

УДК 004.032

ББК 3.32

АНАЛИЗ ЧУВСТВИТЕЛЬНОСТИ ВЕБ-СЕРВИСОВ В ЗАДАЧЕ ВЫБОРА ОПТИМАЛЬНОЙ КОНФИГУРАЦИИ СИСТЕМ С СЕРВИСНО-ОРИЕНТИРОВАННОЙ АРХИТЕКТУРОЙ¹

Душкин Д. Н.²,

*(ФГБУН Институт проблем управления им. В. А. Трапезникова
Российской академии наук, Москва)*

Данная работа посвящена проблеме выбора корректных критериев сравнения веб-сервисов по предпочтению. Проводится обзор и анализ существующих подходов выбора таких критериев. Подробно рассматривается вопрос оценки чувствительности веб-сервисов. Вводится понятие классов чувствительности веб-сервисов, описывается методика выделения и автоматизированного определения классов с помощью техник машинного обучения. Приводятся результаты вычислительного эксперимента, реализующего методику. Обозначаются перспективные направления исследований.

Ключевые слова: веб-сервисы, сервисно-ориентированная архитектура, машинное обучение, анализ чувствительности.

1. Введение

Предоставление сервиса вместо аппаратного и программного обеспечения стало одним из ведущих направлений современной индустрии информационных технологий. Данная тенденция во многом меняет экономику сферы информационных техноло-

¹ Автор признателен Смирновой Н.В., Миркину Б.Г. и Фархадову М.П. за ценное обсуждение содержания статьи.

² Душкин Дмитрий Николаевич, аспирант (ddushkin@asmon.ru, тел. 8(916)978-58-96).

гий и оказывает влияние на формирование информационного общества в целом [17].

В основе сервис-ориентированных систем лежит *веб-сервис* – автономный вычислительный ресурс, предоставляющий свои функции через сеть Интернет (или Интранет) посредством открытых протоколов обмена данными, не зависящих от платформ как самих ресурсов, так и связываемых с ними программных систем. Системы, использующие в своей работе веб-сервисы, называют системами с сервисно-ориентированной архитектурой (СОА).

Часто одну и ту же функцию предоставляют различные сервисы. Например, функцию геокодинга, т.е. конвертацию адреса в географические координаты (например, «Москва, ул. Профсоюзная, 65» в «55.651345, 37.5382»), предоставляют сервисы *Google Maps*, Яндекс Карты, *Bing Maps* (сервис *Microsoft*), *Nokia Maps* и др. В таком случае перед инженерами, проектирующими систему с СОА, ставится многокритериальная задача выбора оптимального по предпочтениям веб-сервиса, предоставляющего функцию геокодинга.

В обозреваемых работах (см. далее) выбор критериев сводится к ряду простых, лежащих на поверхности: среднее время отклика, наработка на отказ и ряду других. Не было найдено таких критериев, которые описывали бы веб-сервис в долгосрочной перспективе, что важно в ситуациях, когда необходимо быть уверенным в предлагаемом уровне производительности при увеличении нагрузки. Определение такого критерия может быть полезно как при выборе оптимальной архитектуры системы с СОА, так и при составлении документов, регламентирующих соглашение об оказании уровня услуг (англ. *service level agreement, SLA*).

Примером такого критерия может служить *чувствительность* веб-сервиса. Чувствительность – критерий, на основе которого может быть оценена возможность обеспечения определенного уровня производительности при возрастающей нагрузке [12].

Работа состоит из четырех основных частей. Первая часть

посвящена обзору современного состояния исследований критериев сравнения веб-сервисов и систем с СОА в целом. Вторая часть посвящена решению задачи оценки чувствительности веб-сервиса. В третьей части кратко описываются результаты вычислительного эксперимента, проведенного на основе выборки из 50 веб-сервисов. В заключительной части делаются выводы по проведенной работе и обозначаются предстоящие направления исследований.

2. Обзор литературы

В российской научной литературе исследованию веб-сервисов уделено мало внимания. В работе [3] веб-сервисы используются как ресурсы, содержащие различные образовательные материалы. В качестве критериев сравнения выбраны класс онтологии и степень близости ресурсов веб-сервиса к заданному классу онтологии. Приводится алгоритм формирования набора веб-сервисов для заданного контекста. Обсуждаемые критерии выбора не являются универсальными и подходят только для узкой предметной области, поэтому результаты этой работы невозможно использовать для решения общей задачи выбора веб-сервисов.

