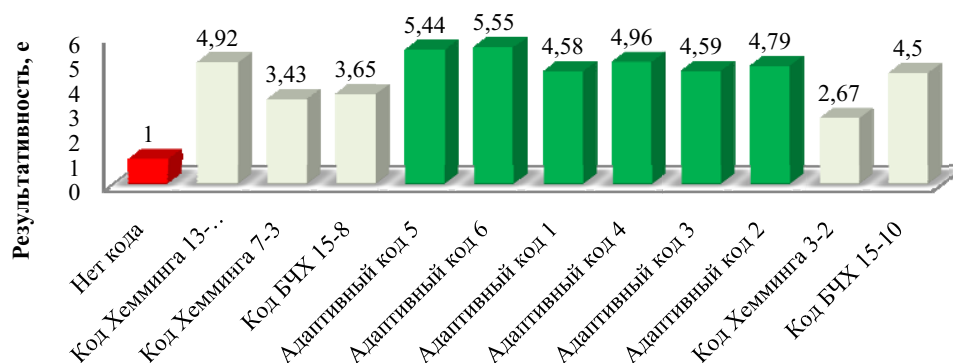


Рисунок. Результаты внедрения для неадаптивных и адаптивных кодов (средний случай)

Выводы

Итак, при оценке результативности предложенного метода и выбора параметров кодов были выявлены преимущества адаптивного матричного встраивания с точки зрения влияния на статистическую структуру контейнера. При этом было экспериментально подтверждено, что матричное встраивание не изменяет распределение вероятностей коэффициентов ДКП изображений. На основе простых помехоустойчивых кодов (Хемминга, БЧХ) были построены адаптивные коды для матричного встраивания стеганографических сообщений, проведено сравнение характеристик имеющихся и предложенных кодов, а также подтверждены предположения о том, что адаптивные коды имеют увеличенную результативность.

Разработано программное обеспечение, реализующее метод адаптивного матричного встраивания данных в изображения JPEG. Кроме того, осуществлен стегоанализ контейнеров по критериям: отношение сигнал-шум, критерий Хи-квадрат, относительная энтропия и др. По результатам исследований при отсутствии оригинального изображения вероятность ошибки I рода для системы стеганографического анализа составила 42 %, II рода – 22,3 % для внедряемого сообщения длиной 64 бита, что выше для

данных критериев по сравнению с неадаптивными методами.

Литература

1. Грибунин В.Г., Оков И.Н., Туринцев И.В. Цифровая стеганография. М., 2016. 262 с.
2. Зеленкин С.Е. Модель робастной стеганографической системы // Наукоемкие технологии. М., 2010. №9. С. 84–86.
3. Fridrich J. *Steganography in Digital Media. Principles, Algorithms and Applications*. Cambridge Univ P, 2010. 462 с.
4. Winkler A. "Advances in Syndrome Coding Based on Stochastic and Deterministic Matrices for Steganography", pp. 25, 37–55, 67–77, Dissertation, Dresden 2012.

References

1. Gribunin V.G., Okov I.N., Turincev I.V. *Cifrovaia steganografiia* [Digital steganography]. Moscow, 2016, 262 p.
2. Zelenkin S.E. Model' robastnoi steganograficheskoi sistemy [Model of robust steganographic systems]. *Naukoemkie tehnologii* [High Tech]. Moscow, 2010, №9, pp. 84–86.
3. Fridrich J. *Steganography in Digital Media. Principles, Algorithms and Applications*. Cambridge Univ P, 2010, 462 p.
4. Winkler A. "Advances in Syndrome Coding Based on Stochastic and Deterministic Matrices for Steganography", pp. 25, 37–55, 67–77, Dissertation, Dresden, 2012.

УДК 004.023

З.К. Кабаков

Череповецкий государственный университет,

Д.В. Храмешин

Выксунский металлургический завод

ИСПОЛЬЗОВАНИЕ ГЕНЕТИЧЕСКОГО АЛГОРИТМА ДЛЯ ОПРЕДЕЛЕНИЯ КЛЮЧЕВЫХ ПАРАМЕТРОВ В ЗАДАЧЕ СТАТИСТИЧЕСКОГО КОНТРОЛЯ МНОГОПАРАМЕТРИЧЕСКОГО ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА

Статья посвящена использованию эвристического алгоритма для быстрого выбора ключевых технологических параметров, наиболее полно определяющих качество поверхности горячего рулонного проката в условиях сильной скоррелированности исходных данных.

Генетический алгоритм, статистический контроль процессов, регрессионный анализ, эвристика, мультиколлинеарность.

