

УДК 004.042:655.2

У статті розроблено математичну модель перетворення даних в поліграфічній системі на додрукарському етапі виробництва. Показано, що дана модель є адекватною і дозволяє проводити імітаційне моделювання інформаційних потоків у поліграфічних системах. Отримані результати можуть застосовуватися для проектування продуктивних, надійних обчислювальних мереж для поліграфічних підприємств і оптимізації робочих потоків на додрукарському етапі виробництва

Ключові слова: математична модель, інформаційний потік, перетворення даних, поліграфічна система

В статье разработана математическая модель преобразования данных в полиграфической системе на допечатном этапе производства. Показано, что данная модель является адекватной и позволяет проводить имитационное моделирование информационных потоков в полиграфических системах. Полученные результаты могут применяться для проектирования производительных, надёжных вычислительных сетей для полиграфических предприятий и оптимизации рабочих потоков на допечатном этапе производства

Ключевые слова: математическая модель, информационный поток, преобразование данных, полиграфическая система

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ПРЕОБРАЗОВАНИЯ ДАНЫХ В ПОЛИГРАФИЧЕСКОЙ СИСТЕМЕ

В. П. Авраменко

Доктор технических наук, профессор*

А. К. Парамонов

Аспирант*

E-mail: paramonov.a.k@bk.ru

*Кафедра медиасистем и технологий

Харьковский национальный университет
радиоэлектроники

пр. Ленина, 14, г. Харьков, Украина, 61166

1. Введение

В настоящее время в полиграфии применяется большое количество самых разнообразных систем обработки информации от простых настольных издательских систем до мощных многомашинных комплексов с разветвленной сетевой структурой. При этом реализуются различные варианты технологических процессов от изготовления оригинал-макета до вывода информации в виде готовой печатной продукции на цифровых печатных машинах.

Современное полиграфическое оборудование автоматизировано и имеет возможность взаимодействовать с компьютерными системами, что позволяет его интегрировать с общей системой управления производственными процессами полиграфического производства. В связи с этим, возрастает значение потоков цифровой информации. Поэтому встаёт проблема анализа потоков информации и разработка математической модели преобразования данных в полиграфической системе.

2. Анализ литературных данных и постановка проблемы

Под требованиями рынка полиграфические системы перешли из узко специализированных в универ-

сальные, выпускающие разнообразную полиграфическую продукцию. В настоящее время полиграфические системы часто являются территориально распределёнными. Обзор научных публикаций показывает, что существующие математические модели не достаточно соответствуют современным полиграфическим системам.

Потоки информации в полиграфической системе разделяют на два типа [1]:

- сигналы контроля и массивы данных управления, которые используются для настройки оборудования и осуществления требуемого техпроцесса,
- отображаемая информация, описывающая содержимое печатной продукции (текст, изображения, данные форматирования и расположения и т.п.).

Объёмы отображаемой информации значительно превосходят объёмы данных контроля и управления, а стремительная тенденция к повышению качества печатной продукции приводит к тому, что вычислительная сеть полиграфического предприятия практически полностью загружена передачей отображаемых данных [2]. Поэтому актуальной является задача разработки и применение адекватных математических моделей информационных потоков при проектировании сетевой инфраструктуры полиграфической системы, что приведёт к оптимальному использованию ресурсов и предотвращению перегрузок на допечатном этапе производства.

3. Цель и задачи исследования

В [3] сформирована схема информационных потоков в полиграфической системе и выполнен их анализ. Целью данной работы является синтез математической модели преобразования данных в полиграфической системе на допечатном этапе производства.

4. Модель преобразования данных

Маршруты потоков отображаемой информации разветвлённые и могут проходить через узлы и их группы по несколько раз вследствие повторного редактирования и итеративности процесса обработки данных. В процессе анализа выделен унифицированный поток (рис. 1), который полностью перекрывает весь технологический процесс и через который может быть выражено всё множество цифровых информационных потоков в полиграфической системе.



