**References**

1. Vodnyj kodeks Ukrainy (2004). Kiev, Vidavnichij Dim “In Jure”, 138.

2. Instrukcija pro porjadok rozrobky ta zatverdzhennja granychno-dopustymyh skydiv (GDS) rechovyn u vodni ob'jekty iz zvorotnymy vodamy (1994). Minpryrody Ukrainy. Kyiv, 89.

3. Vpemennye metodicheskie ukazanija po ppovedeniju paschetov fonovyh koncentpacij himicheskih veshhestv v vode vodotokov (1983). Lviv: Gidpometeoizdat, 52.

4. SNA: Climate, Lakes and Rivers (1995). The National Atlas of Sweden, Almqvist and Wiksell International, Stockholm.

5. Storm, B., Refsgaard, A.; in: Abbott, M. B., Refsgaard, J. Ch. (Eds.) (1996). Distributed physically based modelling of the entire land phase of the hydrological cycle. Distributed Hydrological Modelling, Water Science and Technology Library, 22.

6. Karaushev, A. V. (Ed.) (1987). Metodicheskie osnovy ocenki i reglamentirovanija antropogennogo vlijanija na kachestvo poverhnevih vod. Lvov: Gidrometeoizdat, 285.

7. Otchet o NIR «Razrabotka normativov PDS i VSS veshhestv v vodnyj ob’ekt so stochnymi vodami dlja KP «Luckvodokanal» i predlozhenij k planu meroprijatij po dostizheniju PDS» (2011). Kharkov, UkrNIIJeP, 41.

8. Obobshhennyj perechen' predel'no dopustimyh koncentracij (PDK) i orientirovochno bezopasnyh urovnej vozdejstvija (OBUV) vrednyh veshhestv dlja vody rybohozjajstvennyh vodoemov (1990). Minrybhoz SSSR. Moscow, 44.

9. Proskurnin, O. A. (2012). Problemy normirovanija vodootvedenija stochnyh vod v vodotoki v sluchae nepolnogo razbavlenija. Ekologichna bezpeka: problemi i shljahi virishennja: materiali. Kharkiv, VD „Rajder”, 228–234.

10. Sivapalan, M., Takeuchi, K., Franks, S., Schertzer, D., O’Connell, P. E., Gupta, V. K., McDonnell, J. J., Pomeroy, J. W., Uhlenbrook, S., Zehe, E., Lakshmi, V. (2003). IAHS Science Decade on Prediction in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. Hydrololgoical Science Journal, 48 (6), 857–880.doi: 10.1623/hysj.48.6.857.51421

*Дата надходження рукопису 20.01.2015*

**Адаменко Николай Игоревич**, доктор технических наук, доцент, кафедра теоритической ядерной физики и высшей математики им. О. И. Ахиезера, Харковский национальный университет им. В. Н. Каразина, пл. Свободы, 4, г. Харьков, Украина, 61022
E-mail: nikolajadamenko@mail.ru

**Проскурнин Олег Аскольдович**, кандидат технических наук, старший научный сотрудник, Лаборатория проблем формирования, регулирования качества вод и информационного обеспечения экологического менеджмента, НДУ «Украинский научно-исследовательский институт экологических проблем», ул. Бакулина, 6, г. Харьков, Украина, 61166
E-mail: oaproskurnin@mail.ru

**УДК 004.4**
**DOI: 10.15587/2313-8416.2015.37461**

# ТЕХНИКА ОБНАРУЖЕНИЯ АНОМАЛИЙ В ПРОИЗВОДИТЕЛЬНОСТИ ВЕБ-ПРИЛОЖЕНИЙ С ИСПОЛЬЗОВАНИЕМ КОЭФФИЦИЕНТА РАНГОВОЙ КОРРЕЛЯЦИИ КЕНДАЛЛА

**© А. А. Сытник**

*В данной статье представлена техника обнаружения аномалий в производительности веб-приложений с использованием коэффициента ранговой корреляции Кендалла. Описаны теоретические этапы и проведено имитационное моделирование по определению аномалий в производительности веб-приложения. Данная техника дает возможность обнаружить аномалию производительности ВП, на основе корреляционной связи между величинами, но она не даст информации о том, где именно в исходном коде аномалия возникла и по какой причине*
***Ключевые слова:*** *обнаружение аномалий, веб-приложения, коэффициента ранговой корреляции Кендалла*

