

УДК 004.272

Высочиненко М. С., магистр (Тел.: +380 (67) 139 45 11. E-mail: vysochinenko_m@ukr.net)
(Киевский колледж связи)

Булаковская А. А., аспирант (Тел.: +380 (67) 985 21 90. E-mail: gyhz@yandex.ua)
(Национальный авиационный университет, г. Киев)

ОЦЕНИВАНИЕ ПАРАМЕТРОВ И ОПТИМИЗАЦИЯ ИНФОРМАЦИОННЫХ СИСТЕМ НА ОСНОВЕ МАТРИЦЫ КЛЮЧЕВЫХ ПОКАЗАТЕЛЕЙ ЭФФЕКТИВНОСТИ

Височиненко М. С., Булаковська А. О. Оцінювання параметрів і оптимізація інформаційних систем на основі матриці ключових показників ефективності. Виконаний аналіз основних особливостей матриці ключових показників ефективності для оцінювання параметрів і стану, оптимізації, управління ефективністю і якістю сервісу інформаційної системи мережі. На основі методу множинного кореляційного і регресійного аналізу розроблена методика поточного оцінювання параметрів і оптимізації системи. Приведені методика і результати розрахунків типової кореляційної матриці ключових показників ефективності і проаналізовані статистичні зв'язки між основними параметрами, від яких залежать ефективність системи і якість сервісу.

Ключові слова: інформаційна система, ключові показники ефективності, управління ефективністю, кореляційний і регресійний аналіз, якість сервісу, кореляційна матриця

Высочиненко М. С., Булаковская А. А. Оценка параметров и оптимизация информационных систем на основе матрицы ключевых показателей эффективности. Выполнен анализ основных особенностей матрицы ключевых показателей эффективности для оценивания параметров и состояния, оптимизации, управления эффективностью и качеством сервиса информационной системы сети. На основе метода множественного корреляционного и регрессионного анализа разработана методика текущего оценивания параметров и оптимизации системы. Приведены методика и результаты расчетов типовой корреляционной матрицы ключевых показателей эффективности и проанализированы статистические связи между основными параметрами, от которых зависят эффективность системы и качество сервиса.

Ключевые слова: информационная система, ключевые показатели эффективности, управление эффективностью, корреляционный и регрессионный анализ, качество сервиса, корреляционная матрица

I. Введение

Ключевые показатели эффективности (англ. *Key Performance Indicators, KPI*) – система оценок, которая широко применяется в управлении организациями и предприятиями. Основным результатом применения *KPI* является получение сравнительных оценок степени достижения стратегических и тактических (операционных) целей. Использование ключевых показателей эффективности даёт возможность оценить состояние и помочь в оценке реализации методов управления. Применение оценок по *KPI* в сложных технических системах распространено далеко не так широко. Кроме того, недостаточно развиты математические модели и методы оценки ключевых показателей эффективности, прогноза параметров и состояния, а также оптимизации информационных систем на основе получаемых оценок *KPI*. В данной работе сделана попытка восполнить этот пробел.

Для термина *KPI* зачастую используется русский перевод «ключевые показатели эффективности» (КПЭ), однако это не совсем верно [1]. Дело в том, что слово *performance* имеет множество трактовок. Правильную формулировку можно найти в стандарте ISO 9000:2008. Он разделяет слово *performance* на два термина: результативность и эффективность. По стандарту, результативность – это степень достижения запланированных результатов (способность системы ориентироваться на результат), а эффективность –

соотношение между достигнутыми результатами и затраченными ресурсами (способность системы к реализации своих целей и планов с заданным качественным уровнем, выраженным определёнными требованиями – временем, затратами, степенью достижения цели). Слово *performance* объединяет в себе и результативность, и эффективность. Таким образом, правильным переводом термина *KPI* был бы «ключевой показатель результатов функционирования», так как результат функционирования содержит в себе и степень достижения, и затраты на получение результата [1]. Однако в литературе, в том числе и технической, включая статьи и монографии по связи и информационным технологиям [2-4], под термином *KPI* понимают ключевые показатели соотношений между достигнутыми результатами и затраченными ресурсами. Поэтому, на наш взгляд, целесообразно придерживаться именно такой трактовки термина *KPI*.