The article deals with usage of heuristic algorithm, which makes swift choice of most valuable technological parameters within initial set. It evaluates synergy score of metallurgical parameters affecting surface defect known as 'skin' and compiles set of crucial variables for the needs of Statistical control process.

Введение

В основе статистического контроля технологического процесса заложена идея оказания предупреждающего воздействия на параметры процесса, которые отвечают за его качество при получении сигнала о нарушении. В связи с широким внедрением автоматизации в последние годы объем собираемой информации постоянно растет, и число параметров, подлежащих наблюдению, увеличивается.

Широко применяемые для статистического контроля карты Шухарта и Хотеллинга имеют недостатки в точности обнаружения достоверных выпадов, когда изучаемые с их помощью параметры оказываются коррелированы между собой, что является обычной ситуацией на практике. С увеличением числа таких показателей повышается частота возникновения ложных сигналов о нарушении. Отсюда возникает задача выделения из общей массы контролируемых технологических параметров наиболее значимых в плане оказания влияния на качество готового продукта для окончательного формирования перечня ключевых показателей, подлежащих контролю.

В данной работе производится формальная постановка такой задачи и рассказывается об опыте, полученном при ее решении с использованием генетического алгоритма на АО «Выксунский Металлургический Завод».

Основная часть

Литейно-прокатный комплекс (г. Выкса, Нижегородской области) поставляет для трубного производства АО «ВМЗ» сотни тысяч тонн горячего рулонного проката ежегодно. Высокие требования к качеству поверхности листа делают необходимым тщательный контроль процесса его производства. В 2012 г. перед Инженерно-технологическим центром АО «ВМЗ» была поставлена задача внедрения статистического контроля для осуществления непрерывного слежения за качеством рулонного проката. Целью контроля должно было стать слежение за изменением показателей производства, начиная от выплавки стали и заканчивая смоткой готового рулона, чтобы своевременно обнаруживать нестабильности процесса и оказывать управляющие воздействия для предупреждения брака.

Цеха Литейно-прокатного комплекса оснащены автоматикой, собирающей в базу данных (Oracle) более двух тысяч различных технологических параметров. Изначально предполагалось вести индивидуальный статистический контроль этих параметров при помощи автоматически строящихся карт Шухарта. С помощью таких карт достаточно просто обнаруживать неслучайные шаблоны поведения процесса [1], однако, большое количество контролируемых параметров затрудняет непрерывный контроль и повышает число диагностируемых по статистическим критериям отклонений, по факту не приводящих к браку и являющихся ложными.

В обычных случаях для снижения числа анализируемых параметров применяется хорошо зарекомен-

довавший себя метод главных компонент («При определенных условиях преобразование к главным компонентам позволяет снизить размерность задачи» [2, с. 97]), однако в данной ситуации одним из условий работы было сохранение привязки контролируемых показателей к их физическим измерениям. Иными словами, требовалось не использовать математических преобразований, приводящих несколько различных параметров к одному, а выбрать группу наиболее значимых параметров из предложенного списка.

Для оценки значимости параметров использовался расчет множественной линейной регрессии, поскольку этот метод прямо рекомендуется отраслевым стандартом ОСТ 14-1-34-90 (Статистический приемочный контроль качества металлопродукции по корреляционной связи между параметрами). В качестве независимых переменных уравнения выступали параметры процесса производства проката. В качестве оценки влиятельности параметра выступала вероятность ошибочного принятия гипотезы о его значимости (p -уровень) в том случае, если само уравнение, построенное при помощи выбранных переменных, получалось удовлетворительным.

Возникновение задачи оптимизации. Показателем качества проката был выбран процент металла марки «Сталь 20», имеющего дефект типа «Плена», относительно общего объема проката данной марки за 2011–2012 гг. Поскольку металл на Стан 1950 Литейно-прокатного комплекса идет горячим садом, было принято решение исследовать совокупное влияние всех переделов на качество проката, начиная от выплавки металла в ДСП и заканчивая смоткой. Для независимых переменных совместно с технологами завода из всех параметров производства было отобрано порядка 250 показателей, отслеживающих химический состав отдельных элементов, количество отданных легирующих материалов, температурные режимы, величины обжаты по клетям, скорости прокатки и разливы, расходы воды на охлаждение и др.

Изначально предполагалось нахождение индивидуальных корреляций каждого из технологических параметров с процентом отсортировки металла по плене для последующей оценки значимости через p -уровень коэффициента простой линейной регрессии. Такой подход позволял произвести индивидуальную оценку параметров, но не учитывал их взаимное влияние друг на друга. Для оценки вклада каждой переменной и исключения параметров, зашумляющих выборку, требовалось составление уравнения *множественной* линейной регрессии.

