



где D - множество гармоник, соответствующее полосе пропускания частотного фильтра при снятии круглограммы; E - интервал, равный всему диапазону частот ω_2 модельной поверхности.

Описанная математическая модель использована для расчета характеристик алгоритмической составляющей погрешности метода.

Литература

1. Белопухов, В. Н. Лабораторный макет лазерного кругломера [Текст] / В. Н. Белопухов, М. Ю. Грачев, М. В. Лесников, О. А. Заякин, Ю. О. Понявин // ПИТ-2012: сб. науч. тр. - Самара: Изд. СНЦ РАН, 2012. - С. 10 - 13.
2. Топорец, А. С. Оптика шероховатой поверхности [Текст] / А. С. Топорец. - Л.: Машиностроение. Ленингр. отделение, 1988. - 191 с.
3. Беннет, Дж. М. Шероховатость поверхности и рассеяние [Текст] / Дж. М. Беннет, М. Маттсон. - Вашингтон: Оптическое общество Америки, 1993. - 119 с.
4. Дунин-Барковский, И. В. Измерения и анализ шероховатости, волнистости и некруглости поверхности [Текст] / И. В. Дунин-Барковский, А. Н. Карташова. - М.: Машиностроение, 1978. - 232 с.

И.Д. Зиновьев

ОБНАРУЖЕНИЕ КРИТИЧЕСКИХ СИТУАЦИЙ В РАБОТЕ КОМПЬЮТЕРНОЙ СИСТЕМЫ ПО ДАННЫМ МОНИТОРИНГА НА ОСНОВЕ МНОГОМЕРНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА

(Самарский государственный аэрокосмический университет имени академика
С.П. Королева (национальный исследовательский университет))

Данная работа посвящена проблеме обработки данных мониторинга компьютерной системы, анализа этих данных. Показаны различия между контролем качества серийной продукции и компьютерной системы, предложены пути решения этих проблем.

Цель: сокращение числа контролируемых параметров компьютерной системы и анализ агрегированных показателей

В сфере бизнес решений очень часто создаются большие кластеры из вычислительных машин, выполняющих одинаковые или специфичные для каждой машины функции.

Выход из строя одной из машин может привести к потере производительности или, в случае уникальности узла, вообще к отключению системы. Простой таких систем обходится очень дорого, а в некоторых случаях (например, банки и платежные системы) и вовсе не допустим. Для предотвращения наступления таких ситуаций необходимо вести наблюдение за системой. Для этой цели существуют системы мониторинга.



Для контроля качества продукции при серийном производстве используется статистический контроль технологического процесса на основе карт Шухарта [1]. Обнаружение нарушений процесса, как правило, проводится на основе выхода контролируемой статистики за заданные границы карты. Большинство современных систем мониторинга работают только с простыми картами Шухарта, отслеживая выход контролируемого параметра за границы разрешенной зоны.

Вместе с тем, для одномерного контроля широко используются и другие критерии, в частности, поиск структур специального вида на контрольной карте: если на карте имеет место структура (набор точек), вероятность появления которой близка к вероятности ложной тревоги, то это свидетельствует о нарушении процесса.

Если абстрагироваться от структуры хранения, данные системы мониторинга можно представить в виде матрицы чисел, в которой каждый столбец соответствует какому-то контролируемому параметру, а строка – моменту времени, в которой параметр был получен. Если какой-то параметр в определенный момент времени не собирался, то в соответствующей ячейке стоит значение null. Минимальным шагом по времени является 1 секунда. Пример матрицы приведен в таблице 1.

Таблица 1. Пример матрицы данных системы мониторинга

Время, с	Загрузка процессора, %	Использование ОЗУ, байт	Свободное место на диске, %	...	Параметр №600	...
10	10	null	27	...	null	...
11	null	259652128	null	...
12	null	null	27	...	153,9	...
15	37	null	null	...
...

Однако, ввиду специфики предметной области, большого количества факторов и не учитываемой обычно коррелированности факторов, большинство наработок связанных с продвинутым анализом карт Шухарта нельзя просто применить к задаче мониторинга компьютерной системы. Основные проблемы:

1. Большая размерность. При контроле качества продукции количество измеряемых параметров редко превышает несколько десятков. В случае компьютерной системы это число может измеряться сотнями и тысячами.

2. Несинхронизированность данных. При контроле качества какой-то детали, мы можем снять все показатели сразу. При мониторинге компьютерной системы эти параметры получаются периодически, причем период у каждого параметра свой.

3. Переменный характер корреляций. При контроле производства все корреляции заранее определены производственным процессом. В случае компьютерной системы корреляционная матрица может очень сильно меняться в зависимости от выполняемой системой задачи. Например, в сервисах, предо-



ставляющих услуги облачных вычислений физическая машина может перебрасываться с одной задачи на другую, из одного логического кластера в другой. В такой ситуации значения коэффициентов корреляции поменяются.

