

Анализ эффективности информационных систем

Михайлова Т.В., Коваленко С.В.
Донецкий национальный технический университет
tanya@r5.dgtu.donetsk.ua

Abstract

Mihaylova T., Kovalenko S. Analysis of efficiency of the informative systems. Using of closed network models for studying of the information systems.

Введение

Проблемам оценки эффективности информационных систем (ИС) посвящено большое количество работ [1-3], анализ которых позволяет выделить этапы внедрения и использования.

В источниках предлагаются разнообразные модели исследования эффективности ИС [1,3], однако большинство из них описаны в общем виде и не раскрыты на примерах в полном объеме, либо исследованы отдельные стадии их жизненного цикла.

Актуальным для корректной работы ИС остается соответствие технических средств программному обеспечению. В предлагаемой работе эта проблема решается с использованием результатов приближенных сетевых моделей, так как они менее ресурсоемки и обладают большей погрешностью по сравнению с дискретными Марковскими моделями [4]. Однако при необходимости получения более точных оценок дискретные модели можно считать на многопроцессорной технике, потому что они легко распараллеливаются [5] на различные многопроцессорные структуры.

Методика исследования эффективности ИС включает определение исследуемых параметров и построение модели.

Характеризация и разбиение рабочей нагрузки

Определяются подсистемы, подлежащие анализу, и контрольные точки, в которых будут проводиться измерения, в том числе точка зрения, с позиций которой необходимо характеризовать нагрузку.

Для построения модели нагрузки ИС необходимо: идентификация базовых компонент; выбор параметров, характеризующих нагрузку (интенсивность нагрузки, потребности в обслуживании, задаваемые с помощью вектора $(T_{i1}, T_{i2}, \dots, T_{ik})$, где K – количество рассматриваемых ресурсов); сбор данных, включающий определение временных окон и

измерение работы системы на протяжении выбранных временных окон.

Далее разбивается нагрузка на классы, которая позволяет улучшить представительность характеристики и увеличить возможности модели, связанные с прогнозом.

Используемые атрибуты для разбиения нагрузки на классы:

- по использованию ресурсов;
- по приложениям;
- по типу объектов;
- по функциональному назначению;
- по организационным единицам.

Параметры для каждого класса нагрузки вычисляются методом кластеризации.

Построение закрытой модели с запросами нескольких классов

Для обозначения параметров, относящихся к очереди i и классу r , используются индексы i и r , соответственно. В случае СМО с запросами нескольких классов будем считать, что в каждом классе находится постоянное количество запросов, N_r , которое для удобства будет представляться в виде вектора $\bar{N} (N_1, \dots, N_r, \dots, N_R)$. Для обозначения вектора, все компоненты которого равны нулю, за исключением r -го, равного 1, будем использовать обозначение \bar{l}_r .

В формулах (1 – 4) приведены уравнения анализа средних значений для закрытых СМО с запросами нескольких классов.

Время пребывания в очереди i для запроса класса r :

$$W'_{i,r}(\bar{N}) = \begin{cases} T_{i,r}, \text{ ресурс_с_задержкой} \\ T_{i,r}[1 + n_i(\bar{N} - \bar{l}_r)], \text{ ресурс_с_очередью} \end{cases} \quad (1)$$

Производительность для запросов класса r :

$$X_{0,r}(\bar{N}) = \frac{N_r}{\sum_{i=1}^K W'_{i,r}(\bar{N})} \quad (2)$$

Длина очереди i для запросов класса r :

$$n_{i,r}(\bar{N}) = X_{0,r}(\bar{N}) * W'_{i,r}(\bar{N}) \quad (3)$$

Длина очереди i :

$$n_{i,r}(\bar{N}) = \sum_{r=1}^R n_{i,r}(\bar{N}) \quad (4)$$

В качестве начальных значений удобно использовать длины, получаемые после равномерного распределения всех запросов класса по всем очередям. Таким образом, если запросы класса r могут находиться только в K_r очередях из K , то начальным значением для $n_{i,r}$ будет N_r / K_r . Для получения более точных оценок длин очередей можно воспользоваться уравнениями для времени пребывания запросов в очереди, уравнением для производительности и уравнениями для длин очередей. Далее полученные оценки применяются для уточнения длин очередей. Итерации выполняются до тех пор, пока максимальная относительная ошибка в длинах очередей между двумя последовательными итерациями окажется меньше заданной пользователем максимально допустимой ошибки.