Анализ подготовленной выборки при помощи матрицы корреляций показал присутствие заметной взаимной связи между отдельными переменными. Известно, что высокая коррелированность может приводить к значительным случайным ошибкам [2], [3]. Для снижения зависимости сигнального фактора от мультиколлинеарности было принято решение заменить массив карт Шухарта обобщенной контрольной картой Хотеллинга («С увеличением сте-

пени коррелированности карта Хотеллинга становится эффективней карт Шухарта» [2, с. 40]).

Поскольку коллинеарность переменных в ряде случаев может приводить к невозможности расчета обратной матрицы во время вычисления регрессии методом наименьших квадратов, это может привести к необоснованному исключению независимых переменных из расчета [3]. Для исключения подобного рода ошибок требуется многоплановая проверка поведения каждого параметра в комбинации с другими. С составлением комбинаций параметров для подобного рода проверок хорошо зарекомендовали себя эвристические методы. Поскольку преобразование массива параметров к главным компонентам для снижения размерности регрессионной матрицы исключалось формулировкой задачи, то для нахождения оптимальной выборки было принято решение использовать математический аппарат генетических алгоритмов.

Отсев малозначимых факторов оказался необходим и для расчета обобщенного отклика карты Хотеллинга («С увеличением числа показателей эффективность карты снижается: средняя длина серий без возникновения сигнала о нарушении увеличивается» [2, с. 41]). Таким образом, окончательной задачей генетического алгоритма стало выделение значимых параметров, чье воздействие наиболее полно описывало бы полученное с их помощью регрессионное уравнение.

Максимальный размер регрессионной матрицы для одного прохода позволял включить не более 30 переменных, поскольку к достоверности искомого уравнения предъявлялись повышенные требования. Согласно требованиям ОСТ 14-1-34-90 в исходную выборку были включены данные от 200 различных плавок, произведенных в течение последнего года. Высокое число независимых переменных снижает значимость полученных результатов («Систематическая ошибка приводит к завышению степени тесноты связи при приближении числа оцениваемых параметров к объему выборки» [2, с. 121]). Для учета влияния ошибки остаточной дисперсии в качестве оценки качества регрессионного уравнения использовался скорректированный коэффициент детерминации.

Целью оптимизации стало нахождение такого набора переменных, который обеспечил бы наибольший r -уровень для каждого из отдельных коэффициентов уравнения регрессии при достаточно высоком скорректированном коэффициенте множественной детерминации.

Формализация и описание работы генетического алгоритма. Генетический алгоритм представляет собой модель последовательной эволюции популяции отдельных особей. В каждый момент времени популяция состоит из k особей, главной характеристикой каждой из которых является хромосома. Хромосома представляет собой набор параметров P , фиксированной величины n , определяющий приспособленность данной особи. $f_k = f(P_1, P_2, \dots, P_n)$. Цель работы алгоритма состоит в нахождении за время эволюции такой особи, чья приспособленность ока-

жется максимальной, т.е. оптимизация функции $f_k(P_n)$.

Эволюция популяции происходит путем отсева слабоприспособленных особей и добавления новых, чьи хромосомы получаются путем соединения (скрещивания) разных частей хромосом других особей. В процесс скрещивания свой вклад вносят фактор случайности и фактор мутации, с некоторой вероятностью изменяющий один из параметров P готовой хромосомы.

Работа генетического алгоритма состоит из нескольких шагов.

Шаг 1. Создание начальной популяции, хромосомы особей которых будут сформированы случайным образом.

Шаг 2. Вычисление приспособленности каждой особи популяции $f_k(P)$.

Шаг 3. Отсев особей с наиболее низким показателем $f_k(P)$.

Шаг 4. Скрещивание особей с высоким показателем $f_k(P)$ для получения новых особей до достижения исходного размера популяции k .

Шаг 5. Повтор шагов 1–4 до тех пор, пока не прекратится рост максимального показателя $f_k(P)$ в популяции.

В текущей реализации генетического алгоритма для отсева и отбора особей использовался пропорционально вероятностный отбор, в котором вероятность исключить особь из популяции и использовать ее для скрещивания была пропорциональна ее приспособленности.

В качестве схемы скрещивания использовалось одноточечное скрещивание хромосом, при котором первая и вторая половины хромосомы новой особи берутся от двух разных исходных особей.

Вероятность мутации была установлена на уровне 0,1 %. Мутация предполагала собой замену местами двух случайных параметров P .