4. Переменное число параметров. В любой момент в компьютерную систему может быть добавлена новая машина, параметры которой сразу начинают мониториться. Также машины можно удалять, при этом параметры этой машины перестают рассматриваться.

5. Отсутствие обучающей выборки. Отчасти следует из предыдущего пункта. При добавлении новых параметров отсутствует статистика по ним.

Предлагаемые пути решения проблем

1. Снижение размерности задачи. Предлагается использовать метод главных компонент [2] и анализировать не сами данные, а главные компоненты. Это также позволит нам находить суммарную ошибку, «размазанную» по всем факторам.

2. Если просто посчитать корреляционную матрицу по данным в том виде, в котором они получаются, то все факторы окажутся некоррелированными ввиду их разброса по времени. Поэтому предлагается на всех промежуточных моментах времени использовать предыдущее полученное значение. Также можно при получении следующего значения применять интерполяцию и пересчитывать корреляционную матрицу на этом промежутке.

3. Для своевременной реакции на смену взаимосвязей между параметрами необходимо отбрасывать из рассмотрения устаревшие данные. Для этого предлагается считать корреляционную матрицу не по всем данным системы, а только для некоторого окна, размер которого должен выбираться достаточно большим, чтобы по нему можно было делать статистические выводы, и достаточно малым, чтобы устаревшие данные не оказывали влияния на корреляционную матрицу. По оценкам такой размер составляет 10 минут. Однако из-за введения концепции окна в алгоритм расчета корреляционной матрицы возникает проблема с данными, которые не меняются внутри окна. Дисперсия таких факторов равна нулю, что ведет к неопределенностям. Такие данные мы считаем несущественными и предлагаем не рассматривать.

4. Чтобы накопить определенную статистику по свежедобавленным факторам предлагается вводить их в рассмотрение по прошествии времени, равного одному окну, с момента их появления. В этом случае мы будем располагать достаточно статистикой по ним.

Алгоритм работы

1. Из системы мониторинга получается набор данных определенного размера, равного размеру окна. При этом производится их первичная обработка: null заменяются на предыдущие значения, факторы без изменений отбрасываются.

2. По подготовленным данным считается корреляционная матрица по формуле:



$$r_{i,j} = \frac{\sum_{k=0}^n \left((X_{i,k} - \bar{X}_i) (X_{j,k} - \bar{X}_j) \right)}{\sqrt{\sum_{k=0}^n (X_{i,k} - \bar{X}_i)^2 \sum_{k=0}^n (X_{j,k} - \bar{X}_j)^2}}$$

$$\bar{X}_i = \frac{1}{n} \sum_{k=0}^n X_{i,k}$$

где $X_{i,k}$ – значение i -го фактора в момент времени k ,
 значение i -ой выборки, n – объем выборки (размер окна).

3. Для матрицы r ищутся формулы наибольших главных компонент так, чтобы в них было сконцентрировано заданное количество энергии, выбираемое так, чтобы значительно уменьшить размерность задачи. Для этого необходимо найти p наибольших собственных значения корреляционной матрицы и соответствующие им собственные векторы. Количество необходимых собственных значений p можно найти из соотношения

$$I = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_p}{\lambda_1 + \lambda_2 + \dots + \lambda_n} = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_p}{\text{rank}(r)}$$

где I – наперед заданное значение критерия информативности, λ_i – собственное значение матрицы r .

4. Ищутся значения главных компонент для еще не обработанных моментов времени t .

$$z_{i,t} = \sum_{k=1}^{k=n} l_{i,k} x_{t,k}$$

где z – матрица значений главных компонент, $i = \overline{1, p}$ – номер главной компоненты, l – матрица собственных векторов, где i -ая строка соответствует i -ому собственному вектору, x – матрица исходных данных. Так как исходные данные приходят к нам в «сыром» виде (нецентрированные, выраженные в разных единицах измерения), то их надо предварительно подготовить:

$$\tilde{x}_{t,k} = \frac{x_{t,k} - \bar{x}_k}{\sigma_k}$$

где \bar{x}_k – выборочное среднее, σ_k – выборочное среднее квадратическое отклонение. Таким образом, итоговая формула i -й главной компоненты имеет вид:

$$z_{i,t} = \sum_{k=1}^{k=n} l_{i,k} \tilde{x}_{t,k} = \sum_{k=1}^{k=n} l_{i,k} \frac{x_{t,k} - \bar{x}_k}{\sigma_k}$$

5. По последней формуле также вычисляются значения критических границ для главных компонент и наносятся на контрольную карту. Эти значения считаются, как главные компоненты для вектора законфигурированных в системе критических границ.