*This article presents the anomaly detection technique in web applications performance using Kendall’s rank correlation coefficient. Theoretical stages are described and simulation modeling to detect such anomalies in web applications performance is conducted. This technique makes possible to detect performance anomaly for web applications, based on correlation relationships between variables, but it doesn’t give any details on where exactly the anomaly occurred in the source code and why*
***Keywords:*** *anomaly detection, web applications, Kendall’s rank correlation coefficient*

## 1. Введение

Веб-приложения (ВП) подпадают под класс критически важных бизнес приложений. Они используются в разных организациях как часть бизнес процесса, поэтому сценарии спада производительности, аномалии в работе и недоступности приложения негативно влияют на качестве предоставляемых услуг. Под аномалиями подразумевают закономерности в работе модулей, которые не вписываются в нормальное поведение приложения.

Актуальной является задача своевременного обнаружения и оповещения об аномалиях производительности ВП. Применение методов статистического анализа [1] могут значительно облегчить эту задачу. К одному из таких методов относится определение коэффициента корреляции между двумя величинами. Коэффициенты корреляции используют-ся для определения силы и направления связи между двумя свойствами, измеренными в числовых шкалах (метрических или ранговых). Максимальной силе связи соответствуют значения корреляции +1 (строгая прямая или прямо пропорциональная связь) и –1 (строгая обратная или обратно пропорцио-нальная связь), отсутствию связи соответствует корреляция, равная нулю. В случае с ВП, если сила корреляционной связи между двумя величинами слабая, то существует вероятность деградации производительности приложения и увеличение времени отклика для конечного пользователя.

**2. Анализ литературы**

Фундаментальные труды по определению коэффициента ранговой корреляции (КРК) принадлежат Пирсону [2], Спирмену [3] и Кендаллу [4]. В работе [5] авторы разработали технику, которая позволяет обнаружить аномалии производительности выполняя корреляционный анализ между параметрами приложения, собранных с помощью аспектно-ориентированного программирования [6]. В другой работе [7] авторы изолируют изменения рабочей нагрузки от аномалий производительности путем комбинирования регрессионной модели для транзакций и цифровой подписи приложения. Это необходимо для того, что определить какая транзакция ассоциируется с внезапным повышением нагрузки на центральный процессор (ЦП). Метод. представленный в работе [8], описывает адаптируемую модель мониторинга, которая позволяет без изменений в исходном коде приложения, обнаружить и проанализировать пути прохождения транзакции для отладки производительности между компонентами.

**3. Цель работы**

Разработать и описать технику обнаружения аномалий в производительности ВП с применением коэффициента ранговой корреляции Кендалла.

**4. Задачи исследования**

Для достижения цели были поставлены следующие задачи:

1. Проанализировать преимущества применения КРК Кендалла для обнаружения аномалий в производительности ВП.

2. Описать механизм применения коэффициента ранговой корреляции для обнаружения аномалий в производительности ВП.

3. На основе имитационной модели протестировать разработанную технику и проанализировать полученные результаты.

**5. Преимущества использования коэффициента ранговой корреляции Кендалла для обнаружения аномалий производительность веб-приложений**

Для определения преимуществ использования коэффициента Кендалла, выделим слабые стороны других методов ранговой корреляции при обнаружении аномалий производительности ВП.

КРК Пирсона позволяет определить насколько пропорциональна изменчивость двух переменных и характеризует существование линейной связи между двумя величинами. Слабыми сторонами линейного коэффициента корреляции Пирсона являются: исследуемые переменные $X$ и $Y$ должны быть распределены нормально, неустойчивость к выбросам (один или несколько результатов наблюдений резко выделяются, при сравнении с основной массой данных). В случаем с веб-приложениями, закон распределения исследуемых величин зачастую не известен. Также возможны резкие перепады в результирующих значениях.

При условии отсутствия повторяющихся связей в рангах, по той и другой переменной, формула Пирсона может быть упрощена и преобразована в формулу Спирмена.

