Deep Learning Inspired Hybrid Meta-heuristic Approach for Intelligent Service Composition in Dynamic Cloud Environment

A V L N Sujith [1,*], Dr. A Rama Mohan Reddy [2] and Dr. K Madhavi [3]

[1]Research Scholar, Department of CSE, JNTUA University, Andhra Pradesh, India;
[2]Professor, Department of CSE, S V University, Andhra Pradesh, India;
[3]Associate Professor, Department of CSE, JNTUA University, Andhra Pradesh, India.

Abstract
Current generation computing scenarios evidenced proliferation of cloud services with the rapid growth in the cloud service consumers. Demand for the utilization of cloud services is constantly increasing with complex service requests from the cloud service consumers that lead to the service composition. Functionally similar cloud services are being offered by different cloud service providers that made the selection of optimal cloud service as an multi-objective optimization problem. In this research study a deep learning inspired meta-heuristic approach is devised that with stand the continuous demands from different consumers in the context of selecting optimal composition pattern in dynamic cloud environment. The experimental studies indicate that the integration of deep learning mechanism with meta-heuristic optimization algorithm depicts better performance results than the existing meta-heuristic approaches of cloud service composition.

Key words:
Cloud Service, Service composition, Service oriented architecture, Deep learning, Gated recurrent unit

1. Introduction

Design and development of the automated and optimized service-oriented architecture are considered as a major aspect during cloud service composition. Recent years have evidenced a large number of cloud service providers offering similar cloud services to the vast range of the customers using various service models that include SaaS (Software as a service), IaaS (Infrastructure as a Service) and PaaS (Platform as a Service) [1]. This scenario has laid a path for the development of different cloud services based on user preference. As a single service will not be able to satisfy the user requirement, the development of a composite service on considering their QoS (Quality of service) parameters may persuade the customer needs.
The development of the meta-services is considered a vital mechanism in cloud service composition. Theses meta-services should include a simple function that executes the complex functionality required for the composite service based on user requirements. In general, it is well known that the cloud environment doesn't include any standalone service. Cloud service providers (CSP) enables a variety of cloud services that differ in configuration and functionality based on the principle of utility computing i.e., the customers will be billed based on the usage of the service as included in the service level agreement (SLA). The proliferation of these meta-services obtained from multiple cloud service providers is considered as a major challenge as these meta-services exhibit similar functional properties but they differ in non-functional parametric aspects. Henceforth selecting an absolute cloud service will be considered as an NP-hard problem for cloud consumers from a set of functionally overlapped cloud services that are being offered in different levels of QoS metrics.
Addressing this problem of service composition several mechanisms have been developed and proposed. Most of these approaches prioritized quality of a particular service during the process of composition but only few studies have concentrated on the concept of dynamic functionality and changes that occur in multi-cloud environment. Most of the SLA related to the cloud service will be leveraging the long term business process where in the requirements of the end users will be keep on changing such that a composite service that is bet fit in on situation will be a worst fit in consequent situation over a period of time due to the dynamic nature of the cloud environment. Therefore we could consider a process of service composition as a long term economic aware business process as a convergence of current level web services [2].
Despite the existing cloud service composition techniques, this paper presents a novel deep learning-based framework in which the QoS attributes of the various candidate cloud services are considered as time-series data through which the correlation among the QoS attributes is balanced to predict the QoS attributes that leverages the cloud services to be flexible and reusable. Furthermore, security remains an open challenge in the context of cloud services composition. Even though several researchers addressed the concept of cloud security, very few research studies have quoted about the formulation of security level agreements, service vulnerability during service composition. Addressing

Manuscript received March 5, 2020
Manuscript revised March 20, 2020

these problems the main contribution of this article includes:

- Deep learning inspired novel cloud service composition framework that integrates Gated Recurrent Unit (GRU) and enhances Particle Swarm Optimization algorithm
- Construction of an economic model to evaluate correlation analysis among QoS attributes.
- Optimized multi-objective cloud service composition using PSO

The rest of the article is organized as follows; existing research studies on the integration of deep learning and optimization theory are discussed in section 2; Section 3 details about the problem formulation; Section 4 discusses about the proposed deep learning inspired service composition framework; Section 5 briefs about the experimental results and finally Section 6 summarizes the conclusion along with the future scope.