Пример исследования информационной системы предприятия

На предприятии внедрена информационная система, состоящая из двух систем: 1С-бухгалтерии и учетной системы «Монолит». Учетная система, разработанная пользователем, установлена непосредственно на сервере. Сотрудники фирмы, в количестве 25-ти человек, обращаются к учетной системе на сервере, запрашивая различные документы для работы с ними и для сверки с документами в программе 1С-бухгалтерия. Схема сети данного предприятия относится к закрытым сетям и наиболее подходящей моделью для этой сети будет закрытая модель производительности уровня компонентов [1].

Компоненты сети: узлы сети с сетевой картой Compeх RE100ATX/WOL 10/100Mb with Wake-On-Lan; FSD-803 Planet 12-Port 10/100Mbps Desktop Fast Ethernet Switch (металлический корпус); коаксиальный кабель RG-58 со скоростью передачи 10 Мб/с, 100 м; Сетевой принтер SAMSUNG ML-2151N; А4; 1200dpi; 20ppm; 16Mb; LPT&USB 2,0; Ethernet; Duplex, сервер данных и др.

Для построения закрытой модели производительности уровня компонентов наиболее подходит метод разбиения нагрузки по использованию ресурсов. Атрибутами, по которым будет производиться разбиение нагрузки, служат количество запросов к документу и размер документа (табл.1).

После разбиения нагрузки на некоторое число классов необходимо рассчитать значения параметров, представляющих класс компонентов.

Таблица 1 Выборка из нагрузки

Номер док-та	Вид документа	Размер (Кб)	Число запр. к док-ту
1	Приходные накладные	259	75
2	Расходные накладные	236	70
3	Накл-е на возврат тары	135	46
4	Наклюдные на ввод клиента	112	48
5	Список клиентов	54	21
6	Индивид-еены клиентов	43	18
7	Наценки на клиентов	47	31
8	Ассортимент товара	62	28
9	Книга приобретения	160	64
10	Книга продаж	145	55
11	Главная книга	264	54
12	Книга доходов и расходов	275	59

Для расчета значений параметров широко применяется метод кластеризации [4].

Шаги, которые потребуются выполнить для проведения кластерного анализа данной нагрузки: статистический анализ данных, который включает

- формирование выборки;
- преобразование параметров;
- исключение выбросов.

Число компонентов, из которых состоит нагрузка, зависит от длины сеанса измерения. Поэтому выборка сформирована из измеренных значений, иначе время выполнения кластерного анализа может оказаться недопустимо большим.

Далее анализируется распределение каждого параметра кортежа, чтобы исключить выбросы, которые могут исказить среднее значение. Строится гистограмма распределения частот по данным в табл.1 для выделения малочисленных больших значений, соответствующих малой доле компонентов (1-2%) (рис.1-2).

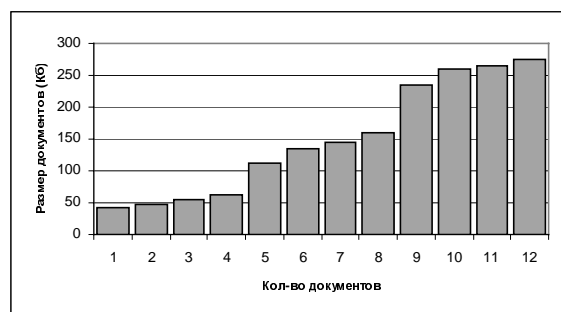


Рисунок 1- Гистограмма распределения частот параметра «Размер документа»