Коэффициент ранговой корреляции Спирмена определяет фактическую степень связи между двумя количественными рядами изучаемых признаков и дает оценку тесноты установленной связи с помощью количественно выраженного коэффициента. Недостатком ранговой корреляции Спирмена выделяют неточные значения при большом количестве одинаковых рангов по одной или обеим сопоставляемым переменным, что ограничивает возможности исследования величин, полученных в результате мониторинга ВП.

Альтернативу корреляции Спирмена и Пирсона для рангов представляет корреляция Кендалла. В основе корреляции, лежит идея о том, что о направлении связи можно судить, попарно сравнивая между собой испытуемых: если у пары испытуемых изменение по $X$ совпадает по направлению с изменением по $Y$, то это свидетельствует о положительной связи, если не совпадает – то об отрицательной связи. Преимущество использования корреляции Кендалла заключается в более точной оценке значений при большом количестве одинаковых рангов. Коэффициент Кендалла независим от закона распределения измеряемых величин, что является оптимальным вариантом для исследования аномалий в ВП.

Также, коэффициент ранговой корреляции Кендалла целесообразно применять при наличии небольшого количества наблюдений, что положительно влияет на производительность ВП в момент анализа данных мониторинга в режиме реального времени.

**6. Механизм применения коэффициента корреляции Кендалла для обнаружения аномалий в производительности веб-приложений**

Коэффициент ранговой корреляции Кендалла это мера линейной связи между случайными величинами. В случае с ВП эти величины могут быть

получены посредством мониторинга компонентов приложения, а именно: времени отклика сервера на пользовательские транзакции и количество транзакций, обработанных за этот интервал времени.

Коэффициент корреляции Кендалла представлен формулой (1).

$$\tau = \frac{P(p) - P(q)}{N\frac{(N-1)}{2}} \quad (1)$$

и упрощенной формулой (2)

$$\tau = \frac{4P}{N(N-1)} - 1, \quad (2)$$

где $P(p)$ – число совпадений; $P(q)$ – число инверсий; $N$ – объем выборки.

Механизм обнаружения аномалий в призводительности ВП с использованием КРК Кендалла состоит из следующих этапов:

1. Мониторинг ВП и формирования вектора $X$ из последовательности величин суммарного времени отклика по транзакции пользователя и вектора $Y$, величины которого, равны количеству пользовательских транзакций обработанных за этот период.

2. Сортировка вектора $X$ в порядке убывания.

3. Вычисление числа совпадений $P$.

4. Для этого по переменной $Y$ вычтем ранг текущего испытуемого с рангом испытуемого находящего ниже на одну строку. Положительная разность и будет отражать число совпадающих рангов. Отрицательная разность приравнивается к нулю, поскольку в том случае совпадающих рангов нет.

5. Расчет коэффициента корреляции $\tau$-Кендалла используя упрощенную формулу (2).

6. Определение уровня значимости коэффициента.

Для того чтобы при уровне значимости $\alpha$ проверить нулевую гипотезу о равенстве нулю генерального коэффициента ранговой корреляции Кендалла при конкурирующей гипотезе $H_1 : \tau \neq 0$, надо вычислить критическую точку:

$$T_{kp} = z_{kp}\sqrt{\frac{2(2n+5)}{9n(n-1)}}, \quad (3)$$

где $n$ – объем выборки; $z_{kp}$ – критическая точка двусторонней критической области, которую находят по таблице функции Лапласа по равенству:

$$\Phi(Z_{kp}) = \frac{(1-\alpha)}{2}. \quad (4)$$

1. Сопоставление значения коэффициента ранговой корреляции $\tau$-Кендалла и уровня его значимости.

Если $|\tau| < T_{kp}$ – нет оснований отвергнуть нулевую гипотезу. Ранговая корреляционная связь между качественными признаками незначима. Если $|\tau| > T_{kp}$ – нулевую гипотезу отвергают. Между качественными признаками существует значимая ранговая корреляционная связь.

