

УДК 519.6

А.Н. Пасхин

ИСПОЛЬЗОВАНИЕ КРИТЕРИЯ СПРИМЕНА ДЛЯ ВЫБОРА ПАРАМЕТРОВ УПРАВЛЕНИЯ ТЕЛЕКОММУНИКАЦИОННОЙ СЕТЬЮ

Аннотация. В данной работе выделены основные характеристики телекоммуникационных сетей и предложена возможность использования методов непараметрической статистики для выделения среди возможных способов подстройки параметров, только тех, которые оказывают влияние на сеть и могут повысить эффективность ее работы.

Одной из наиболее важных задач, которые стоят перед проектировщиками телекоммуникационных систем, систем поддержки принятия решений является определение и корректировка ключевых параметров систем. Контроль показателей этих параметров является необходимым процессом, гарантирующим заблаговременное принятие решений о расширении, модернизации сетевого оборудования; перенастройка параметров системы позволяет разрешить проблемы связанные с критическими, уязвимыми параметрами и гарантировать целостность данных в системе.

Для определения и отбора настроек системы, необходимо сначала определиться с теми параметрами, управление которыми будет осуществляться, для чего необходимо исследовать соответствующую модель. В случае телекоммуникационной сети подойдет подход анализа очередей.

Для систем с очередями необходимо, прежде всего, оценить их загруженность. Самой простой мерой загруженности может выступать нагрузка ρ , выступающая как отношение среднего времени обслуживания заявки к среднему времени между моментами прихода заявок. Величины в числителе и знаменателе равны соответственно $1/\lambda$ и $1/\mu$. Тогда, $\rho = \lambda / \mu$.

Если нагрузка больше единицы, то это означает что заявки приходят быстрее, чем обрабатываются. В системе массового обслуживания (СМО) с l параллельными обслуживающими

© Пасхин А.Н., 2010

устройствами, на каждый из них приходится в среднем около λ / l заявок в единицу времени. Поэтому, нагрузка в этой СМО может быть поднята до величины 1.

С нагрузкой тесно связан такой показатель качества как коэффициент использования или коэффициент загрузки обслуживающего устройства, который обозначим u . Это показатель частицы часового измерения, когда обслуживающее устройство занято. Рассмотрим достаточно длительный промежуток времени T . В СМО с l обслуживающими устройствами на каждый из этих устройств в среднем за время T приходится по $\lambda T / l$ заявок, если допустить, что поток заявок равномерно распределяется по l устройств. Поскольку каждая заявка требует в среднем время обслуживания $1/\mu$, то общее среднее время занятости обслуживающего устройства составит $\lambda T / \mu l$. Поделив эту величину на T , получим $u = \lambda / l\mu$. При том, что обслуживающее устройство не может быть занято более чем на 100% времени, коэффициент использования не может быть выше единицы. Значит получаем выражение для СМО с обслуживающими устройствами:

$$u = \min(\lambda / l\mu, 1).$$

Одним из основных показателей качества моделей телекоммуникационных систем является пропускная способность. Эта величина определяется как среднее число заявок, обслуживающихся в единицу времени. В СМО в каждую единицу времени в среднем завершается обслуживание $l\mu$ заявок, из чего следует, что пропускная способность равняется:

$$l \mu = \min(\lambda, l\mu).$$

Очевидно, что пропускная способность будет совпадать с интенсивностью вхождения заявок λ до тех пор, пока λ будет меньшей, чем максимальная интенсивность обслуживания $l\mu$, выше которой пропускная способность не поднимается.

Касаясь заявки, основным показателем качества ее обслуживания, очевидно, является время, в течении которого она находится в ожидании обработки. Определим, что время ожидания W_j заявки j равняется отрезку времени от момента ее входа в систему до момента начала ее обслуживания, а время ответа T_j равняется отрезку времени от момента входа заявки в систему до момента завершения ее обслуживания. Как следствие имеет такое простое

отношение:

Час ответа (T) = время ожидания (W) + время обслуживания (S).

Чтобы оценить загруженность системы, введем еще такое понятие, как “длина очереди”. Пусть случайный процесс $Q(t)$ - это число заявок, которые ожидают обслуживания на момент времени t . Аналогично, определим $N(t)$ как число заявок, которые находятся в системе в очереди или на обслуживании. Процесс $N(t)$ называют длиной очереди. В системах массового обслуживания с l обслуживающими устройствами $Q(t)$ и $N(t)$ связаны соотношением

$$Q(t) = \max(0, N(t) - l).$$

Таким образом, выделены основные параметры телекоммуникационных сетей - загруженность, час ответа, пропускная способность, коэффициент загрузки.

