METRICS DATA ANALYSIS FOR WEB SERVICE SELECTION

Hristian Dimitrov, Olga Georgieva

Faculty of Mathematics and Informatics
Sofia University "St. Kliment Ohridski"
Sofia, Bulgaria
hristian@fmi.uni-sofia.bg, o.georgieva@fmi.uni-sofia.bg

***Abstract**: The paper presents three different approaches of web service assessment based on metrics data of their qualitative features. The advantages and drawbacks of the approaches are analyzed for the purposes of service selection. The investigation of assessment results of real time data of two meteorological forecasting services are considered as a solid base to reveal ready-made solutions for a reliable decision support tool.*

***Keywords**: web service quality, metrics data, data analysis, web service selection, meteorological forecasting services*

1. Introduction

Nowadays, software systems providing functionality ensured by other services are increasingly available. Basically these solutions allow re-use of ready-made services that are well established and popular in their application. The open question is how to choose an appropriate service among the others with same or similar functionalities so to deliver the best result for the consumers. An accurate response needs time, people and resources for data analysis and for consequent argued selection of appropriate service. Most organizations or particular clients would not devote such resources to addressing this task.

Service selection among services providing the same functionality could be based on analysis of their qualities (noted as QoS selection) using a proper software tool. For such tool the analysis of metrics data of the service qualities is crucial [1,2]. The elaborated solution has to be able to provide generalized service estimation based on all interested quality features. The overall assessment of QoS is important and difficult, as some qualities have the opposite effect to others in shaping the overall behavior of the service.

The paper goal is to present different assessment approaches for web service selection based on metrics data of their qualitative features. Analysis of the advantages and drawbacks and assessment results of their application towards real time services' data are investigated and underlined. The work is considered as a solid base to reveal ready-made solutions for building a reliable decision support tool of QoS selection.

2. Software Services Quality Assessments

Web services are a widespread approach to providing functionalities that are available from any other system most often through the request-response mechanism. They have a standardized way of exchanging data through the Internet. Product-oriented non-functional features are of a special attention as these represent the quality of the need communication. Appropriate metrics data for quality assessment algorithms is important stage of the web service selection. The most frequently used qualities are:

- Response time (RT) – time in seconds or milliseconds from the sending of the request until the response arrives
- Throughput (T) – the frequency of data transfer in bytes or kilobytes per second
- Number of Completed requests (CR) in requests per second.

However, metrics related to measurement of other web services features are on the focus. Thus, the most contemporary used web browsers follow several different times that characterize the services [10]:

- Initial connection time
- Time of SSL (Secure Sockets Layer)
- Time to send a request
- Time to first byte (TTFB)
- Time to download response content

The dynamic environment in which services operate needs to apply temporal data analysis where not only current quality data but also past data are taken into account for assessing the web service [3]. Relying on the fact that the history log file could accumulate metrics data of several qualities in time then a reliable, accurate and dynamic service selection solution could be searched by a mechanism assessing the integrated quality of a service within certain time period. Based on this idea different methods exist. Three different approaches are explained further and discussed in the sense to be incorporated in a software tool for generalized QoS estimation.

2.1. Statistical Analysis of Metrics Data

Statistical analysis can be applied separately to each of the quality metrics data as a powerful tool for analysis of the available data in the history log file. Information about qualities such as minimum value, maximum value, average, dispersion as well as first, second (median), third quartiles can provide essential knowledge about the service being evaluated. Successful and unsuccessful queries as percentage values are also valuable. Quadrant information is preferred because it responds much less than outliers data, unlike average and dispersion. Providing this information the client has a clear view about the levels and dynamic changes of each of the interested feature of the services being investigated.

A drawback of this approach for service evaluation is that assessments reveal qualities separately. The approach has not a mechanism to account for the quality interdependences and their contradictory effects.

2.2. Apdex Index

Another important aspect of QoS assessment is also the account of users' opinion of the system in respect to the studied characteristics. *Apdex* (Application Performance Index) is a standardized method developed by a group of companies that serves for reporting, measuring, and tracking the performance of web applications. The method transforms the service measurements to one value index over a given time interval. The index belongs to the range 0 to 1, with the value 0 of not satisfied by the quality of the system under consideration with respect to this characteristic, and at value 1 if it is fully satisfied [4].

In this paper we apply *Apdex* to evaluate the preference for maximum response time of the service requests. For this, it is necessary to determine in which area each of the measured response time data belong. Three zones - Satisfied, Tolerating, Frustrated are defined. Each zone corresponds to user satisfaction through this metric. It is calculated according to a predefined threshold value T presented in Fig.1.