Рис. 1. Схема информационного потока на допечатном этапе

Данные поступают от заказчика через интернет в базу данных, которая находится на сервере хранения и систематизации информации. В простейшем случае это FTP сервер. На рабочих станциях установлена настольная издательская система или клиентские компоненты серверной издательской системы. Сотрудник полиграфического предприятия обрабатывает поступивший в базу данных заказ на рабочей станции, где происходит преобразование данных. Далее информация о печатной продукции отправляется на сервер обработки изображений, где она подготавливается для выводного устройства. На схеме (рис. 1) этот сервер обозначен как RIP (англ. обработчик растрового изображения), т. к. растривание наиболее ресурсоемкая операция.

Основной метод исследования информационных систем состоит в моделировании процессов. Модели должны отражать структуру системы и воспроизводить основные её характеристики. Модель представляет собой специально синтезированный для удобства исследований объект, обладающий необходимой степенью подобия исходному объекту. При этом степень подобия должна быть адекватна целям исследования [4]. Адекватной считается модель, которая в определённой мере соответствует объекту, приводит к нужной цели и позволяет получить необходимые на практике результаты [5]. Количественным и качественным отражением степени достижения системой поставленных перед ней целей являются критерии. Критериев может быть несколько. Многокритериальность - один из способов повышения адекватности описания цели.

Для синтеза математической модели численной характеристикой объёма информации выбран размер

файлов, входящих и передаваемых по сети полиграфической системы. Размер файлов, приходящих из Интернета и соответствующий одному заказу, является случайной величиной. На основе собранной статистики найдена функция распределения размера данных для одного заказа. В качестве нулевой гипотезы о распределении случайной величины рассмотрены те распределения, у которых аргумент принимает только положительные значения, так как размер данных не может быть отрицательным. Параметры теоретического распределения вычислены по принципу максимума правдоподобия для всех подходящих непрерывных распределений и занесены в табл. 1.

Проверка правильности подбора теоретического распределения произведена по критерию согласия Колмогорова. Распределение, которое лучше всего согласуется с выборкой, имеет наибольшее критическое значение (р-значение) – вероятность совершить ошибку первого рода при отклонении нулевой гипотезы.

Критические уровни значимости для соответствующих распределений представлены в табл. 1. Также произведена проверка по критерию согласия Пирсона с уровнем значимости 0.1. Разности между критическим и выборочным значениями χ^2 занесены в табл. 1 [6, 7].

Таблица 1

Подбор теоретического распределения

Распределение	Параметры		Критерий Колмогорова	Критерий Пирсона	
			Р-значение	$\chi^{2кр} - \chi^2$	
Экспоненциальное	$\mu=53759.52$		0.00000	-41.72063	
Гамма	$a=0.44602$	$b=120531.41$	0.07066	-2.79118	
Логнормальное	$\mu=9.4424$	$\sigma=1.87548$	0.98617	5.70197	
Рэлеевское	$b=82220.91$		0.00000	-311.036	
Вейбулловское	$a=31867.68$	$b=0.575918$	0.52051	2.96224	
Экстремальных значений обобщённое	$k=1.544$	$\sigma=9525$	$\mu=5727$	0.53879	4.87752
Логлогистическое	$\mu=9.46938$	$\sigma=1.09385$	0.98247	5.16387	

Следовательно, для генерирования размера входных данных лучше всего подходит логнормальное распределение с вероятностью 0.98617 по критерию Колмогорова не совершить ошибку первого рода и наибольшей разностью по критерию Пирсона. Плотность вероятности и функция распределения логнормальной случайной величины соответственно имеют вид:

$$F_X(x; \mu, \sigma) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{\ln x - \mu}{\sigma\sqrt{2}}\right),$$

$$f_X(x; \mu, \sigma) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

где erf() – функция ошибки, вычисляемая по формуле $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$.

На рис. 2 приведена гистограмма статистических данных и плотность вероятности подобранного распределения. На рис. 3 показаны практическая и теоретическая функции распределения.