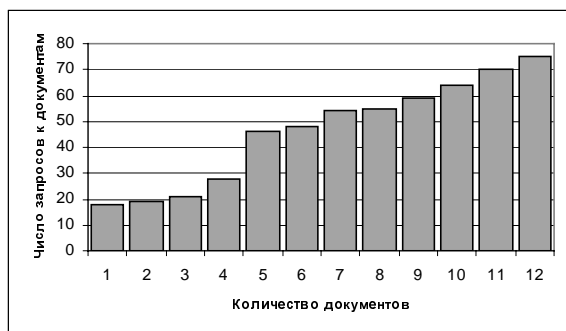


Рисунок 2- Гистограмма распределения частот параметра «Число запросов к документу»

Анализируя рис.1-2 можно сделать вывод, что компоненты различаются друг от друга незначительно, следовательно, аномальных значений или выбросов здесь нет.

Далее выделяются естественные группы компонентов, идентичных по требованиям к ресурсам с использованием алгоритма кластеризации, конкретная реализация которого зависит от многих факторов, например, количества данных, опыта работы с этими моделями, цели разбивки и т.д..

Так как он трудоемкий, его можно заменить в нашем случае геометрической интерпретацией (рис.3).

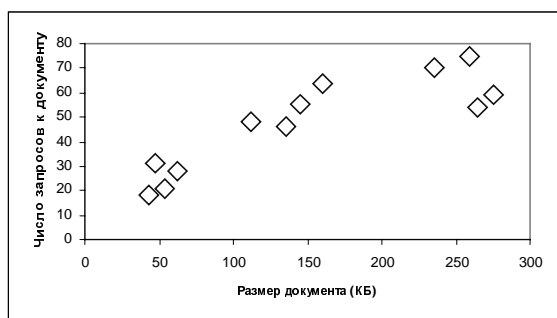


Рисунок 3- Запросы документов различного размера

Реальная нагрузка, представленная в табл.1, теперь разделена на три класса схожих запросов. Каждый класс представлен значениями параметров, равными среднему арифметическому значений параметров компонентов класса. В табл.2 показаны результаты процесса кластеризации. Она содержит описание центроидов, представляющих каждый кластер, и количество компонентов класса.

В табл. 2 приведены значения параметров после кластеризации. Каждый кластер соответствует одному классу нагрузки. Например, класс 2 соответствует среднему размеру документа 136,9 Кб и среднему количеству запросов к документу 52,8. В этот класс входит 4 документа.

Таблица 2 Результаты процесса кластеризации

Тип	Обозначение соотв-го класса	Размер документа (Кб)	Число запросов к док-ту	Число компонент
Класс 1	C1,2,11,12	258,1	64,0	4
Класс 2	C3,4,9,10	136,9	52,8	4
Класс 3	C5,6,7,8	51,0	23,9	4

В табл.3 – 4 приведены времена передачи запросов сетью и диском, соответственно, посчитанные по формулам [1,2].

Таблица 3 Результаты расчета времени передачи запросов по сети

	Размер файла (байт)	K1	K2	D1
Класс 1	258099	177	10253	0,21
Класс 2	136855	94	5437	0,11
Класс 3	51002	35	2026	0,04

Таблица 4 Результаты расчета времени обработки запросов диском

	Размер файла (байт)	Скорость обработки запросов диском (Мб/с)	D2
Класс 1	258099	10	0,026
Класс 2	136855	10	0,014
Класс 3	51002	10	0,005

Результаты модели производительности с запросами нескольких классов с помощью (1-4) (реализованной в системе Mathematica) представим в таблице 5. Аналогично рассчитывается производительность для запросов второго третьего классов.

Из табл.5 видно, что после седьмой итерации максимальная относительная ошибка для значений длин очередей не превышает 0,05%.