В работе [1] предлагается алгоритм выбора резервных веб-сервисов на основе вычисления интегральной оценки веб-сервиса — весовой функции, вычисляемой по следующим критериям: количество отказов, деленное на количество вызовов (другими словами доступность), среднее время обслуживания запроса и стоимость одного запроса, заданная экспертом. Каждый из критериев умножается на весовой коэффициент, отражающий важность критерия в данный момент времени, и полученные значения складываются. Путем сравнения таких интегральных оценок выбирается альтернативный веб-сервис. Стоит заметить, что в данном подходе не рассматриваются функциональные требования к веб-сервису (т.е. не учитываются предоставляемые веб-сервисом функции); при составлении интегральной оценки веб-сервиса критерии не приводятся к единой шкале для осуществле-

ния возможности их сравнения по предпочтению; не рассмотрены способы получения весовых коэффициентов.

В работе [2] рассматривается применение теории нечетких чисел в задаче выбора сервисов для реализации определенных бизнес-процессов в рамках корпоративной информационной системы. Используя терминологию настоящей работы, можно сказать, что в [2] выделяют ряд функциональных критериев (критерии реализации бизнес-процессов) и нефункциональных. К нефункциональным относятся оценки экономических затрат: единовременных, периодических и косвенных. Другие нефункциональные критерии не рассматриваются. Оценки качества реализации функциональных требований определяются экспертно. В целом, в работе в основном рассматривается экономический аспект выбора веб-сервисов, в то время как техническая сторона вопроса, касающаяся надежности и производительности веб-сервисов, практически не освещена.

Большое количество исследований по сервисно-ориентированным вычислениям проведено в рамках европейского научно-исследовательского проекта *SENSORIA*. Часть результатов этих исследований посвящена методам оценки качества веб-сервисов (см., например, [5, 7, 9, 11, 16, 18]). Необходимым условием применимости данных методов является наличие модели оцениваемого веб-сервиса в формализме одной из алгебр случайных процессов. Разработка такой модели требует определенных затрат времени и ресурсов, которые имеет смысл затрачивать разработчику веб-сервиса, но не внешнему пользователю, периодически сталкивающемуся с необходимостью выбора одного из существующих на рынке веб-сервисов для реализации одной из многих функций разрабатываемой им системы массового обслуживания.

Особого внимания заслуживает работа [18], в которой предлагается алгоритм формирования коллективной оценки веб-сервиса, использующий в качестве мнений экспертов (выборщиков) степени доверия одних веб-сервисов к другим. Этот алгоритм схож с алгоритмом *PageRank* [4], используемым в поиско-

вом сервисе *Google* для оценки степени полезности сайтов в зависимости от количества и качества внешних ссылок на эти сайты. Данный подход имеет следующие недостатки: 1) для определения производительности веб-сервиса используется только два числовых показателя (время ответа и доступность); 2) для использования подхода необходимы сведения о том, как ранжируемые веб-сервисы связаны между собой, при этом предполагается, что топология композиции веб-сервисов известна, т.е. доступна информация о том, как веб-сервисы взаимодействуют друг с другом. Второй недостаток существенен, поскольку данные предположения противоречат принципу слабой связности веб-сервисов, являющемуся одним из основных положений подхода к организации сервисно-ориентированных вычислений.

Стоит отметить статью [12], в которой рассматривается такая характеристика, как чувствительность веб-сервиса, отражающая изменение числовых показателей системы при варьировании нагрузки на нее. В статье [12] рассмотрен только случай линейного увеличения нагрузки на систему, при котором отмечается логарифмическое ухудшение пропускной способности сервиса. Кроме того исследовался только «идеальный» лабораторный вариант, когда одна машина в рамках локальной сети тестировала другую. Делалось предположение о схожести полученных результатов исследований для всех веб-сервисов, что неверно при определенных конфигурациях низлежащей платформы (например, в случае «облачной» инфраструктуры) и характере дополнительного внешнего трафика веб-сервиса.