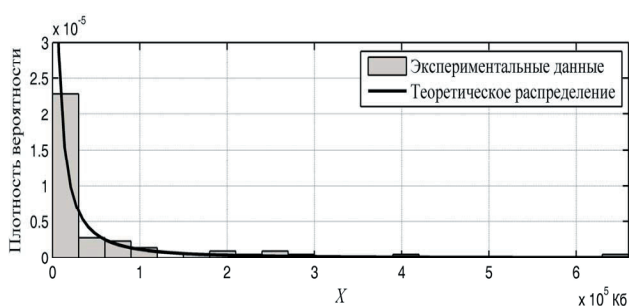


Рис. 2. Гистограмма статистики и плотность вероятности теоретического закона распределения

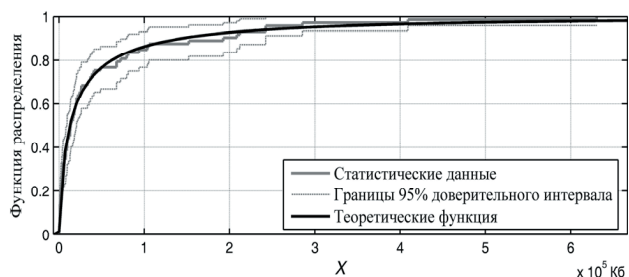


Рис. 3. Практическая и теоретическая функции распределения

Во многих языках программирования стандартным является только генератор равномерно распределённых псевдослучайных чисел, поэтому для удобства моделирования на ПК множество поступающих в полиграфическую систему заказов с логнормально распределёнными данными представляется через равномерное распределение. По свойству логнормального распределения, если случайная величина X распределена логнормально с параметрами μ и σ , то она связана с нормально распределённой переменной Z с теми же параметрами соотношением $X = \exp(Z)$.

Нормально распределённая случайная величина моделируется с помощью равномерного распределения согласно преобразованию Бокса – Мюллера [8], которое является точным в отличие от методов, основанных на центральной предельной теореме. Для повышения быстродействия применена моди-

фикация данного метода с меньшей вычислительной сложностью [9, 10].

Таким образом, функция для генерирования размера входных данных, связанных с одним заказом имеет вид:

$$X = \exp\left(\mu + \sigma \cdot a \cdot \sqrt{\frac{-2 \cdot \ln s}{s}}\right)$$

$$\text{и } X = \exp\left(\mu + \sigma \cdot b \cdot \sqrt{\frac{-2 \cdot \ln s}{s}}\right),$$

$$s = a^2 + b^2, \quad 0 < s \leq 1,$$

где a и b – независимые случайные величины, равномерно распределённые на отрезке $[-1, 1]$.

Анализ плотности вероятности подобранного теоретического распределения показал, что вероятность появления заказа с размером данных большим $6 \cdot 10^5$ Кб равна 0.0197, поэтому для удобства построения математической модели размер данных для одного заказа был ограничен полузакрытым диапазоном $(0; 6 \cdot 10^5]$ Кб.

На первом этапе обработки заказа данные оператором приводятся к необходимому виду, форме представления и преобразуются в формат, совместимый с сервером обработки изображений. Был проведён регрессионный анализ преобразования данных. Для оценки качества подбора функциональной зависимости применён коэффициент детерминации. Коэффициент детерминации (R-квадрат) – это доля дисперсии зависимой переменной, объясняемая рассматриваемой моделью зависимости, то есть объясняющими переменными. Коэффициент детерминации для модели с константой принимает значения от 0 до 1.

Чем ближе значение коэффициента к 1, тем сильнее зависимость. При оценке регрессионных моделей это интерпретируется как соответствие модели данным.

В данной работе, чтобы число регрессоров (факторов) не влияло на статистику R^2 , качество подбора анализировалось по скорректированному коэффициенту детерминации (R^2_{adj}), который рассчитывается по формуле:

$$R^2_{adj} = 1 - \frac{SSE/(n-k)}{SST/(n-1)} \leq R^2,$$

где SSE – сумма квадратов остатков регрессии, SST – общая сумма квадратов, n – количество наблюдений, k – количество параметров.

Коэффициент детерминации не может в полной мере отображать качество модели, поэтому логическая и смысловая адекватность модели имеют первостепенную важность.

При выборе наилучшего варианта учитывалось также поведение функции в области и за областью определений модели, уход в отрицательные значения и степень полинома.

В табл. 2 приведены значения скорректированного коэффициента детерминации моделей для первого и второго этапа преобразования данных.