В данной работе рассматривается система ключевых показателей эффективности для большой информационной системы (ИС). При оптимизации характеристик ИС необходимо учитывать параметры, от которых зависит качество сервиса, и взаимосвязь между этими параметрами. Поскольку параметры отдельных сетевых узлов и элементов, состояние системы в целом изменяются в процессе функционирования сети случайным образом, необходимо применять методы математической статистики. Оценка множества параметров и состояния системы в процессе ее функционирования, по существу, представляет собой задачу многомерного статистического (корреляционно-регрессионного) анализа. Для правильного выбора приемлемой размерности задачи с получением достаточных статистик о техническом состоянии системы можно предложить следующий подход.

Определяются цели, основанные на экономических ключевых показателях. На их основе ставится задача глобальной оптимизации и частные технические задачи оценки поведения ИС и результатов ее функционирования. Например, для инфокоммуникационной сети это производительность и уровень ошибок и потерь при речевых вызовах и пересылке пакетов. Эти услуги будут оптимизированы, и обнаруженные ошибки будут исключены (или хотя бы сведены к допустимому уровню).

II. Постановка задачи

При оптимизации параметров и структуры ИС в состав целевой функции входит большое количество основных и дополнительных параметров, от которых зависит качество сервиса *QoS*. Рассмотрим в качестве примера постановку задачи оптимизации коммуникационной сети. Последовательность этапов оптимизации можно представить в виде алгоритма, изображенного на Рис. 1.

Для решения задач текущего управления необходим системный подход, поскольку критерии оптимизации ключевых параметров функционирования ИС и текущего управления являются неоднозначными и, как правило, противоречивыми. Учет этих противоречий и поиск компромиссных решений возможен при использовании статистических методов, согласования достоверности и детальности исходных данных с физическим смыслом решаемых задач.

Процессы изменения *KPI*, с одной стороны, являются нестационарными на достаточно длительных интервалах наблюдения, но с другой – тенденции их изменений весьма схожи. Поэтому представляет интерес исследование характеристик их статистической связи во времени и ковариации между процессами в разных объектах в одни и те же моменты времени. Этот интерес имеет не только теоретический, но и практический характер. В

качестве основных характеристик статистической связи обычно используют матрицы коэффициентов множественной корреляции и системы уравнений множественной линейной или полиномиальной регрессии [5, 6]. Кроме того, для автоматизации измерений и расчетов необходимо выбрать метод аппроксимации кривых повторяемости изменений КПЭ. Наиболее гибкими и точными методами являются аппроксимация полиномами по минимуму среднего квадрата ошибки [7] или аппроксимации Паде [8].

Рассмотрим процесс прогноза параметров ИС как задачу предсказания k -й переменной Y_k , $k = \overline{1, N}$ по M переменным X_m , $m = 1, 2, \dots, M$; $m \neq k$. В общем случае $M \neq N$. При $m = 1$ имеем уравнение линейной или полиномиальной регрессии независимой переменной X_m на зависимую переменную Y_k , при $m > 1$ имеем систему m уравнений множественной регрессии переменных X_1, X_2, \dots, X_m на Y_k . (Имеется в виду функциональная, а не статистическая зависимость.) В рассматриваемой задаче независимые переменные X_1, X_2, \dots, X_m – это случайные величины, которые не обязательно являются статистически независимыми.



Рис. 1. Алгоритм оптимизации абстрактной ИС по ключевым показателям $KPIs$

Переменную Y_k аппроксимируем функцией регрессии $\psi(\cdot)$, содержащей оценки КПЭ и неизвестные коэффициенты $\{a_0, a_1, \dots, a_m\}$. Уравнение модели линейной регрессии

независимых переменных X_1, X_2, \dots, X_m на зависимую переменную Y_k запишем в следующем виде:

$$Y_k = a_{0k} + a_{1k}X_1 + \dots + a_{mk}X_m + \varepsilon, \quad (1)$$

где ε – ошибка аппроксимации.