В статьях [11, 13] предлагаются онтология и расширение языка *XML*, которые должны использоваться: 1) провайдерами сервисов для описания качества предоставляемого обслуживания; 2) потребителями сервисов для описания требуемого качества обслуживания. Предлагаемая онтология разделена на три уровня. С помощью верхнего уровня описываются сведения об основных концепциях, связанных с измерением показателей качества обслуживания, таких как способ оценки показателя (субъективный или объективный), наличие корреляций между рас-

смаатриваемыми показателями и т.д. С помощью среднего уровня онтологии описываются сведения о наиболее значимых по мнению авторов показателях качества обслуживания распределенных систем, таких как доступность сервиса, производительность, надежность и т.д. Нижний уровень предназначен для описания сведений о значениях показателей, специфичных для предметной области сервиса. Авторы предлагают использовать многоагентную систему, в которой должны регистрироваться провайдеры сервисов. Предполагается, что они в дальнейшем будут отсылать ей значения показателей качества обслуживания, а потребители веб-сервисов будут обращаться к ней для выбора наиболее подходящего веб-сервиса. Предлагается способ выбора веб-сервиса для двух показателей качества обслуживания, учитывающий корреляции между показателями. Недостатком подхода является необъективность оценок показателей качества, основанная на доверии к заявляемым провайдерами показателям.

В работе [6] описан способ вычисления интегральной оценки качества веб-сервисов, основанный на использовании утилиты *ping*, посылающей запросы по протоколу ICMP. Несмотря на то, что функционал утилиты *ping* является частью стека *TCP/IP*, часто сервера настраивают на игнорирование подобных запросов с целью предотвращения проведения атак злоумышленником. Кроме того, в качестве показателей, на основе которых формируется интегральная оценка качества, авторы предлагают использовать только среднее время отклика, время отклика на последний запрос и т.д. При этом не учитывается возможность периодического изменения значений показателей в зависимости от времени суток.

В [8] подробно рассматривается проблема измерения и предсказания производительности систем. При этом изучаемые системы представлены в виде «черного ящика». С помощью методов статистического машинного обучения формируется модель, способная спрогнозировать последующее поведение системы при различных условиях, например, при изменении числа одновременных пользователей системы. Данная работа близка по методо-

логическому содержанию к настоящей работе, но с тем отличием, что в ней в качестве систем рассматриваются только реляционные базы данных с привязкой к синтаксису языка *SQL*.

Из приведенного обзора можно сделать вывод о малом внимании к критериям, описывающим производительность и надежность веб-сервисов при растущей нагрузке. Такие сведения важны при проектировании системы с СОА на этапе определения максимальной загруженности, при которой система сможет обеспечить приемлемый уровень предоставляемых услуг при данном наборе используемых веб-сервисов. В настоящей работе исследуется критерий чувствительности, позволяющий решить описываемую проблему.

3. Определение чувствительности веб-сервиса

Чувствительность веб-сервиса – критерий, на основе которого может быть оценена возможность обеспечения определенного уровня производительности веб-сервиса при возрастающей нагрузке.

В следующих подразделах работы освещена процедура выделения классов чувствительности и распределения классов по предпочтению. Класс чувствительности определяет принадлежность веб-сервиса к группе, характеризуемой определенными диапазонами значения показателей. Автоматизированное определение класса чувствительности веб-сервисов с помощью алгоритма машинного обучения «с учителем» – методом опорных векторов (англ. *Support vector machine, SVM*), а также описание системы, осуществляющей все шаги определения чувствительности, будут описаны в следующих работах.

Выделения классов чувствительности веб-сервисов реализуется следующим образом:

- 1) формируется случайная выборка веб-сервисов (см. подраздел 3.1);
- 2) проводится тестирование выборки по определенному плану (см. подраздел 3.2);

- 3) составляется матрица «объект–признак». С этой целью полученные данные преобразовываются для выделения ряда признаков (факторов), отражающих характер изменения производительности при возрастающей нагрузке (см. подраздел 3.3);
- 4) проводится эвристическое разделение множества веб-сервисов на классы по уровню чувствительности, экспертно классы упорядочиваются по предпочтению (см. подраздел 3.4);
- 5) для проверки корректности эвристического разделения веб-сервисов на классы проводится кластеризация данных с помощью алгоритма машинного обучения «без учителя» – методом k -средних (см. подраздел 3.5).