2. Оповещение администратора ВП об аномалии если, если сила корреляционной связи слабая (<0,3);

На практике, если взаимосвязь между суммарным временем отклика и количеством обработанных транзакций пользователя стабильно удерживается, то корреляционная связь высокая. Ослабевание силы связей между суммарным време-нем отклика и количеством обработанных транза-кций пользователя, дает основание предполагать о возможной аномалии в производительности ВП.

**7. Имитационное моделирование. Анализ результатов**

Имитационная модель состоит из нескольких уровней инфраструктуры, а именно:

- веб-приложения JPetStore6 [9], которое реализует прототип электронного магазина с основными функциями покупки-продажи товаров;
- нагрузочного приложения TPC-W [10], основная задача которого заключается в имитации активности в магазине. Приложение позволяет эмулировать несколько одновременных сессий пользователей, покупку разных товаров в один момент времени и тд;
- базы данных MySQL [11] для хранения данных мониторинга;
- инструменты мониторинга на уровне приложения, а именно: jeeObserver [12], InfraRED [13], JavaMelody [14]. С их помощью были собраны метрики необходимые для анализа;
- модуль анализа данных мониторинга с помощью КРК Кендалла. Цель данного модуля заключается в определении аномалии производительности ВП, на основе оценки силы корреляционной связи между значениями;
- модуль электронного оповещения администратора ВП в случае возникновения аномалии.

Имитационное моделирование проходило по следующему сценарию: с помощью нагрузочного приложения TPC-W была с эмулирована активность на веб-приложении JPetStore6, далее с помощью инструментов мониторинга jeeObserver, InfraRED, JavaMelody были собраны метрики необходимые для анализа и сохранены в БД. После этого данные мониторинга были проанализированы и определены силы корреляционной связи между значениями, на основе которых, были детерминированы аномалии в производительности ВП.

Имитационное моделирование учитывает разное время отклика сервера на пользовательские транзакции и разное количество пользовательских транзакций. Для анализа были использованы 10 пар значений $X$ и $Y$ с учетом увеличения задержки времени отклика сервера на данные транзакции.

В табл. 1, 2 представлены фактические значения величин, полученные в результате мониторинга, а также предварительный анализ входящих данных для определения КРК Кендалла. Рассмотрим данные результаты подробнее, используя технику обнаружения аномалий, изложенную выше.

По формуле (2) вычислим коэффициент ранговой корреляции $\tau$ -Кендалла:

$$\tau = 0.82$$

Далее нужно определить уровень значимости полученного коэффициента, для этого необходимо вычислить критическую точку, по формулам (4,3), где

$$\Phi\left(Z_{kp}\right) = 0,3,$$

$$T_{kp} = 0,21.$$

Так как $|\tau| > T_{kp}$ – ранговая корреляционная связь между оценками по двум тестам значимая. На основании этого делаем вывод, что аномалии в производительности не обнаружены.

Однако, в табл. 2 приведены величины, между которыми корреляционная связь не значимая, что дает основания полагать о возникновении аномалии в производительности ВП.

Таблица 1

Результаты моделирования при сильной корреляционной связи

| Задержка времени отклика (мсек), $X$ | Количество обработанных транзакций, $Y$ | ранг $X$, $d_x$ | ранг $Y$, $d_y$ | $P$ |
|---|---|---|---|---|
| 200 | 10 | 1 | 1 | 9 |
| 265 | 12 | 2 | 2 | 8 |
| 287 | 13 | 3 | 3 | 7 |
| 345 | 14 | 4 | 4 | 6 |
| 354 | 17 | 5 | 6 | 4 |
| 365 | 23 | 6 | 9 | 1 |
| 458 | 16 | 7 | 5 | 3 |
| 654 | 18 | 8 | 7 | 2 |
| 743 | 19 | 9 | 8 | 1 |
| 873 | 30 | 10 | 10 | 0 |
| | | | | 41 |

Таблица 2

Результаты моделирования при слабой корреляционной связи

| Задержка времени отклика (мсек), $X$ | Количество обработанных транзакций, $Y$ | ранг $X$, $d_x$ | ранг $Y$, $d_y$ | $P$ |
|---|---|---|---|---|
| 654 | 18 | 1 | 9 | 1 |
| 765 | 10 | 2 | 5 | 4 |
| 1265 | 12 | 3 | 7 | 2 |
| 1345 | 6 | 4 | 3 | 4 |
| 1361 | 4 | 5 | 1 | 5 |
| 1365 | 5 | 6 | 2 | 4 |
| 1542 | 8 | 7 | 4 | 3 |
| 1845 | 10 | 8 | 6 | 2 |
| 2776 | 29 | 9 | 10 | 0 |
| 3458 | 16 | 10 | 8 | 0 |
| | | | | 25 |