Пусть $X_{1j} = X_1^j$. Тогда можно записать уравнение полиномиальной регрессии в виде

$$Y_k = a_{0k} + a_{1k}X_1 + a_{2k}X_1^2 + \dots + a_{mk}X_1^m + \varepsilon. \quad (2)$$

Параметры модели регрессии оцениваются по выборке объема N , взятой из некоторой генеральной совокупности. Теоретически генеральная совокупность имеет бесконечный объем или представляет собой весь набор данных, который существует в принципе.

Выборка формируется следующим образом. По результатам теста функционирования сети фиксируем первую выборку независимых переменных $X_{11}, X_{12}, \dots, X_{1m}$ и рассчитываем зависимую переменную Y_1 . Затем фиксируем вторую выборку независимых переменных $X_{21}, X_{22}, \dots, X_{2m}$ и рассчитываем зависимую переменную Y_2 . Продолжаем процедуру до получения N переменных Y_k , $k = \overline{1, N}$. Получаем выборку из N наблюдений

$$\{Y_1 : X_{11}, X_{12}, \dots, X_{1m}\}, \{Y_2 : X_{21}, X_{22}, \dots, X_{2m}\}, \dots, \{Y_N : X_{N1}, X_{N2}, \dots, X_{Nm}\}.$$

Система уравнений множественной линейной регрессии принимает вид:

$$\left. \begin{aligned} Y_1 &= a_{01} + a_{11}X_{11} + \dots + a_{m1}X_{1m} + \varepsilon_1 \\ Y_2 &= a_{02} + a_{12}X_{21} + \dots + a_{m2}X_{2m} + \varepsilon_2 \\ &\dots \\ Y_k &= a_{0k} + a_{1k}X_{k1} + \dots + a_{mk}X_{km} + \varepsilon_k \\ &\dots \\ Y_N &= a_{0N} + a_{1N}X_{N1} + \dots + a_{mN}X_{Nm} + \varepsilon_N \end{aligned} \right\}, \quad (3)$$

где $\{a_{0k}, a_{1k}, \dots, a_{mk}\}$, $k = \overline{1, N}$ – неизвестные коэффициенты, а $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k, \dots, \varepsilon_N\}$ – случайные ошибки, которые логично считать нормальными одинаково распределенными с параметрами $\{0, \sigma_\varepsilon^2\}$.

Для получения оценок по методу наименьших квадратов необходимо минимизировать сумму S_k квадратов отклонений в каждой точке. Наилучшее приближение соответствует минимальной величине выражения

$$S_k = \sum_{k=1}^N (Y_k - a_{0k} - a_{1k}X_{k1} - \dots - a_{mk}X_{km})^2. \quad (4)$$

Величина S_k является мерой ошибки, связанной с привязкой имеющихся данных к выбранной модели регрессии. Минимум S_k достигается дифференцированием последнего выражения по коэффициентам $\{a_{0k}, a_{1k}, \dots, a_{mk}\}$, $k = \overline{1, N}$, приравниванием соответствующих производных нулю и решением системы уравнений относительно $\{a_{0k}, a_{1k}, \dots, a_{mk}\}$. Получаем систему уравнений для оценки частных коэффициентов регрессии:

$$\left. \begin{aligned} \hat{Y}_1 &= \alpha_{01} + \alpha_{11}X_{11} + \dots + \alpha_{m1}X_{1m} \\ \hat{Y}_2 &= \alpha_{02} + \alpha_{12}X_{21} + \dots + \alpha_{m2}X_{2m} \\ &\dots \\ \hat{Y}_k &= \alpha_{0k} + \alpha_{1k}X_{k1} + \dots + \alpha_{mk}X_{km} \\ &\dots \\ \hat{Y}_N &= \alpha_{0N} + \alpha_{1N}X_{N1} + \dots + \alpha_{mN}X_{Nm} \end{aligned} \right\}. \quad (5)$$