6. После этого к полученным контрольным картам главных компонент можно применять уже существующие алгоритмы анализа структур карты Шухарта с поправкой на то, что границы у наших графиков «прыгающие».

7. Окно сдвигается на заранее определенное значение сдвига и шаги 1-6 повторяются для новых данных.



На рисунке 1 приведен примерный график изменения главной компоненты z_1 и соответствующей ей критической границы в зависимости от времени.

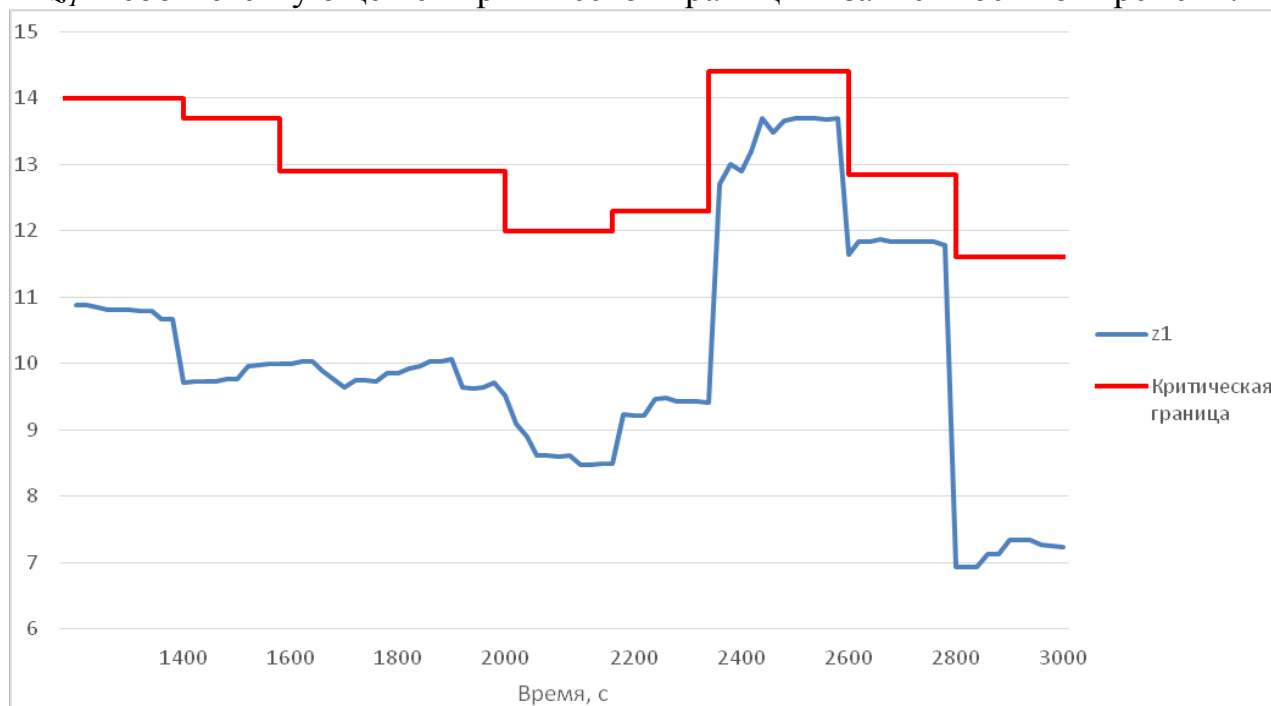


Рис. 1. Примерный вид графика главной компоненты

Заключение

В процессе работы были указаны различия между контролем качества серийной продукции и компьютерной системы, предложены пути решения этих проблем. Были получены данные системы мониторинга, проведен их предварительный анализ, первичная обработка.

Литература

1. ГОСТ Р 50779.42–99 (ИСО 8258-91) Статистические методы. Контрольные карты Шухарта. М.: Издательство стандартов, 2004. 36 с.
2. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: Классификация и снижение размерности. М.: Финансы и статистика, 1989. 607 с.

З.Ф. Камальдинова, К.В. Сырчикова

КОМПЕТЕНТНОСТНЫЙ АНАЛИЗ ВЫПУСКНЫХ КВАЛИФИКАЦИОННЫХ РАБОТ ПО НАПРАВЛЕНИЮ «ИНФОРМАЦИОННЫЕ СИСТЕМЫ»

(Самарский государственный архитектурно-строительный университет)

Компетентность специалиста с высшим образованием - это проявленные им на практике стремление и способность (готовность) реализовать свой потенциал (знания, умения, опыт, личностные качества и др.) для успешной творческой деятельности в профессиональной и социальной сфере, осознавая