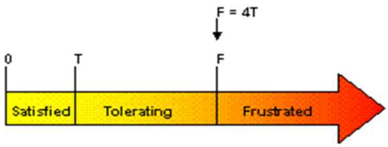


Fig. 1. The three zones of Apdex index [4] (http://apdex.org/overview.html)

The *Apdex* is calculated by the number of satisfied samples ($C_{Satisfied}$), half of the tolerating samples ($C_{tolerating}$) and by the total number the samples (C_{Total}) as follows:

$$Apdex_T = \frac{C_{Satissfied} + \dfrac{C_{Tolerating}}{2}}{C_{Total}}. \qquad (1)$$

The $Apdex_T$ value (1) further is classified according to the scale created for this metric as a way to evaluate the quality of the web application against its Response time. According to it at values above 0.94 the web application is classified as excellent. The problem with working with this metric is to find an appropriate *T* value. It is most determined by the user personal preferences and understanding of the speed with which he wants to return a response to a query. Based on experiments with 500 independent web applications [5] it is found that using the value of the 90th percentile of response time data there is 95% probability of *Apdex* ranging from 0.91

to 0.96. This can be used as the starting point for evaluating a web service against the response time for requests. In addition, the average value for the *Apdex* score is also evaluated and provided as a valuable service measure.

The *Apdex* reflects the user preferences according to the requested property. However, it does not account for other interested qualities.

2.3. Integrated Quality Assessment

The criticism of the methods presented above is in the fact that the extreme value of each quality is determined independently from the values of other qualities. The service overall appearance is estimated based on the individual extremes not accounting for their interdependence. In answering to this problem we apply a newly proposed method based on cluster analysis, which was reported as a successful approach for integrated service quality assessment [7].

Cluster analysis is well known unsupervised method that defines specific data groups in a multidimensional data space. In our case each dimension corresponds to an interested quality. Each cluster comprises service data that present services having similar QoS. The method explores Subtractive clustering (SC) algorithm [6]. The method does not necessarily apply the whole SC algorithm as we are not interested in all clusters of the data sets. As most of the service data sets do not presents diversity in their data structure it is valuable to define densest cluster of each service set represented by respective cluster center v_j, where j is the index of the service. For a selection purpose we have to compare the cluster centers in terms of their coordinates and respective compactness. The preference is given to a service, which center of the densest cluster optimizes the interested quality values and has most compact cluster.

Consider that we have m different web services S_j, j = 1, …, m performing the same functionality. Let the service quality is characterized by n quality properties x_i, i=1,…, n that are measured for QoS selection. For a certain time window their metrics data were collected in the history log. Each row of a particular data set corresponds to a point of the data space.

Cluster density Dj, j=1,…, m is important index as if the data present a compact and dense data group then the service has a stable and reliable behavior with low noisy data within the evaluated time.

$$D_j = \frac{\sum_{i=1}^{N} d_i}{N}, \qquad (2)$$

Where $d_i = | x_k - v_j |$ is the distance between the point x_k of the data set and the cluster center v_j, j = 1,…, m, N is the number of data in the data space. Most preferable service minimizes the density measure (2).

The other measure of the cluster compactness is the cluster potential that is estimated as:

$$\mathrm{P}_j = \sum_{k=1}^{\mathrm{N}} \mathrm{e}^{-\frac{4}{\mathrm{r_a}^2}\left|\mathrm{x}_k - v_j\right|^2}, \qquad (3)$$

where the radius r_a is a value that specifies a cluster center's range of influence. It defines the amount of neighbors of the cluster as the points outside this range contribute slightly to the potential value. Most preferable service maximizes the potential measure (3).

The three types of service quality assessments presented above could be provided to the client as they reveal different aspects of the services. Careful consideration of specific qualities could be done on the basis of statistical analysis. If more weight to the user preferences should be given then *Apdex* score is very useful. The integrated quality assessment is mostly reliable in sense that all interested qualities are accounted for.

3. QoS Analysis of Meteorological Forecasting Services

In order to estimate the usefulness of the proposed methods we carried out experiments for two services that provide meteorological forecast for the next 5 days from the day of the sent request. The selected web services are *apixu* [8] and *metaweather* [9]. The services are appropriate for experimenting because they are used by various other systems. In addition, their data are periodically changed, and so different dynamics in the information could be monitored.

As often the web services have limitations of the daily or monthly number of requests that can be sent to them free of charge, we restricted up to 5 HTTP connections to web services of the 5-seconds interval, for which the data of the web services is taken. This reduces throughput but still the assessment abilities could be explored.

Particular experiment was conducted on February, 2019 at a time interval of 10 minutes accumulating 120 data. The collected metrics data of Response Time, Requests per seconds of Successful (SR) and Failed Requests (FR) and Received KBps are shown in Fig. 2. There are no unsuccessful queries. Also no deviation of the received kilobytes data per second for both services is observed. It means that the third service feature is not valuable for integrated quality estimation but only respective levels should be compared. After displaying the reference lines of *metaweather* data graph, it can be assumed that most of the data of RT and SR are in the vicinity of the data median. There are only several outliers. On the *apixu* data graphs, more peak values can be identified. This is particularly noticeable on the RT graph. However, it is not possible to clearly identify which service is preferable especially comparing to all monitored features. Further we will consider in detail the quality estimated based on the three considered assessment approaches.