Таблица 2

Значения скорректированного коэффициента детерминации

Тип уравнений	Число коэффициентов	R ² _{adj}
Первый этап преобразования		
Полином 1 степени	2	0.84809
Полином 2 степени	3	0.96866
Полином 3 степени	4	0.97199
Полином 4 степени	5	0.97556
Полином 5 степени	6	0.97553
Второй этап преобразования		
Полином 1 степени	2	0.98067
Полином 2 степени	3	0.99494
Полином 3 степени	4	0.99194
Полином 4 степени	5	0.98883
Полином 5 степени	6	0.95730

На рис. 4 приведена найденная функция изменения размера данных после обработки заказа оператором. Коэффициенты подобранного полинома 4-й степени $Y_1(x) = a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0$ с 95% доверительными интервалами:

$$a_4 = -6.016 \cdot 10^{-17} (-9.622 \cdot 10^{-17}, -2.41 \cdot 10^{-17}),$$

$$a_3 = 5.704 \cdot 10^{-11} (1.785 \cdot 10^{-11}, 9.65 \cdot 10^{-11}),$$

$$a_2 = -6.647 \cdot 10^{-6} (-1.938 \cdot 10^{-5}, 6.085 \cdot 10^{-6}),$$

$$a_1 = 1.171 (-0.135, 2.478),$$

$$a_0 = 20220 (-2390, 4.283 \cdot 10^4).$$

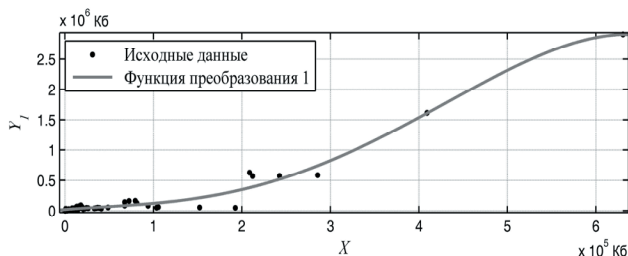


Рис. 4. Функция преобразования данных после обработки оператором

На следующем этапе поток информации поступает на сервер обработки изображений, где претерпевает значительные изменения. После первого преобразования входной размера данных для одного заказа на втором этапе ограничен диапазоном (0; 3·10⁶] Кб.

Из табл. 2 видно, что экспериментальные данные лучшим образом описывает полином второго порядка, поэтому для модели использовано квадратное уравнение вида $Y_2(y_1) = a \cdot y_1^2 + b \cdot y_1 + c$ с коэффициентами и 95% доверительными интервалами:

$$a = -8.772 \cdot 10^{-7} (-9.029 \cdot 10^{-7}, -8.515 \cdot 10^{-7}),$$

$$b = -2.881 (2.816, 2.946),$$

$$c = 1424.1 (-7438, 8286).$$

На рис. 5 показана функциональная зависимость преобразования потока информации сервером обработки изображений.

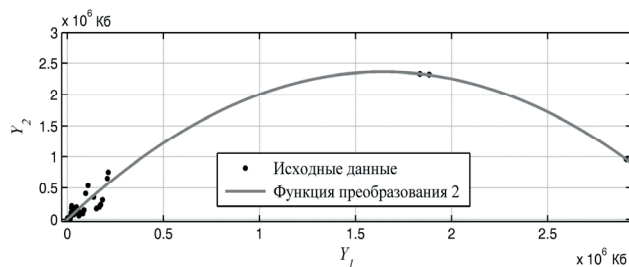


Рис. 5. Функция преобразования данных сервером обработки изображений

5. Выводы

В данной работе выполнен синтез математической модели преобразования данных в полиграфической системе на допечатном этапе производства, оценена адекватность полученной модели, проанализировано влияние параметров потоков информации на объемы информации, передаваемой на допечатном участке полиграфической системы, произведена адаптация модели для моделирования в вычислительных средах общего назначения.

