

6. Косолап А.И. Обобщение симплекс-метода для решения задач полуопределенной оптимизации / А.И. Косолап // Математичне та компютерне моделювання. – 2010. С. 99–106.

УДК 004.94

## **ІМІТАЦІЙНЕ МОДЕЛЮВАННЯ СИСТЕМИ РОЗПОДІЛЕНОГО ЗБЕРІГАННЯ ФАЙЛІВ З РЕГУЛЬОВАНОЮ НАДЛИШКОВІСТЮ**

Г.В. Кузьменко

*Національний технічний університет України «Київський політехнічний  
інститут»*

На тлі бурхливого розвитку Internet-технологій, все більш актуальною стає задача побудови системи, яка б забезпечувала можливість зберігання величезних обсягів інформації і в той же час гарантувала стабільний час доступу до інформаційного ресурсу, а також включала в себе механізми забезпечення безвідмовного доступу до інформації. Більшість існуючих систем, що надають різні Internet-сервіси, як правило, управляють різними інформаційними ресурсами, які розподілені між різними комп'ютерами мережі (далі ці комп'ютери, що належать системі, будемо називати серверами). При побудові таких систем однією з головних проблем є забезпечення стабільного часу задоволення клієнтських запитів в умовах завантаженої мережі. По-перше, через високу зовнішню завантаженість мережі частина серверів системи може бути просто недоступна. По-друге, через нерівномірний доступ користувачів до інформації, в деякий момент часу може скластися ситуація, при якій дуже багато клієнтських запитів, будуть направлені до одного й того ж інформаційного ресурсу, що призведе до істотного завантаженні мережі, навколо серверів, що містять бажаний ресурс. Це призводить до того, що деякі сервери системи стають недоступними, хоча до інших доступ вільний.

Кожен файл пропонується зберігати у вигляді деякого набору частин (шматків), кількість яких може змінюватися в часі, проте завжди в будь-який момент часу для існуючих в системі частин виконується умова, що з будь-яких  $k$  частин можна повністю зібрати файл. Цю умову будемо називати далі умовою надлишковості. Розмір кожної частини  $\sim 1/k$  від розміру файлу що зберігається. Кількість частин файлу -  $n$ , що містяться в системі може змінюватися в залежності від її конфігурації, від кількості комп'ютерів в ній, інтенсивності роботи з даним файлом і т.д. причому, при зміні (збільшенні) числа  $n$  вже існуючі частини ніяк не змінюються, тобто нові частини будуються таким чином, щоб як і раніше виконувалася умова надлишковості[2].

Побудова такої системи вимагає значних зусиль та ресурсів, тому перед її побудовою доцільно провести моделювання її роботи, з ціллю визначення ефективності її роботи, знаходження недоліків в моделі та усу-

нення їх, визначення параметрів для конфігурування системи, що буде побудована. Зважаючи на обсяг задач які покриває система та різноманітність цих задач (кодування та декодування файлів, визначення попиту на файл в частинах мережі, генерація додаткових частин для файлів з високим попитом та знищення зайвих для файлів з низьким, передача даних) доцільно використати імітаційне моделювання.

Імітаційну модель пропонується будувати у вигляді повної програмної емуляції компонентів системи. Основними випадковими величинами що впливають на роботу системи є:

- інтервал між надходженнями запитів на файл, що має враховувати можливість надвисокого попиту на деякі файли (середнє значення інтервалу менше секунди);
- інтервал між виконанням викладання файлу в мережу;
- розмір файлу що викладається;
- кількість доступних серверів.

Інтервал між надходженнями запитів, інтервал між виконанням викладання та кількість доступних серверів залежать від часу доби, оскільки вночі кількість користувачів в мережі зменшується. Також деякі з них вночі вимикають комп'ютери, які є серверами системи, що зменшує кількість доступних частин деяких викладених файлів. Отже система повинна моделювати зміну часу доби у різних частинах мережі.

З побудованої імітаційної моделі можна отримати наступні значення:

- середню швидкість завантаження файлів;
- мінімальну та максимальну швидкість завантаження;
- кількість випадків неможливості задоволення запиту на завантаження файлу.

За отриманими даними конфігурація системи змінюється таким чином, щоб збільшити середню швидкість завантаження. Окремо розглядаються випадки з мінімальною швидкістю завантаження для знаходження, та у разі можливості усунення, причин. Кількість випадків неможливості задоволення запиту на завантаження файлу має бути рівною нулю, в іншому випадку змінюються налаштування надлишковості таким чином щоб гарантувати доступність файлу.

Отже, побудовано імітаційну модель системи розподіленого зберігання файлів з регульованою надлишковістю, що дає можливість отримати інформацію необхідну для побудови такої системи, а також виконувати конфігурацію системи для більш ефективної роботи.