3.1. ФОРМИРОВАНИЕ ВЫБОРКИ

Выборка веб-сервисов формируется на основе данных из каталога *API Directory* [15], содержащего информацию о более чем 5000 различных веб-сервисах. Выбираются разнородные сервисы, реализующие функции картографии и геокодинга, предоставляющие информацию о различных показателях торговых бирж, о погоде, новостях и т.д. Все сервисы реализуют архитектуру *REST*, предоставляя свои функции по протоколу *HTTP*.

3.2. ПЛАН ТЕСТА

Задача теста – определить величину среднего времени обработки запросов, стандартное отклонение времени обработки запросов и количество необработанных запросов при заданной нагрузке. В процессе тестирования осуществляется последовательное выполнение итераций, отличающихся числом запросов в секунду. В рамках каждой итерации в течение секунды отправляется определенное число запросов к веб-сервису по протоколу *HTTP*. Запросы равномерно распределены в рамках секунды.

Пусть λ_{max} – максимальное число запросов в секунду, S – шаг теста (число, на которое увеличивается количество запросов

в секунду в последующей итерации). Тогда можно посчитать общее число итераций в тесте N_{iter} :

$$(1) \quad N_{iter} = \left\lceil \frac{\lambda_{max}}{S} \right\rceil$$

Пусть $\mathbf{r} \in \mathbb{R}^k$ — вектор, содержащий время обработки запросов всего теста, где k — общее число отосланных запросов за все итерации. Для простоты дальнейших вычислений, если i -ый запрос не был обработан или был обработан с ошибкой, то $r_i = 0^3$.

Пусть $Iter \in \mathbb{R}^{N_{iter} \times k}$ — матрица, содержащая информацию о принадлежности запросов к определенным итерациям. $Iter_{i,j} = 1$, если запрос j выполнялся в рамках итерации i , иначе $Iter_{i,j} = 0$. Например,

$$Iter = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

В примере запрос 1 отправлен в первой итерации, запросы 2 и 3 — во второй итерации, 4-ый запрос — в третьей.

3.3. ПОЛУЧЕНИЕ МАТРИЦЫ «ОБЪЕКТ-ПРИЗНАК»

Матрица «объект-признак» $X \in \mathbb{R}^{m \times n}$, где m — число объектов, n — количество признаков, формируется на основе результатов тестов веб-сервисов. Результат теста состоит из вектора времен обработки запросов \mathbf{r} и матрицы распределения запросов по итерациям $Iter$. Ряды матрицы соответствуют объектам (веб-сервисам), столбцы — признакам. Обозначим $\mathbf{x}^{(i)}$ i -ый ряд матрицы X . Вектор $\mathbf{x}^{(i)}$ является объединением трех векторов:

$$\mathbf{x}^{(i)} = \mathbf{t}^{(i)} \cup \mathbf{d}^{(i)} \cup \mathbf{e}^{(i)}$$

, где вектор $\mathbf{t}^{(i)} \in \mathbb{R}^{N_{iter}}$ содержит среднее время обработки запросов по итерациям i -го веб-сервиса, $\mathbf{d}^{(i)} \in \mathbb{R}^{N_{iter}}$ — стандартное отклонение времен обработки запросов по итерациям,

³ При вычислении среднего арифметического и стандартного отклонения исключаются те запросы, время обработки которых равно нулю ($r_i = 0$).

$\mathbf{e}^{(i)} \in \mathbb{R}^{N_{iter}}$ – количество ошибочных или необработанных запросов по итерациям (далее для краткости такие запросы будем называть просто «необработанные запросы»). Таким образом общее количество признаков $n = 3 * N_{iter}$.

Вектор $\mathbf{t}^{(i)}$ вычисляется следующим образом:

$$t_j^{(i)} = \frac{1}{j \cdot S} \sum_{l: Iter_{j,l}^{(i)}=1} r_l^{(i)}$$

где $i = \overline{1, \dots, n}$ – номер веб-сервиса, $j = \overline{1, \dots, N_{iter}}$ – номер итерации, $l = \overline{1, \dots, k}$ – номер запроса, $\mathbf{r}^{(i)}$ – вектор, содержащий время обработки запросов i -го веб-сервиса, $Iter^{(i)}$ – матрица, содержащая информацию о принадлежности запросов к определенным итерациям. Запись $\sum_{l: Iter_{j,l}^{(i)}=1} r_l^{(i)}$ следует понимать как «сумма по элементам $r_l^{(i)}$, где l такое, что $Iter_{j,l}^{(i)} = 1$ », или «сумма по элементам, принадлежащим i -ой итерации».