Также, как и в первом случае вычислим $\tau$ -Кендалла, который равен:

$$\tau = 0.11$$

И определим уровень значимости используя формулы (3,4):

$$\Phi\left(Z_{kp}\right) = 0,3,$$

$$T_{kp} = 0,21$$

Так как $|\tau| < T_{kp}$ – ранговая корреляционная связь между оценками по двум тестам незначимая. В данном случае корреляционная связь слабая, а значит существует вероятность аномалии в производительности приложения.

Анализируя величины с табл. 2, прослеживается взаимосвязь между задержкой времени отклика и уменьшением количества обработанных транзакций, что ведет к ослаблению корреляционной связи и возникновению аномалии в производительности. Аномалии обычно проявляются деградацией производительснти для конечного пользователя и увеличивают время на выполнение заданых операций, по этому своевременное обнаружения и оповещение администратора ВП крайне важно.

**8. Выводы**

В данной статье описана и проанализирована техника обнаружения аномалий в производи-тельности ВП с помощью коэффициента ранговой корреляции $\tau$ -Кендалла. Для анализа были взяты значения двух

метрик: суммарного времени отклика по транзакции пользователя и количества обработанных пользовательских транзакций за этот период. Далее с помощью КРК Кендалла была определенна сила корреляционной связи между данными величинами с целью детерминировать аномалию в производительности ВП. Если сила связи слабая, то высока вероятность аномалии в производительности ВП, что также подтверждается имитационном моделированием. Данное утверждение справедливо и для данных с табл. 2, где с увеличением времени отклика и уменьшением количества обработанных транзакций сила корреляционной связи слабеет и увеличивается вероятность возникновения аномалии. Техника, описанная в данной статье, может помочь обнаружить аномалию производительности ВП по факту возникновения, но она не даст информации о том, где именно в исходном коде аномалия возникла и по какой причине.

**Литература**

1. Thomas, P. R. Modern engineering statistics [Text] / P. R. Thomas. – Wiley-Interscience, 1st edition, 2007. – 736 p. doi: 10.1002/9780470128442

2. Benesty, A. A. Pearson Correlation Coefficient [Text] / A. Benesty, A. Chen. – Springer Berlin Heidelberg, 2009. –326 p.

3. Jerrold, H. Z. Significance Testing of the Spearman Rank Correlation Coefficient [Text] / H. Z. Jerrold // Journal of the American Statistical Association, 1972. –Vol. 67, Issue 339. – P. 578–580. doi: 10.2307/2284441

4. Харченко, М. А. Корреляционный анализ [Text] / М. А. Харченко, – Учебное пособие для вузов. – Воронеж: Изд-во ВГУ, 2008. – 31 с.

5. Magalhaes, J. P. Anomaly Detection Techniques for Web-Based Applications: An Experimental Study [Text] / J. P. Magalhaes, L. M. Silva // 11th IEEE International Symposium on Network Computing and Applications, 2012. – P. 181–190. doi: 10.1109/nca.2012.27

6. Kiczales, G. Aspect-Oriented Programming [Text] / G. Kiczales, J. Lamping, A. Mendhekar, C. Maeda, C. L. Videira, J. M. Loingtier, J. Irwin // In Proceedings of the 11th European Conference on Object Oriented Programming, 1997. – P. 220–242. doi: 10.1007/bfb0053381

7. Cherkasova, L. Anomaly? Application Change? or Workload Change? Towards Automated Detection of Application Performance Anomaly and Change [Text] / L. Cherkasova, K. M. Ozonat, M. Ningfang, J. Symons, E. Smirni // In Proceedings of the International Conference on Dependable Systems and Networks, 2008. – P. 452–461. doi: 10.1109/dsn.2008.4630116

8. Aguilera, M. K. Performance debugging for distributed systems of black boxes [Text] / M. K Aguilera, J. C. Mogul, J. L. Wiener, P. Reynolds and A. Muthitacharoen, – In Proceedings of the nineteenth ACM symposium on Operating Systems Principles, 2003. –74–89.