2. Related Work

Heuristic algorithms are the algorithms derived from experience with particular optimization problems. Usually, they find an optimal solution by a trial-and-error method in an adequate amount of time by taking full advantage of the particularities of the problem. The solutions by these algorithms may be better then an educated guess but may not be optimal" [4]. Since non-heuristic algorithms take a massive amount of time to procure the optimal solution, heuristic algorithms are preferred because these algorithms can procure near-optimal solutions in an acceptable amount of time. Yuan-sheng et al. [5] introduced an improved heuristic algorithm for selection of service compositions with multiple constraints.
Klein et al. [10] applied the hill-climbing approach and compared it with LIP to minimize the time complexity in finding a near-optimal solution. Qi et al. [11] proposed a local optimization method for finding the local candidate services and then combining them so that the optimal solution is obtained. Jing et al. [12] presented a distributed heuristic approach to solving web service composition problems with a high approximation ratio, enabling adaptability and a near-optimal solution. Ying et al. [13] introduced a win–win strategy approach for obtaining a QoS-aware web service composition. They applied game theory in developing a mathematical model of service composition and a genetic algorithm for service composition with Pareto optimum.
Diana et al. [14] developed two service selection algorithms to satisfy certain QoS requirements based on weight factors and priority factors. Li et al. [15] presented an efficient and reliable method for selection of the reliable services in a QoS-aware web service composition. Luo et al. [6] presented a heuristicenhanced cross-entropy algorithm for QoS-aware web service composition satisfying end-to-end QoS constraints. Seog-Chan et al. [7] proposed a BF ∗-based graph search algorithm to solve the QoS-aware web service composition problem. Niu et al. [8] proposed the heuristic graph search and the "breadth heuristic uncertain composition" approaches for solving uncertain web service composition problems to reduce the search space and for high efficiency in obtaining a service composition solution. Lin et al. [9] discussed a relaxable QoS-based service selection algorithm to attain the optimal solution. This method recommends prospective candidate services to users by relaxing the QoS constraints if no suitable web service can fulfill the user requirements exactly

3. Service Composition to process big data in Dynamic Cloud Environments:

Service computing is considered as the most powerful mechanism in the development of intelligent and automated systems. The most vital concept in service computing is service orchestration such that it enables the construction of the required composite service based on the complex priorities involved in customer preferences [3]. The primary objective of developing a composite service is to satisfy the QoS (Quality of Service) constraints based on customer requirements. The selection of optimal cloud services for composition is considered as a challenging aspect in cloud service composition. In specific, the QoS centric service composition is considered as an NP-Hard problem as it could be formulated based on the multi-choice and multi-objective optimization.

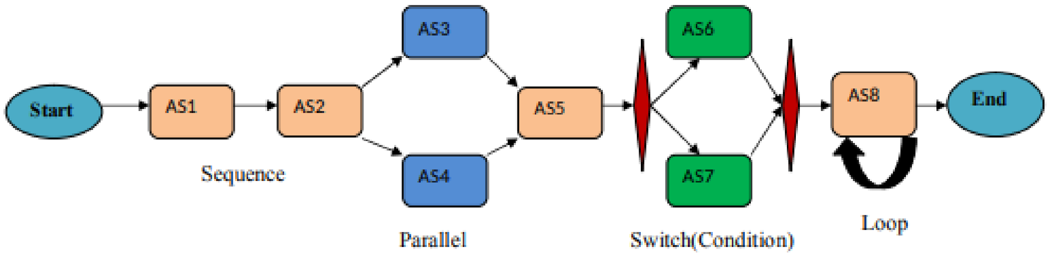

Fig. 1 Service Composition patterns

The meta-service classes MSC are defined as

$$M_{SC} = [CS_1, CS_2, CS_3, \ldots CS_k, \ldots CS_n] \quad (1)$$

In which CS is represents the candidate service and 'n' refers to the total number of the service candidates and CSk refers to a single candidate service as defined in Eq.2

$$CS_k = \{S_1, S_2, S_3, \ldots S_j\} \quad (2)$$

Such that every single cloud service include 'j' number of services that are functionally equivalent. Further the QoS attributes of the services are described as

$$QoS(M_{SC}) = [QoS(CS_1), QoS(CS_2), QoS(CS_3), \ldots QoS(CS_k), \ldots, QoS(CS_n)] \quad (3)$$