Мониторинг и управление значениями этих параметров является ключевым в процессе управления сетью, настройки параметров сети с целью оптимизации процессов в ней протекающих.

Задача выбора эталонных значений этих параметров целиком лежит на проектировщике, так как именно он является экспертом в области.

Основной проблемой в управлении параметрами телекоммуникационной сети является изменение тех настроек, которые действительно улучшают эффективность работы сети. Например, для улучшения пропускной способности, можно расширить канал, изменить топологию сети, закупить новые маршрутизаторы и так далее. На самом деле, при изменении некоторых конфигурационных настроек целевые, “эталонные” параметры сети не изменятся.

Решение задачи корреляции двух выборок - эталонной и реальных значений может быть получено с помощью непараметрических методов оценки различных двух групп наблюдений. Непараметрическая статистика рассматривает те ситуации, в которых о функциональном виде распределения ничего не известно. Единственной априорно доступной информацией является характер случайной величины - дискретна ли она или непрерывна, каков тип различия между величинами. Сами распределения могут быть любыми, интерес представляет лишь ответ, сдвинуты ли они относительно друг друга; различаются ли они

масштабами; наличием или отсутствием симметрии или зависимости; наконец, есть ли между ними вообще какое бы то ни было отличие, или они могут считаться одинаковыми.

Среди существующих критериев наиболее точными являются ранговые. В отличие от частотных, ранговые критерии оперируют небольшими выборками (до 20-30 вариантов в каждой из сравниваемых групп). Процедура использования ранговых критериев полагает составление упорядоченного (ранжированного) ряда всех анализируемых вариантов. Для применения этих методов, величины ранжируют по восходящей градации в едином ряде. Обычно этот ряд пишут в виде двух параллельных столбцов: в левом столбце варианты одной выборки X , в правом другой совокупности Y . Количество столбцов при анализе соответствует числу сопоставляемых выборок.

Процедура применения непараметрических методов состоит из двух этапов. Первый этап заключается в вычислении эмпирического значения критерия в соответствии с используемым методом или критерием. На втором этапе по эмпирическому значению критерия определяют вероятность статистической гипотезы (о различии, о связи, согласованности); такое определение осуществляют двумя способами. Более простой способ - сравнение эмпирических значений критериев с их критическими значениями, имеющимися в статистических таблицах, причем принцип сравнения у части критериев отличается от такого у параметрических тестов.

Для решения задачи анализа нахождения настроек, оказывающих положительное влияние на работу сети и приближенность значений ее количественных параметров к эталонным предлагается использование критериев различия по форме распределения по форме. Эти критерии могут вскрыть различие между выборками с одинаковыми средними в том случае, когда выборки различаются только характером распределения вариантов (дисперсиями, эксцессами, направлением асимметричности). Для нашей задачи использование именно таких методов является оптимальным, так как позволит отфильтровать только те настройки, изменение которых поможет достичь ожидаемых параметров конфигурации телекоммуникационной сети даже в том случае, если в усредненные показатели останутся неизменными.

Например, изменение ширины канала может улучшить либо

ухудшить пропускную способность сети, в то время как, среднее значение пропускной способности не изменится. Именно непараметрические критерии различия помогут найти внутренние зависимости между настройками системы и значением параметров сети и выделить только те настройки, от которых параметры сети зависят.

Чаще всего для оценки взаимосвязей между двумя выборками используют линейный коэффициент корреляции Пирсона. Эффективность оценки связи между переменными при помощи этого коэффициента зависит, как и вообще любая статистическая характеристика, от определенных особенностей сопоставляемых совокупностей. Линейный коэффициент корреляции оказывается наилучшей оценкой связи, если x и y являются нормальными переменными, то есть их распределение является нормальным с устойчивыми средними значениями.

Если это не так, то обычный коэффициент корреляции может служить лишь приближенной характеристикой меры связи, и это приближение тем хуже, чем больше отличаются от нормального распределения наблюдаемые выборки. Если проанализировать эту закономерность более точно, то решающим здесь окажется не столько вид распределения, сколько форма связи (так называемая форма уравнения регрессии) между сопоставляемыми величинами.

В нашем случае, соотношение между параметрами и выборками не отвечает условию “нормальности” и, очевидно, данный метод применять нельзя.

Для тех случаев, когда ничего нельзя сказать о нормальности, но наблюдается монотонное, но не обязательное, линейное изменение линии регрессии (мы не можем гарантировать, что увеличение объема буферов обмена линейно уменьшит длину очереди пакетов в маршрутизаторе), предлагается получать коэффициенты корреляции не для самих выборок, но для их рангов.