Здесь $\alpha_{0k}, \alpha_{1k}, \dots, \alpha_{mk}$ – оценки для $\{a_{0k}, a_{1k}, \dots, a_{mk}\}$. Оценки являются несмещенными и эффективными, т.е. имеют минимальную дисперсию для выборки X_1, X_2, \dots, X_m среди всех линейных оценок для предсказания переменных $Y_k, k = \overline{1, N}$.

III. Результаты численного анализа

В рассматриваемом случае проще вместо оптимальной оценки как выходной величины, определяемой матричным уравнением, искать оптимальную оценку как решение двойственного ему разностного уравнения. Коэффициенты разностного уравнения определяются статистикой наблюдений и помех и в общем случае являются переменными величинами, зависящими от времени. Преимуществом такого подхода является то, что если даже не удастся получить аналитическое решение разностного уравнения, то всегда можно получить его численное решение на вычислительной машине. Более того, решение можно получать в реальном времени с учетом вновь получаемой информации об изменениях параметров наблюдений и помех.

Как отмечено выше, в настоящее время широко используется понятие ключевых параметров эффективности сложных систем, в частности, компьютерных и телекоммуникационных сетей [2]. Ключевыми параметрами являются задержка передачи, пропускная способность, потери пакетов и уровень безопасности. Эти параметры оказывают наибольшее влияние на результирующее качество сервиса.

В качестве оптимизируемых параметров задачи выбраны следующие:

- задержка передачи τ ;
- пропускная способность C_p ;
- потери пакетов при передаче данных L_p ;
- уровень безопасности и защиты данных при передаче по сети D_{sp} ;
- качество *Web*-сервиса;
- качество передачи аудио (звуковые файлы, обычная и *IP*-телефония);
- скорость и надежность обмена файлами по протоколу *FTP*;
- скорость и надежность работы электронной почты (*E-mail*);
- качество передачи видео.

Рассмотрена гипотетическая сеть *WiMax*, данные для расчета параметров которой взяты из работы [9]. Для расчетов использовалась ФОРТРАН-программа множественного корреляционного анализа, приведенная в [10] и модифицированная для рассматриваемой задачи.

В Табл. 1 приведенные частные коэффициенты корреляции оптимизируемых параметров, по которым в дальнейшем с использованием уравнений (1-5) можно рассчитывать частные коэффициенты регрессии.

Табл. 1. Коэффициенты взаимной корреляции оптимизируемых параметров

Параметр	Коэффициенты корреляции									
τ	1,0									
C_p	0,98	1,0								
L_p	0,69	0,68	1,0							
D_{sp}	0,89	0,86	0,69	1,0						
<i>Web</i>	0,75	0,76	0,36	0,77	1,0					
Аудио	0,85	0,64	0,50	0,56	0,30	1,0				
<i>FTP</i>	0,27	0,75	0,63	0,61	0,57	0,44	1,0			
<i>E-mail</i>	0,17	0,22	0,34	0,78	0,30	0,36	0,16	1,0		
Видео	0,87	0,89	0,84	0,82	0,53	0,67	0,79	0,30	1,0	
	τ	C_p	L_p	D_{sp}	<i>Web</i>	Аудио	<i>FTP</i>	<i>E-mail</i>	Видео	

Между основными ключевыми параметрами обнаруживается сильная корреляция. Это объясняется тем, что они оказывают значительное влияние на требования к качеству сервиса. Исключение составляет электронная почта, поскольку, в отличие от потокового аудио, видео, *Web*-сервиса и передачи файлов по протоколу *FTP*, для нее не критичны ни полоса пропускания канала, ни задержка доставки. Однако необходимо отметить, что параметр D_{sp} – уровень безопасности и защиты данных является критичным практически для всех представленных приложений, поскольку даже для таких видов эластичного трафика, как электронная почта, защита данных является неотъемлемым требованием обеспечения качества сервиса *QoS*.