В табл. 6 содержатся времена пребывания запросов в очереди и времена отклика для одного класса. Аналогично времена пребывания запросов в очереди и времена отклика можно получить для остальных классов. Ввиду большого количества расчетов, таблицы для второго и третьего классов не приводятся.

Таблица 5 Результаты расчета производительности для запросов 1-го класса

Итерация	Запросы 1-го класса		
	Длина очереди		Производитель. (запр/с)
	Сеть	Диск	
1	2,0000	2,0000	2,61
2	3,5593	0,4407	1,72
3	3,9002	0,0998	1,60
4	3,9467	0,0533	1,59
5	3,9526	0,0474	1,59
6	3,9533	0,0467	1,59
7	3,9534	0,0466	1,59

Таблица 6 Результаты расчета средних времен и времени отклика

Итерация	Запросы 1-го класса		
	Время пребывания в очереди		Время отклика (с)
	Сеть	Диск	
1	1,365	0,169	1,534
2	2,262	0,058	2,32
3	2,462	0,033	2,495
4	2,489	0,03	2,519
5	2,492	0,029	2,521
6	2,493	0,029	2,522
7	2,493	0,029	2,522

Анализируя данные, можно сделать вывод, что в данной модели узким местом является сеть. Изменив пропускную способность сети с 10Мб/с на 100Мб/с, модель заново рассчитывается. Результаты моделирования с модифицированными данными приведены в табл.7.

Таблица 7 Результаты расчета средних времен и времени отклика

Итерация	Запросы 1-го класса		
	Время пребывания в очереди		Время отклика (с)
	Сеть	Диск	
1	0,137	0,169	0,306
2	0,123	0,185	0,308
		...	
24	0,073	0,247	0,32
25	0,073	0,247	0,32

Таким образом, информационная система более эффективна при преобразовании сети. При этом производительность сети увеличивается в 8 раз при уменьшении времени отклика в 8 раз.

Выводы

На основе предложенных моделей проанализирована работа информационной системы на конкретном предприятии, выявлены недостатки ее работы, предложены способы улучшения производительности ИС.

Результаты работы могут применяться на предприятиях, которые имеют такие информационные системы, и которые стремятся улучшить использование рабочего времени.

Также такие модели могут использоваться при проектировании ИС, при расчетах экономической эффективности использования IT-технологий.

Предложенные модели могут быть базовыми при разработке планов внедрения информационных систем на предприятиях, в консультационной деятельности по использованию информационных технологий.

В перспективе к этой модели можно добавить модель внедрения ИС и прогноза, которые являются составной частью методики анализа эффективности ИС.

Современные информационные технологии осуществляют не только автоматизацию уже существующих процессов на предприятии, но и становятся своеобразным носителем и катализатором распространения передовых технологий менеджмента. Новые информационные системы воплощают в себе передовой опыт управления.

Литература

1. Менаске Дэниел, Алмейда Виргилио. Производительность Web-служб. Анализ, оценка и планирование. СПб: ООО «ДиаСофтЮП», 2003.
2. Спортак М., Франк Ч., Паппас Ч. и др. Высокопроизводительные сети. Энциклопедия пользователя.-К.: "ДиаСофт", 1998.-432с.
3. Скрипник К. Г. Экономическая эффективность информационных систем. – М.: ДМК Пресс, 2002. – 256с.: ил.
4. Михайлова Т.В. Оценка точности непрерывной и дискретной моделей Маркова. / Научные труды Донецкого государственного технического университета. Серия «Информатика, кибернетика и вычислительная техника»(ИКВТ-2005), выпуск 93.- Донецк: ДонГТУ.- 2005.
5. Фельдман Л.П., Михайлова Т.В. Параллельный алгоритм построения дискретной модели Маркова. Искусственный интеллект. Интеллектуальные и многопроцессорные системы. Материалы Международной научно-технической конференции, 25-30 сентября 2006г., Таганрог-Донецк-Минск, 2006.