В качестве параметров P использовались отдельные технологические параметры. Количество элементов хромосомы n было установлено равным 30, согласно ранее определенному максимальному числу независимых переменных. Алгоритм начального формирования хромосом был составлен таким образом, чтобы назначать с определенной долей вероятности вместо параметров пустые значения. Кроме того, в результате воздействия мутаций и скрещиваний у новых особей возникали дубли параметров P внутри одной хромосомы (при вычислении функции приспособленности f повторные и пустые параметры игнорировались), поэтому фактическое число «уникальных» параметров в каждой новой хромосоме изначально и в процессе эволюции могло оказаться и менее 30. Размер популяции был установлен в 20 особей для того, чтобы каждый из 250 параметров встречался как минимум у 2 особей начальной популяции.

Критерием оптимизации функции f выступало произведение коэффициента множественной корреляции линейного уравнения регрессии с параметрами P и средней скорректированной суммы r -уровней независимых переменных и свободного члена полученного уравнения. Дополнительным условием было

резкое снижение приспособленности особи, если коэффициент множественной корреляции уравнения оказывался ниже 0,4.

Информация о технологических параметрах производства проката литейно-прокатного комплекса и результатах аттестации качества поверхности готового листа, необходимая для создания исходной выборки, хранится в базе данных Oracle. Работа алгоритма подразумевала многократные вычисления уравнения регрессии на отобранных данных. Для вычисления результатов работы генетического алгоритма в базе данных был создан программный пакет, моделирующий эволюцию популяции по указанному ранее алгоритму, написанный на языке PL/SQL. После формирования текущей популяции для каждой особи k производилось преобразование параметров хромосомы в набор переменных (технологических параметров), после чего происходило вычисление функции f , включающее следующие шаги:

Шаг 1. Выбор данных по указанным параметрам и сведение их в единую таблицу.

Шаг 2. Фильтрация некорректных значений и составление регрессионной матрицы.

Шаг 3. Расчет уравнения регрессии методом наименьших квадратов при помощи математического пакета для работы с матрицами Oracle UTL_NLA.

Шаг 4. Отсев переменных с высоким значением p -уровня (выше 0,6) и переменных, не удовлетворяющих проверки на критическое значение t -статистики.

Шаг 5. Повтор шагов 3, 4 до отсева всех неудовлетворяющих поставленным условиям переменных.

Шаг 6. Проверка значимости полученного уравнения по критерию Фишера. При удовлетворительном результате вычисление коэффициента приспособленности.

Результаты работы алгоритма. После запуска генетический алгоритм совершил 41 цикл вычислений. Причиной остановки было выполнение одного из заранее введенных условий: достижение значения

критерия оптимизации и скорректированного коэффициента множественной детерминации на уровне 0,6 (после округления) в течение трех циклов подряд. Сокращенные результаты работы алгоритма приведены в таблице.

На приведенном рисунке видно, что скорректированный коэффициент множественной детерминации возрастал с самого начала работы алгоритма, критерий оптимизации, по которому оценивалась текущая популяция, хотя и проявил заметный тренд к возрастанию, периодически снижался. Такая нестабильность объясняется попаданием в набор независимых переменных отдельных особей параметров, не оказывающих заметное влияние на количество отсортировки по плене, либо зашумляющих выборку и снижающих значимость других переменных. Со временем особи, содержащие в хромосомах такие параметры, были исключены из популяции, и критерий оптимизации стал вести себя более предсказуемо.

Оценка качества рассчитанного уравнения регрессии производилась визуально по отсортированному графику остатков. Алгоритм был настроен таким образом, чтобы при наличии единичных выпадов расчетных и фактических значений за границы 6 стандартных отклонений исключать породившие их записи из расчета для улучшения исходной выборки. Связь скорректированного коэффициента множественной детерминации и критерия оптимизации генетического алгоритма представлена на рисунке.

По результатам работы было отобрано 30 параметров, оказывающих ключевое воздействие на количество брака, и рассчитана статистика Хотеллинга для вывода предупреждающей контрольной карты. В качестве дополнения к ключевым параметрам было выделено еще 7 технологических параметров, не попавших в финальный список, взятых из хромосом других высокоприспособленных особей финальной популяции.