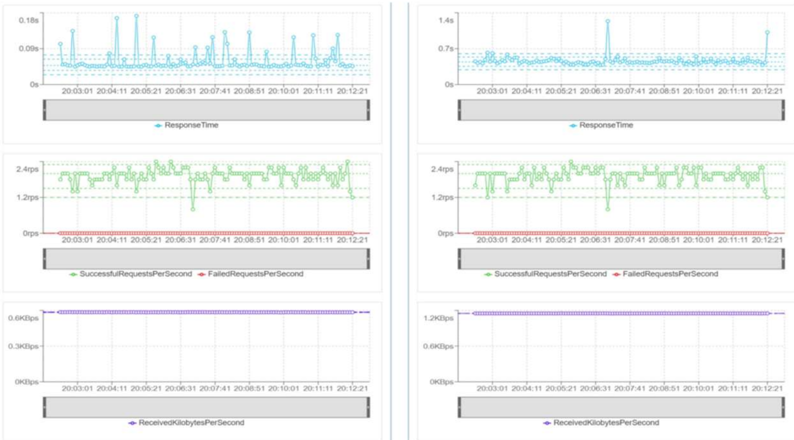
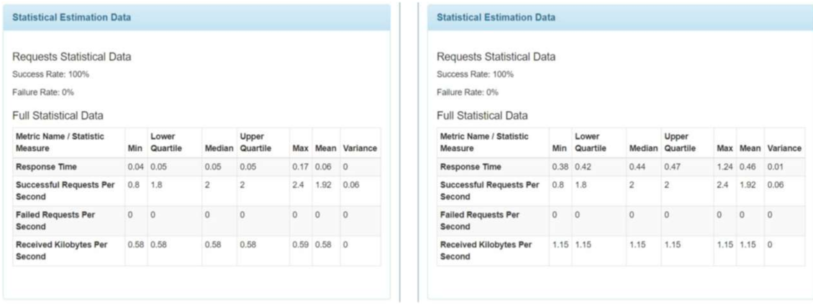


Fig 2. Metrics data of web services apixu (left) and metaweather (right) of first experiment

The statistics analysis data (Fig. 3) confirm that both services have 100% successful queries. The average response time of *apixu* is 0.06 seconds and for *metaweather* service it is much more - 0.46 seconds. The same tendency is maintained for the minimum and maximum value, and the proportion increases to about 8 times difference. It can be concluded that *apixu* requests are received faster than the *metaweather* ones for the examined time interval. In terms of throughput it could be seen that the respective statistical values for each service remain constant as no dynamics of this property is observed. However, comparing the two web services the respective values differ significantly as data from *apixu* arrive at a frequency of about 0.58 Kbps, whereas from *metaweather* approximately twice as much - 1.15 Kbps.

The *Apdex* estimate (Fig. 4) the requested upper Response time limit is not set. Therefore, the 90th percentile of all response time data is used. For the first web service, this value is 0.09 seconds, for the second one - 0.52 seconds. The average value of the *Apdex* of both web services is very similar as for *apixu* it is 93.32% and for *metaweather* - 93.51%. Based on these values, the two web services are classified as *good* services. This means that, according to this metric, the two services are equivalent.


Statistical Estimation Data

Requests Statistical Data

Success Rate: 100%

Failure Rate: 0%

Full Statistical Data

Metric Name / Statistic Measure	Min	Lower Quartile	Median	Upper Quartile	Max	Mean	Variance
Response Time	0.04	0.05	0.05	0.05	0.17	0.06	0
Successful Requests Per Second	0.8	1.8	2	2	2.4	1.92	0.06
Failed Requests Per Second	0	0	0	0	0	0	0
Received Kilobytes Per Second	0.58	0.58	0.58	0.58	0.59	0.58	0

Statistical Estimation Data

Requests Statistical Data

Success Rate: 100%

Failure Rate: 0%

Full Statistical Data

Metric Name / Statistic Measure	Min	Lower Quartile	Median	Upper Quartile	Max	Mean	Variance
Response Time	0.38	0.42	0.44	0.47	1.24	0.46	0.01
Successful Requests Per Second	0.8	1.8	2	2	2.4	1.92	0.06
Failed Requests Per Second	0	0	0	0	0	0	0
Received Kilobytes Per Second	1.15	1.15	1.15	1.15	1.15	1.15	0

Fig. 3 Statistical analysis results of metrics data of apixu (left) and metaweather (right) web service

Fig. 4 Apdex Scoring of apixu (left) and metaweather (right) from first experiment

Integrated assessment provided by the cluster analysis (Fig. 5) shows that only one valuable cluster was identified for both services. In addition, *apixu* has a potential of 115.26, and a *metaweather* of 110.62. This means that the *apixu* cluster contains more data in its vicinity.