9. MyBatis – MyBatis-Spring Sample Code [Electronic resource] / Available at: https://mybatis.github.io/spring/sample.html – 18.01.2015. – Загл. с экрана.

10. TPC-W – Benchmarking an Ecommerce Solution [Electronic resource] / Available at: http://www.tpc.org/tpcw/ – 18.01.2015. — Загл. с экрана.

11. MySQL: The World's Most Popular Open Source Database [Electronic resource] / Available at: http://www.mysql.com/ – 18.01.2015. – Загл. с экрана.

12. jeeObserver. J2EE performance monitoring tool [Electronic resource] / Available at: http://www.jeeobserver.com – 18.01.2015. – Загл. с экрана.

13. InfraRED. Opensource J2EE Performance Monitoring Tool. [Electronic resource] / Available at: http://infrared.sourceforge.net/ – 18.01.2015. – Загл. с экрана.

14. JavaMelody. Monitoring of JavaEE applications [Electronic resource] / Available at: http://code.google.com/p/javamelody/ – 18.01.2015 г. – Загл. с экрана.

**Refernces**

1. Thomas, P. R. (2007). Modern engineering statistics.Wiley-Interscience, 1st edition, 736. doi: 10.1002/9780470128442

2. Benesty, A. A. (2009). Pearson Correlation Coefficient. Springer Berlin Heidelberg, 326.

3. Jerrold, H. Z. (1972). Significance Testing of the Spearman Rank Correlation Coefficient. Journal of the American Statistical Association, 67 (339), 578–580. doi: 10.2307/2284441

4. Harchenko, M. A. (2008). Correlation analiz. Manual for students, 31.

5. Magalhaes, J. P., Silva, L. M. (2012). Anomaly Detection Techniques for Web-Based Applications: An Experimental Study. In Proceedings of the 11th IEEE International Symposium on Network Computing and Applications, 181–190. doi: 10.1109/nca.2012.27

6. Kiczales, G., Lamping, J., Mendhekar, A., Maeda, C., Videira, C. L., Loingtier, J. M., Irwin, J. (1997). Aspect-Oriented Programming. In Proceedings of the 11th European Conference on Object Oriented Programming, 220–242. doi: 10.1007/bfb0053381

7. Cherkasova, L., Ozonat, K. M., Ningfang, M., Symons, J., Smirni, E. (2008). Anomaly? Application Change? or Workload Change? Towards Automated Detection of Application Performance Anomaly and Change. In Proceedings of the International Conference on Dependable Systems and Networks, 452–461. doi: 10.1109/dsn.2008.4630116

8. Aguilera, M. K., Mogul, J. C., Wiener, J. L., Reynolds, P., Muthitacharoen, A. (2003). Performance debugging for distributed systems of black boxes. In Proceedings of the nineteenth ACM symposium on Operating Systems Principles, 74–89.

9. MyBatis – MyBatis-Spring Sample Code. Available at: https://mybatis.github.io/spring/sample.html

10. TPC-W: Benchmarking an Ecommerce Solution. Available at: http://www.tpc.org/tpcw/

11. MySQL : The World's Most Popular Open Source Database. Available at: http://www.mysql.com/

12. jeeObserver. J2EE performance monitoring tool. Available at: http://www.jeeobserver.com/

13. InfraRED. Opensource J2EE Performance Monitoring Tool. Available at: http://infrared.sourceforge.net/

14. JavaMelody. Monitoring of JavaEE applications. Available at: http://code.google.com/p/javamelody/

*Рекомендовано до публікації д-р техн. наук Первунінський С. М.*
*Дата надходження рукопису 27.01.2015*

**Сытник Антон Александрович**, аспирант, кафедра программного обеспечения автоматизированных систем, Черкасский государственный технологический университет, бул. Шевченка, 460, г. Черкассы, Украина, 18000,
E-mail: sytnik.anton@gmail.com