Вектор $\mathbf{d}^{(i)}$:

$$d_j^{(i)} = \sqrt{\frac{1}{j \cdot S - 1} \sum_{l: Iter_{j,l}^{(i)}=1} (r_l^{(i)} - t_j^{(i)})^2}$$

Вектор $\mathbf{e}^{(i)}$:

$$e_j^{(i)} = \sum_{l: Iter_{j,l}^{(i)}=1, r_l=0} 1$$

Нормализуем полученные данные. Для этого зададим функцию нормализации g_j :

$$g_j : x_{i,j} \mapsto \frac{x_{i,j} - \min_j x_{i,j}}{\max_j x_{i,j} - \min_j x_{i,j}}$$

Примем за $X' = g(X)$ нормализованную по признакам матрицу «объект-значение».

3.4. ЭВРИСТИЧЕСКОЕ ВЫДЕЛЕНИЕ КЛАССОВ ЧУВСТВИТЕЛЬНОСТИ ВЕБ-СЕРВИСА

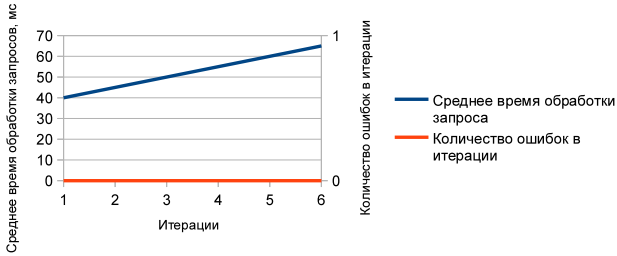
Анализируя визуальное представление матрицы «объект-признак» (см. рис. 3), полученной в ходе тестирования 50 веб-сервисов, было выявлено, что реальные веб-сервисы имеют большой разброс значений критериев. В ходе анализа нетрудно определить случаи низкой чувствительности, когда повышение нагрузки практически не влияет на значения показателей, и случаи высокой чувствительности, когда небольшое повышение нагрузки значительно увеличивает среднее время обработки запросов и часто ведет к отказу в обслуживании большей части запросов. Однако большинство веб-сервисов демонстрируют промежуточное поведение и возникают сложности с определением их класса чувствительности.

Ниже представлено эвристическое разделение веб-сервисов на классы чувствительности: от предпочтительной низкой чувствительности к высокой. При эвристическом разделении учитывались только показатели среднего времени обслуживания запроса и количество необработанных или ошибочно обработанных запросов на итерацию. Стандартное отклонение в большинстве случаев имеет линейную корреляцию со средним временем обработки запросов.

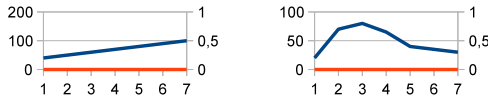
Первый класс (рис. 1(a)) — низкая чувствительность, характеризуется медленным повышением (иногда отсутствием повышения) среднего времени обслуживания запросов, отсутствием необработанных запросов.

Второй класс (рис. 1(b)) – средняя чувствительность, характеризуется более быстрым повышением среднего времени обслуживания запросов по сравнению с первым классом, отсутствием необработанных запросов.

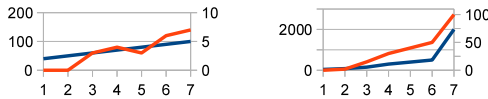
Третий класс (рис. 1(c)) – повышение с последующим понижением среднего времени обслуживания запросов, отсутствие необработанных запросов. Такое поведение характерно для ряда «облачных» веб-сервисов (при увеличении нагрузки динамически увеличивается мощность обслуживающего узла) или веб-



(a) 1 класс чувствительности (низкая чувствительность)



(b) 2 класс чувствительности (c) 3 класс чувствительности



(d) 4 класс чувствительности (e) 5 класс чувствительности (высокая чувствительность)

Рис. 1. Концептуальное определение классов чувствительности

сервисов с адаптивным распределителем нагрузки (англ. *Load balancer*) (при высокой утилизации ресурсов одного обслуживающего узла часть запросов передается на обслуживание узлам с меньшей утилизацией).