Результаты корреляционного анализа служат также ключевым индикатором мониторинга и регулирования потоковых данных и *Web*-сервиса. Это необходимо для обеспечения безопасной передачи информации по сети, прогнозирования и предотвращения перегрузок контролируемого сетевого фрагмента. Таким образом, текущий мониторинг и управление уровнем безопасности в сети, которые являются неотъемлемой частью задачи общего управления качеством сервиса, можно успешно осуществлять статистическими методами, в частности, методом корреляционно-регрессионного анализа.

Кроме того, необходимо отметить, что полностью скомпилированная программа расчетов занимает в памяти вычислительного устройства от 80 до 500 килобайт в зависимости от масштаба сети и объема обрабатываемой выборки. Поскольку в настоящее время практически любой сетевой узел, по существу, представляет собой специализированный вычислитель или даже многопроцессорную систему, задача аппаратной реализации предложенного метода может решаться сравнительно просто.

IV. Выводы

В работе проведен анализ системы ключевых параметров эффективности и особенностей их применения для управления качеством сервиса информационной системы. Показано, что при использовании статистического подхода можно выделить зависимости между ключевыми параметрами сети, что дает возможность построения системы управления качеством сервиса.

При использовании ключевых параметров эффективности сложной системы с задержками сигнальной и управляющей информации можно обеспечить предсказание ее состояния и решать задачи управления качеством сервиса в реальном времени. В дальнейшем планируется исследовать задачи приоритизации частных показателей эффективности, например, методом анализа иерархий, для оптимизации информационных систем по многим, в том числе противоречивым критериям.

Литература

1. Parmenter D. Key Performance Indicators (KPI): Developing, Implementing, and Using Winning KPIs, 2nd Edition / D/ Parmenter. – John Wiley & Sons, 2010. – 320 p.
2. Kreher R. UMTS Performance Measurement: A Practical Guide to KPIs for the UTRAN Environment / R. Kreher. – John Wiley & Sons, Ltd, 2006. – 227 p.
3. Shahin A. Prioritization of key performance indicators. An integration of analytical hierarchy process and goal setting / A. Shahin, M. A. Mahbod // International Journal of Productivity and Performance Management. – 2007. – Vol. 56, No. 3. – P. 226-240.
4. Ye Ouyang, M. Hosein Fallah. An Analysis of Traffic and Throughput for UMTS Packet Core Networks / Ye Ouyang, M. Hosein Fallah // International Journal of Interdisciplinary Telecommunications and Networking (IJITN). – 2010. – Volume 2, Issue 2. – P. 1-26.
5. Evans M. Statistical distributions / M. Evans, N. Hastings, B. Peacock. – 2nd ed. – John Wiley & Sons, Inc, 1993. – 186 pp.
6. Афифи А. Статистический анализ: Подход с использованием ЭВМ ; пер. с англ. / А. Афифи, С. Эйзен. – Москва : Мир, 1982. – 488 с.
7. Nash J. C. Compact numerical methods for computers: linear algebra and function minimisation. – 2nd ed. / J. C. Nash. – Adam Hilger, Bristol and New York, 1990. – 288 p.
8. Бейкер Дж. Аппроксимации Паде / Дж. Бейкер, П. Грейвс-Моррис. – Москва : Мир, 1988. – 502 с.
9. Вишнеvский В.М. Широкополосные беспроводные сети передачи информации / В. М. Вишнеvский, А. И. Ляхов, С. Л. Портной, И. В. Шахнович. – Москва: Техносфера. – 2005. – 592 с.
10. Мак-Кракен Д. Численные методы и программирование на Фортране ; пер. с англ. / Д. Мак-Кракен, У. Дорн. – Москва : Мир, 1977. – 584 с.

Дата надходження в редакцію: 22.01.2015 р. Рецензент: д.т.н., проф. Н. А. Віноградов