To improve the results' readability, the obtained density values are transformed to a growing function by subtracting a unit of its value and then result is converted to a percentage. In this case, a preference service has a higher value on this index. The obtained cluster density of *apixu* is 96.25% and for the other web service it is 94.78%. Therefore, it could be concluded that *apixu* has a denser cluster. Important information can also be derived from the percentage of data that belong to a cluster in the vicinity of r_a=0.5. For the first service the scope of the cluster is 100% and for the second one 99.17% of the data. The conclusion is that *apixu* is the preferred web service as it has more stable behavior in this time interval than *metaweather*.

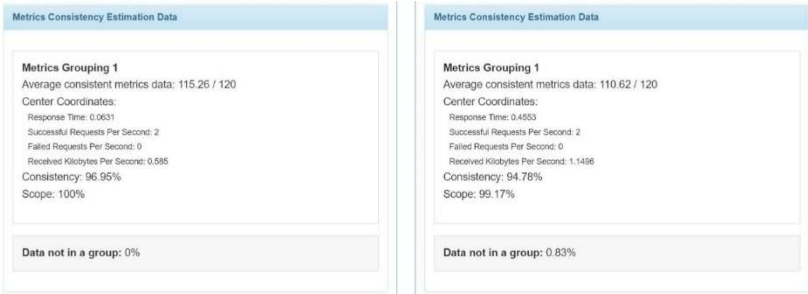


Fig. 5 Indexes of cluster analysis of apixu (left) and metaweather (right) services from first experiment

Conclusion

Three different approaches of web service assessment based on metrics data of their qualitative properties are presented in detail. The analysis of respective advantages and disadvantages for quality of service selection task are revealed. The results by using real-time data of two meteorological forecasting services are discussed as a good basis for further developing ready-made deployment solution. This solution is aimed at creating a reliable tool to help users to make reasonable and optimized decisions which web service to use.

Acknowledgements

This work was supported by the Bulgarian National Scientific Fund within the project no. DN 12/9. The authors also gratefully acknowledge the support provided by project UNITe BG05M2OP001-1.001-0004 funded by the OP "Science and Education for Smart Growth", co-funded by the EU through ESIFs.

Bibliography

1. R.J.R. Raj and T. Sasipraba, 2010, Web service selection based on QoS Constraints, 2nd IEEE International Conference on Trendz in Information Sciences & Computing, 156-162.
2. D. Sachan et al., 2013, A System for Web Service Selection Based on QoS, In Proc. ofInternational Conference on Information Systems and Computer Networks, 139-144.
3. Jinfang Zhang, 2012, A QoS-Aware Computation Model for Dynamic Web Service Selection, In proc. IEEE 12th International Conference on Computer and Information Technology, 230-235.
4. Apdex Overview, APDEX: Measuring What Matters, apdex.org/overview.html (last seen 16.04.2019).
5. Kayser, Bill. “How to Choose Apdex T: The Final Word.” New Relic Blog, 29 June 2017, blog.newrelic.com/product-news/how-to-choose-apdex-t/.

6. Chiu, S., "Fuzzy Model Identification Based on Cluster Estimation," Journal of Intelligent & Fuzzy Systems, Vol. 2, No. 3, Sept. 1994.
7. Olga Georgieva, 2016, Dynamic Service Quality Selection, ACM International Conference Proceeding Series, 18th International Conference on Information Integration and Web-based Applications & Services (iiWAS2016), Nov.28 - 30, Singapore, pp.401-404.
8. Apixu, Mzemo DWC LLC, www.apixu.com/.
9. Cartwright, Jason. MetaWeather, 2017, www.metaweather.com/.
10. Garbee, Jonathan, Understanding Resource Timing. Tool for Web Developers (Google, 12.02.2019, developers.google.com/web/tools/chrome-devtools/network-performance/understanding-resource-timing)

ИЗБОР НА УЕБ УСЛУГИ ЧРЕЗ АНАЛИЗ НА ДАННИ ОТ МЕТРИКИ

Християн Димитров, Олга Георгиева

Резюме: *Представени са три различни подхода със съответни индекси за оценка на уеб услугите, базирани на данни за метрики на техни качествени характеристики. Анализът на предимствата и недостатъците на различните подходи, както и резултатите от приложението им към данни в реално време на две услуги, предоставящи метеорологична прогноза, са добра база за разработване на готови решения за внедряване. Тези решения са насочени към създаване на надежден инструмент за подпомагане на потребителите при вземане на решения за ползване на уеб услуги.*