Четвертый класс (рис. 1(d)) по характеру повышения среднего времени обслуживания похож на второй, но с повышением нагрузки появляется небольшое число необработанных запросов.

Пятый класс (рис. 1(e)) — высокая чувствительность, характеризуется резким повышением среднего времени обслуживания, большим числом необработанных запросов.

3.5. АВТОМАТИЗИРОВАННОЕ ВЫДЕЛЕНИЕ КЛАССОВ ЧУВСТВИТЕЛЬНОСТИ

С целью проверки обоснованности эвристического разделения веб-сервисов, а также с учетом последующей автоматизации процесса определения класса чувствительности используется кластеризация данных с помощью алгоритма k -средних [10]. В качестве входных данных используются нормализованная матрица «объект-признак», 5 центроидов (по количеству эвристически определенных классов), каждую итерацию центроиды выбираются случайно, всего проводится 50 итераций, в качестве меры расстояния используется расстояния Евклида.

Результаты кластеризации на экспериментальной выборке приводятся в следующем разделе.

4. Проведение тестов и анализ полученных данных

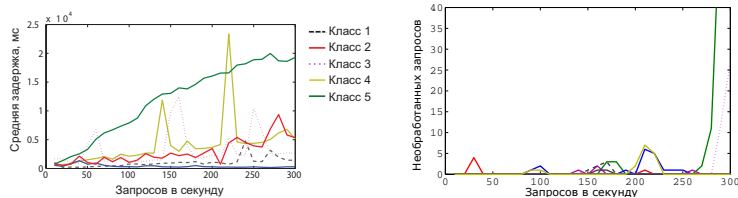
Зададим план теста: максимальное число запросов в секунду $\lambda_{max} = 300$, шаг теста $S = 10$. Такие значения параметров выбраны экспериментально, т.к. было установлено, что начиная с 280–290 запросов в секунду большинство веб-сервисов демонстрируют устойчивое поведение. Выборка состоит из 50 веб-сервисов: Google Maps, Яндекс Карты, Bing Maps, Nokia Maps, Twitter, Factolex, Quora и др. Исходя из плана теста несложно определить размер матрицы $X \in \mathbb{R}^{50 \times 90}$, т.е. матрица представляет 50 объектов, каждый из которых характеризуется 90 показателями.

На рис. 2(а) показана⁴ первая группа признаков «среднее время отклика по итерациям» 5-ти веб-сервисов⁵. Данные выбраны таким образом, чтобы веб-сервисы наглядно можно было бы отнести к различным классам. Каждая линия соответствует одному веб-сервису.

⁴Для удобства вместо номера итерации на оси x показано число запросов в данной итерации.

⁵Если отобразить на графике все 50 веб-сервисов, то график станет слишком зашумленным.

Иллюстрация группы признаков «стандартное отклонение времен обработки запросов по итерациям» отсутствует в силу достаточно устойчивой корреляции с первой группой признаков. На рис. 2(b) показана третья группа признаков «число ошибочных запросов по итерациям».



(а) Группа признаков «среднее время (б) Группа признаков «число ошибочных запросов от числа запросов в секунду» бочных запросов от числа запросов в секунду»

Рис. 2. Ненормализованные группы признаков

На рис. 3 представлены результаты кластеризации. Каждая линия обозначает один веб-сервис, линии одного цвета принадлежат одному кластеру. Также результаты кластеризации можно визуализировать с помощью метода главных компонент [14], уменьшив размерность с 90 до 2 признаков, наиболее полно объясняющих изменчивость и взаимосвязи исходных данных. На графике каждый плюс обозначает один веб-сервис; плюсы, закрасненные одним цветом, принадлежат одному кластеру.

На рис. 4 хорошо видна разделяемость множества веб-сервисов на различные кластеры. Сопоставляя рис. 4 и рис. 3, а также руководствуясь эвристической классификацией, описанной в подразделе 3.4, можно соотнести кластеры с классами чувствительности.