Таблица

Результаты работы генетического алгоритма

№ шага	Коэффициент корреляции множественной регрессии	Наивысшее значение критерия оптимизации в популяции	Стандартное отклонение уравнения регрессии	Скорректированный коэффициент детерминации множественной регрессии
1	0,4	0,06	0,232034	0,11186
2	0,412	0,105	0,212857	0,139456
3	0,414	0,21	0,836805	0,143575
4	0,417	0,092	0,806045	0,145189
5	0,424	0,108	0,659647	0,137768
...				
37	0,746	0,468	7,726967	0,519546
38	0,755	0,437	3,967811	0,47936
39	0,791	0,582	3,807414	0,603822
40	0,811	0,603	5,49596	0,632505
41	0,87	0,598	2,50577	0,635471

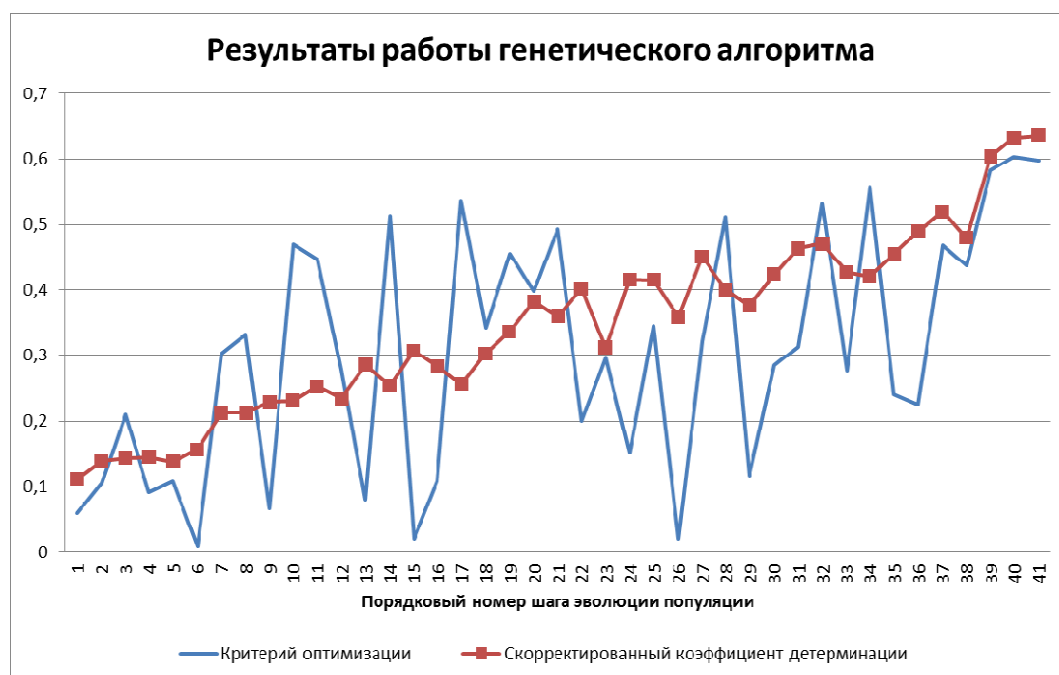


Рисунок. Изменение критерия оптимизации и скорректированного коэффициента множественной детерминации регрессионного уравнения по шагам

Выводы

Использование генетического алгоритма в работе позволило в приемлемый срок провести статистический анализ большого объема технологических параметров, многие из которых были заметно коррелированы между собой, и с достаточной достоверностью установить форму зависимости 30 из них на количество возникающего брака.

Эта информация была использована для расчета статистики Хотеллинга, что позволило ввести начальный уровень статистического контроля в виде единой контрольной карты обобщенного отклика. Информация о форме зависимости, представленная в виде уравнения множественной линейной регрессии, была использована в дальнейшей работе для построения модели прогнозирования качества металла. Программа, использованная для расчета ключевых параметров, получила дальнейшее развитие в моделях неразрушающего контроля механических свойств горячего проката и модели прогнозирования возникновения коррозионноактивных неметаллических включений для других марок стали.

Литература

1. Государственный стандарт Российской Федерации ГОСТ Р 50779.42 – 99 (ИСО 825 – 91) Статистические методы. Контрольные карты Шухарта.
2. Клячкин В.Н. Модели и методы статистического контроля многопараметрического технологического процесса. М., 2011. 195 с.
3. Belsley D. Conditioning diagnostics: collinearity and weak data in regression. N.Y., 1991.

References

1. Gosudarstvennyi standart Rossiiskoi Federacii GOST R 50779.42–99 (ISO 825–91) Statisticheskie metody. Kontrol'nye karty Shuharta [State standard of the Russian Federation GOST R 50779.42-99 (ISO 825-91) Statistical methods. Shewhart control charts].
2. Kliachkin V.N. Modeli i metody statisticheskogo kontrolya mnogoparametricheskogo tehnologicheskogo processa [Models and methods of statistical control of multivariable process]. Moscow, 2011, 195 p.
3. Belsley D. Conditioning diagnostics: collinearity and weak data in regression. N.Y., 1991.