Такой критерий носит название “Коэффициент корреляции рангов Спирмена”. Для его использования выборки надо сначала проранжировать. Обычно ранжируют в порядке убывания значения признака, так что самому большому значению будет соответствовать единица.

На следующем этапе находят значения разностей рангов $d = r_x$

- гу в каждой паре. Далее разности рангов возводят в квадрат и суммируют квадраты разностей d^2 . Коэффициент корреляции рангов Спирмена чаще всего вычисляют по формуле

$$r = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

При наличии в одной формуле одинаковых рангов значение r заметно занижается, и его необходимо вычислять по формуле.

Введение поправки на одинаковые ранги существенно повышает мощность коэффициента Спирмена. Для этого, отдельно в выборке x и y рассчитывают независимые поправки по формуле

$$\sigma = \sum \frac{\tau^3 - \tau}{12}$$

где τ – количество одинаковых рангов с одним и тем же значением.

Вычисленное значение r оценивают по трем параметрам: знаку, абсолютной величине и достоверности. Знак коэффициента корреляции показывает форму связи. Знак $+$ характерен для прямой связи, отрицательное значение коэффициента говорит об обратной связи, ноль в свою очередь, означает отсутствие связи между выборками.

Абсолютная величина отражает тесноту прямолинейной связи (степень приближения эмпирических точек на графике взаимозависимости рангов к прямой линии). Большинство авторов считают, что при значениях меньших 0,3 теснота связи слабая, при значениях от 0,3 до 0,5 – умеренная, при 0,5 до 0,7 – заметная, и, наконец, при значениях больше 0,7 – высокая. Квадрат коэффициента корреляции показывает, какая доля изменчивости одного признака определяется изменчивостью другого признака.

Достоверность коэффициента корреляции можно определить через формулу вычисления стандартной ошибки m

$$m = \frac{\sqrt{1 - r^2}}{\sqrt{n - 2}}$$

Эмпирическое значение t -критерия достоверности сравнивают с табличными значениями критерия Стьюдента для $n - 2$ степеней свободы, где n – количество разностей рангов.

$$t = \frac{r}{m} = \frac{r\sqrt{n - 2}}{\sqrt{1 - r^2}}$$

Таким образом, использование коэффициента корреляции рангов Спирмена проектировщик сети получает мощный аппарат фильтрации конфигурационных настроек, не влияющих реально на параметры сети. С помощью данного метода можно промоделировать поведение телекоммуникационной сети и, применив критерий Спирмена, оптимизировать процесс подстройки сети, оставив лишь те настройки, которые влияют на выделенные основные параметры сети. Также, учитывая знак коэффициента корреляции, можно будет определить тип зависимости (прямая или обратная). Абсолютная величина коэффициента покажет насколько тесно параметр сети зависит от того или иного настроечного параметра, что позволит дополнительно отфильтровать те, зависимость от которых слабая или незначительная, тем самым еще более оптимизировать и упростить процесс работы с сетью и, следовательно, увеличить скорость подстройки сети до достижения эталонных значений параметров.

С помощью предложенного подхода были проанализированы такие параметры телекоммуникационной сети как пропускная способность, загруженность, час ответа, коэффициент загрузки.

Был рассмотрен набор возможных конфигурационных настроек и построены соответствующие выборки значений настроек, а так же выборки соответствующих им значений параметров системы. В результате оценивания были выделены настройки, влияющие на эффективность, и доказана высокая точность оценивания.

ЛИТЕРАТУРА

1. И.И. Коваленко, А.П. Гожий Нетрадиционные методы статистического анализа данных. Николаев: «Илион», 2006. – стр. 10-70
2. Л.С. Глоба. Математичні основи побудови інформаційно-телекомунікаційних систем. - Київ, “Норіта-Плюс”, 2007. - 400 с.
3. Рунион Ф. Справочник по непараметрической статистике. Пер с англ. – М.: Финансы и статистика, 1982. – 198с.
4. Тарасенко Ф.П., Непараметрическая статистика. – Томск, : ТГУ, 1976.
5. Тюрин Ю.Н., Непараметрические методы статистики. – М.: «Знание», 1978.
6. Холлендер М., Вулф Д.А. Непараметрические методы статистики - М.: Финансы и статистика 1983 г.
7. Малета Ю.С., Тарасов В.В. Непараметрические методы статистического анализа в биологии и медицине. – М.: Издательство Московского университета, 1982. – 178 с.
8. Гублер Е.В., Генкин А.А., Применение непараметрических методов статистики в медико-биологических исследованиях. – Л.: «Медицина», 1973.

Получено 22.01.2010г.