5. Выводы

На основе проведенных исследований можно сделать вывод: реальные веб-сервисы демонстрируют различную произво-

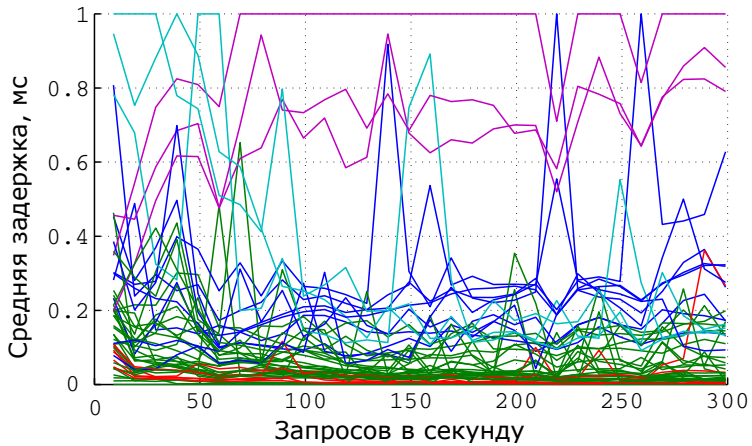


Рис. 3. Нормализованная группа признаков «среднее время отклика от числа запросов в секунду» с распределенными по кластерам веб-сервисами

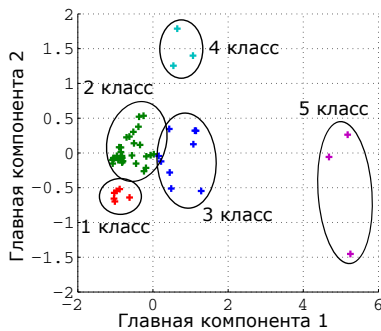


Рис. 4. Применение метода главных компонент к результатам кластерного анализа

дительность и надежность при повышении нагрузки. Выявлению такого показателя, который смог бы охарактеризовать различное поведение веб-сервисов при повышении нагрузки, посвящена настоящая работа.

Используя корректное программное обеспечение для проведения тестов и сбора данных, методы анализа данных и машинного обучения, возможно автоматизировать процесс определения класса чувствительности веб-сервиса. Автоматизация такого процесса позволит учитывать перспективную производительность и надежность веб-сервисов на этапе проектирования систем с сервисно-ориентированной архитектурой, что особенно полезно при создании систем с высокой нагрузкой.

Видится перспективным продолжение исследований по данной тематике. Возможно выделение новых признаков (и отбор наиболее значимых из имеющихся), характеризующих чувствительность. Полезна формализация плана проводимого теста с учетом закона распределения среднего времени обработки запросов и погрешностей, возникающих вследствие неизвестного внешнего трафика веб-сервиса. В виду отсутствия программного обеспечения для формирования оптимального по предпочтениям набора веб-сервисов, используемых системой с сервисно-ориентированной архитектурой, а также с учетом растущего рынка программного обеспечения как услуги (англ. *Software as a Service, SaaS*), видятся перспективными работы по формированию требований к такой системе и её реализация.

В ходе проведения исследований на языке программирования *Erlang*⁶ реализован автономный сервер, проводящий необходимые тесты, с возможностью получения текущего состояния по сети Интернет. Анализ данных, кластерный и факторный анализы выполнены в математической среде *Matlab*.

⁶ *Функциональный язык программирования Erlang был выбран из соображений минимизации системной погрешности при проведении тестов, поскольку данная технология обладает эффективной архитектурой параллелизации вычислений.*

Литература

1. БАБОШИН А.А., КАШЕВНИК А.М. *Подход к организации взаимодействия веб-сервисов на основе модели потока работ* // Труды СПИИРАН. — 2007. — №. 5. — С. 247–254.
2. ЗАТЕСА А.В. *Нечеткая модель стоимости в рамках сервисно-ориентированного подхода к архитектуре информационных систем* // Экономика, статистика и информатика. Вестник УМО. Изд. МГУЭСИ. — 2011. — №. 1. — С. 162–164.
3. СМИРНОВ А.В., ЛЕВАШОВА Т.В., ШИЛОВ Н.Г. *Конфигурирование сервис-ориентированных сетей ресурсов для интеллектуальной поддержки дистанционного образования* // Открытое образование. — М.: Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования ИТМО; Московский государственный университет экономики, статистики и информатики ИТЭИ, 2010. — С. 111-117.
4. BRIN S., PAGE L. *The anatomy of a large-scale hypertextual Web search engine* // Comput. Netw. ISDN Syst. — 1998. — Т. 30, №. 1-7. — С. 107–117.
5. CAPPELLO I., CLARK A., GILMORE S. И ДР. *Quantitative analysis of services* // Rigorous software engineering for service-oriented systems / Под ред. Martin Wirsing, Matthias Holzl. — Berlin, Heidelberg: Springer-Verlag, 2011. — С. 522–540.
6. EGAMBARAM DR., ПАВАРАСАН, VADIVELOU G., SIVASUBRAMANIAN S. *QoS based web-service selection*. — 2010.
7. FOSTER H., LÁSZLÓ G., KOCH N. И ДР. *UML extensions for service-oriented systems* // Rigorous software engineering for service-oriented systems / Под ред. Martin Wirsing, Matthias Holzl. — Berlin, Heidelberg: Springer-Verlag, 2011.