Литература:

- Гехман, Ч. Рабочий поток [Текст]: пер. с англ. / Чак Гехман. – М: Издательство МГУП, 2004. – 252 с.
- Дурняк, Б. В. Системний аналіз та оптимізація параметрів книжкових видань [Текст]: Монографія. / Б. В. Дурняк, І. В. Піх, В. М. Сенківський. – Львів: Українська академія друкарства, 2006. – 197 с.
- Парамонов, А. К. Анализ потоков информации в полиграфической системе на допечатном этапе производства [Текст] / А. К. Парамонов // 17-й Международный молодежный форум «Радиоэлектроника и молодежь в XXI веке». Сб. материалов форума. Т.6. – Харьков: ХНУРЭ. 2013. – С. 468-469
- Миротин, Л. Б. Системний аналіз в логістическі [Текст]: Учебник. / Л. Б. Миротин, Ї. Э. Ташбаев. – М.: Издательство «Экзамен», 2002. – 480 с.
- Сорока, К. О. Основи теорії систем і системного аналізу [Текст]: Навч. посібник / К. О. Сорока. – Х.: Тимченко, 2005. – 288 с.

6. Ниворожкина, Л. И. Теория вероятностей и математическая статистика [Текст] / Л. И. Ниворожкина, З. А. Морозова. – М.: Эксмо, 2008. – 432 с.
7. Иглин, С. П. Теория вероятностей и математическая статистика на базе MATLAB [Текст]: Учебное пособие / С. П. Иглин. – Харьков: НТУ "ХПИ", 2006. – 612 с.
8. Box, G. E. P. A Note on the Generation of Random Normal Deviates [Текст] / G. E. P. Box and Mervin E. Muller // The Annals of Mathematical Statistics. – 1958. – Vol. 29, No. 2. – pp. 610-611.
9. Bell, J. 'Algorithm 334: Normal random deviates' [Текст] / J. Bell Communications of the ACM. – vol. 11, No. 7. – 1968.
10. Knopp, R. 'Remark on algorithm 334 [G5]: normal random deviates' [Текст] / R. Knopp // Communications of the ACM. – vol. 12, No. 5. – 1969.

Розглянуто дві структурні моделі інформаційних перетворень вихідних сигналів дубльованих вимірювальних перетворювачів. Теоретично обґрунтовано диференційний метод бездемонтажного контролю похибок перетворювачів по різниці залишкових дисперсій структурних моделей. Представлено результати статистичного моделювання інформаційних перетворень при контролі похибок по випадковим вимірювальним сигналам

Ключові слова: вірогідність, інформація, контроль, модель, дисперсія, фактор, вимірювальний перетворювач, невизначеність

Рассмотрены две структурные модели информационных преобразований выходных сигналов дублированных измерительных преобразователей. Теоретически обоснован дифференциальный метод бездемонтажного контроля погрешностей преобразователей по разности остаточных дисперсий структурных моделей. Представлены результаты статистического моделирования информационных преобразований при контроле погрешностей по случайным измерительным сигналам

Ключевые слова: достоверность, информация, контроль, модель, дисперсия, фактор, измерительный преобразователь, неопределенность

УДК 621.317

ОЦЕНКА ЭФФЕКТИВНОСТИ ПАРАМЕТРИЧЕСКОГО КОНТРОЛЯ МЕТОДОМ СТАТИСТИЧЕСКОГО МОДЕЛИРОВАНИЯ

П. Ф. Шапов

Доктор технических наук, профессор*

E-mail: shapov.p.f@gmail.com

И. И. Камбаев

Аспирант*

*Кафедра «Информационно-измерительные технологии и системы»

Национальный технический университет
«Харьковский политехнический институт»

ул. Фрунзе, 21, г. Харьков, Украина, 61002

E-mail: vanadiy2311@ukr.net

И. В. Тищенко

Кафедра искусственного интеллекта

Харьковский национальный университет

радиоэлектроники

пр. Ленина, 16, г. Харьков, Украина, 61166

E-mail: lvanstud91@gmail.com

1. Введение

Основой повышения эффективности промышленного производства, качества продукции, надежности промышленного оборудования является совершенствование методов неразрушающего контроля и функциональной диагностики объектов и параметров контроля [1]. В данной статье обсуждается применение статистических методов анализа информационных систем, используемых для контроля нарушений точности первичных измерительных преобразователей.

2. Постановка проблемы

Контроль точности измерительных преобразований, осуществляемый традиционными методами практической метрологии, наиболее эффективен в условиях лабораторного метрологического эксперимента [2]. При этом поверка или калибровка средств измерений проводится по эталонным входным сигналам, режим измерений – статистический, а само средство измерения с эксплуатации, фактически, снимается. Если для контроля точности используют структурно-алгоритмические методы [3], то для