## **Література**

1. Томашевський В.М. – Моделювання систем / Томашевський В.М. // Видавнича група ВНУ 2005. – 352с
2. Тормасов А.Г. – Модель распределенного хранения данных с регулируемой избыточностью [Электронный ресурс] / А.Г. Тормасов, М.А. Хасин,

Ю.И. Пахомов // Режим доступа:  
<http://zhurnal.ape.relarn.ru/articles/2001/035.pdf>  
3. Курош А.Г. – Курс высшей алгебры / Курош А.Г. // М.:Наука 1975. – 423 с.

УДК 519.237.5

## **ОСНОВНЫЕ ПРОБЛЕМЫ ПОСТРОЕНИЯ РЕГРЕССИОННЫХ МОДЕЛЕЙ**

Лапач С.Н., Радченко С.Г.

*Национальный технический университет «Киевский политехнический институт»*

Регрессионный анализ (РА) является одной из наиболее распространенных и мощных методов многомерного статистического анализа.

Широкое применение РА в настоящее время обуславливается следующими факторами: быстрая смена изучаемых областей (технологий, материалов, условий эксплуатации), которая не позволяет тратить много времени на исследования и получение требуемых зависимостей; расширения областей применения (социология, история и пр.); повсеместная автоматизации, требующая множества моделей, применимых в данном конкретном случае с заданной точностью. Тем не менее, следует признать, что как в теории, так и практическом применении регрессионного анализа положение далеко не лучшее.

Применение регрессионного анализа переросло свои разработанные теоретические основания. Например, в [1] 1198с., а в [2] – 912с. Это хорошие книги, но научного работника или инженера, для которого регрессионный анализ всего лишь один из инструментов, такой объем может лишь отпугнуть.

Планирование эксперимента (ПЭ), которое должно было бы стать основанием и неотъемлемой частью РА само находится в таком же состоянии и рассматривается как самостоятельная научная дисциплина с сильным уклоном от математической статистики и выборочного метода в сторону «чистой» математики. Кроме того, некоторые представители научных кругов считают ПЭ устаревшим и ненужным, т.е. наблюдается использование РА и ПЭ формально, без понимания сущности, ограничений и предпосылок.

Многие проблемы в построении эмпирических моделей есть следствие незавершенного исторического пути построения соответствующей теории. Регрессионный анализ и теория планирования экспериментов ни по отдельности, ни вместе не представляют собой единых теорий: это множество отдельных теорем и методов, собранных вместе постепенно и по необходимости решать реальные задачи, на что указывал еще Налимов В.В. Очень часто при решении конкретных задач происходит конфликт

различных частей, входящих в эти теории. Для решения каждой проблемы разрабатывается много средств, ad hoc, т.е. для каждого конкретного случая. Такой подход только усугубляет ситуацию в связи с остающейся в целом несогласованностью как между ПЭ и РА, так и между отдельными частями РА.

Современное состояние дел в данной области требует построение новой теории, объединяющей регрессионный анализ и планирование эксперимента и разрешающей существующие противоречия.

При подходе к построению регрессионной модели следует исходить из цели исследования. Использование регрессионного анализа в терминах прикладной статистики возможно со следующими целями: выяснение наличия статистической связи между случайными переменными (раньше называлось нелинейной корреляцией); аппроксимация; построение математических моделей [3]. При построении моделей обычно требуется семантическая, т.е. возможность объяснения с помощью модели происходящего процесса или явления. Это требует, чтобы структура уравнения регрессии и свойства коэффициентов в некотором смысле соответствовали исследуемому процессу. И в этом заключается главная проблема.

Вся традиционная теория ПЭ и РА опирается на постулат априорной известности «истинной» структуры регрессионной модели. Сами свойства оценок сохраняются при соответствии структуры «истинной» [4]. Оптимальность плана, свойства регрессионных оценок, исследование уравнение на наличие гетероскедастичности, выбросов и прочее, все опирается на указанный постулат. В прикладных исследованиях в подавляющем большинстве ситуаций указать эту структуру невозможно. Более того, исследователь во многих случаях как раз и желает получить эту структуру как результат исследования.

Рассмотрим построение линейных по параметрам (и в общем случае нелинейных по факторам) регрессионных моделей, считая структуру заранее неизвестной. Полагаем также, что имеющаяся в нашем распоряжении информация о процессе ограничена, т.е. мы имеем дело с «черным» или «серым» ящиком. Сразу скажем, что регрессионные модели не могут описывать любые процессы, а только те, которые можно представить в виде  $\hat{Y} = f(X) + \varepsilon$ , где  $f(X)$  – некоторая детерминированная функция от множества факторов, а  $\varepsilon$  – случайная величина.

В общем виде построение регрессионной модели проходит следующие этапы.

1. Формирование выборки (плана эксперимента).
2. Предварительная спецификация модели.
3. Окончательная спецификация модели.
4. Идентификация модели.
5. Оценка качества полученной модели.