- С. 35–60.
8. GANAPATHI A. *Predicting and Optimizing System Utilization and Performance via Statistical Machine Learning*: тезисы кандидатской диссертации / EECS Department, University of California, Berkeley. – 2009. – С. 97.
 9. GILMORE S, LÁSZLÓ G. И ДР. *Non-functional properties in the model-driven development of service-oriented systems* // Software Systems Modelling – 2011. – Т. 10, №. 3. – P. 287–311.
 10. MACQUEEN J.B. *Some Methods for Classification and Analysis of MultiVariate Observations* // Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability / Под ред. L. M. Le Cam, J. Neyman. – Т. 1. – University of California Press, 1967. – С. 281–297.
 11. MAXIMILIEN E.M., SINGH M.P. *A Framework and Ontology for Dynamic Web Services Selection* // IEEE Internet Computing. – 2004. – Т. 8, №. 5. – С. 84–93.
 12. MAXIMILIEN E.M., SINGH M.P. *Toward autonomic web services trust and selection* // Proceedings of the 2nd international conference on Service oriented computing, ICSOC '04. – New York: ACM, 2004. – С. 212–221
 13. MAXIMILIEN E., SINGH P. *Agent-based trust model involving multiple qualities* // Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems. – AAMAS '05. – New York, NY, USA: ACM, 2005. – С. 519–526.
 14. PEARSON K. *On lines and planes of closest fit to systems of points in space* // Philosophical Magazine. – 1901. – Т. 2, №. 6. – С. 559–572.
 15. PROGRAMMABLE WEB API Directory URL: <http://www.programmableweb.com/apis/directory>.
 16. TRIBASTONE M., GILMORE S. *Scaling performance analysis using fluid-flow approximation* // Rigorous software engineering for service-oriented systems / Под ред. Martin

- Wirsing, Matthias Holzl. — Berlin, Heidelberg: Springer-Verlag, 2011. — С. 486–505.
17. WIRSING M., HOLZL M., KOCH N., MAYER P. *Sensoria — software engineering for service-oriented overlay computers*. — Berlin, Heidelberg: Springer-Verlag, 2011. — С. 1-14
18. WU Q., IYENGAR A., SUBRAMANIAN R. И ДР. *Combining Quality of Service and Social Information for Ranking Services // Service-Oriented Computing* / Ed. by David Hutchison, Takeo Kanade, Josef Kittler et al. — Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. — Т. 5900. — С. 561–575.

SENSITIVITY ANALYSIS OF WEB SERVICES IN THE PROBLEM OF CHOOSING THE OPTIMAL SYSTEM CONFIGURATION WITH A SERVICE-ORIENTED ARCHITECTURE

Dmitry Dushkin, Institute of Control Sciences of RAS, Moscow, Postgraduate (ddushkin@asmon.ru).

Abstract: We study the problem of choosing proper measurement criteria for Web services' ranking. Existing approaches to selecting such criteria are reviewed and analyzed. The issue of Web services' sensitivity evaluation is studied in detail. The concept of sensitivity classes for web services is introduced. We suggest a routine based on machine learning techniques for isolation and automated class definition. We also present the results of a computer simulation that implements the suggested routine. In conclusion we outline promising lines of research.

Keywords: web-services, service-oriented architecture, machine learning, sensitivity analysis.

Статья представлена к публикации членом редакционной коллегии М. В. Губко