The main objective is to devise an optimal service composition approach through based on 'p' QoS where P=3 that includes a cost (Ct), throughput (Tp) and response time (Rt) such that on substituting the QoS parameters considered in the current scenario using Eq.3 QoStot(MSC) is derived as follows

$$QoS_{tot}(M_{SC}) = [QoS_{Rt}(CS), QoS_{Ct}(CS), QoS_{Tp}(CS)] \quad (4)$$

The aggregation functions for various composition patters depicted in figure 1 for various QoS parameters are devised as indicated in table 1.In specific the aggregate functions are applied in a recursive manner while determining the global QoS parameters for various composition patterns.

Table 1: Aggregate functions of QoS attributes for various Composition patterns

QoS attributes	*Parallel*	*Conditional*	*Sequence*	*Loop*
Cost(Ct)	$Agg_{Pc} = \sum_{i=1}^{n} Pc(S_{ij})$	$Agg_{Pc} = PR_i * \sum_{i=1}^{n} Pc(S_{ij})$	$Agg_{Pc} = \sum_{i=1}^{n} Pc(S_{ij})$	$Agg_{Pc} = K * \sum_{i=1}^{n} Pc(S_{ij})$
Throughput(Tp)	$Agg_{Tp} = \min_{i=1}^{n} Tp(S_{ij})$	$Agg_{Tp} = PR_i * \pi_{i=1}^{n} Tp(S_{ij})$	$Agg_{Tp} = \min_{i=1}^{n} Tp(S_{ij})$	$Agg_{Tp} = K * \pi_{i=1}^{n} Tp(S_{ij})$
Response time(Rt)	$Agg_{Rt} = \min_{i=1}^{n} Rt(S_{ij})$	$Agg_{Rt} = PR_i * \sum_{i=1}^{n} Rt(S_{ij})$	$Agg_{Rt} = \sum_{i=1}^{n} Rt(S_{ij})$	$Agg_{Rt} = K * \sum_{i=1}^{n} Rt(S_{ij})$

In this context as the service composition in the cloud is considered as the multi-objective optimization problem the main challenge is to enhance the parameterized vector of throughput and minimize cost and response time. Therefore the parameterized quality vector of the composed service over a period of time series $Q^{CSk}(m)$ related to the throughput, response time and cost are indicated as

$$Q^{CSk}(m) = [\, Q_{rt}(m), Q_{tp}(m), Q_{ct}(m)] \quad (5)$$

The cloud service user indicates his preferences as a service request using a weighted function over the time series 'm' where the weighting attribute Wt (m) is denoted as

$$Wt(m) = [Wt_{rt}(m), Wt_{ct}(m), Wt_{tp}(m)] \quad (6)$$

The functional score F_s is determined by the sum of the products of the QoS parameterized values along with the weights assigned by the consumer on par with the varying time period 'm' is derived as follows

$$F_s(m) = [Wt_{rt}(m)*Q_{rt}(m) + Wt_{ct}(m)*Q_{ct}(m) + Wt_{tp}(m)*Q_{tp}(m)] \quad (7)$$

Henceforth the functional score required to build a composite service over a term of varying time frequency is derived as the summation of multiple QoS values such that

$$F_s(p) = \int_1^M F_s(m)\, dm \quad (8)$$

As the multi-cloud environment includes large number of services the composition plan varies with the dynamic nature of the environment. The main objective of the work is to develop a service composition plan that optimally maximize the value of $F_s(p)$ as it leverages the accuracy of the service composition activity based on the aggregation of real-time QoS values that are utilized in the process.

4. Proposed Mechanism

The exponential growth of the usage of cloud service models enables the researchers in cloud environment to develop an optimized, accurate and expectant economic model that predicts the behavioral patterns of the consumers. In this context this section discuss about a security aware deep learning inspired service composition mechanism for multi-cloud environment that integrates Gated Recurrent Units and Enhances PSO for the optimal selection of the service provider based on the constraints in the SLA(Service Level Agreement). Fig.2 Details the flow chart of the proposed mechanism.

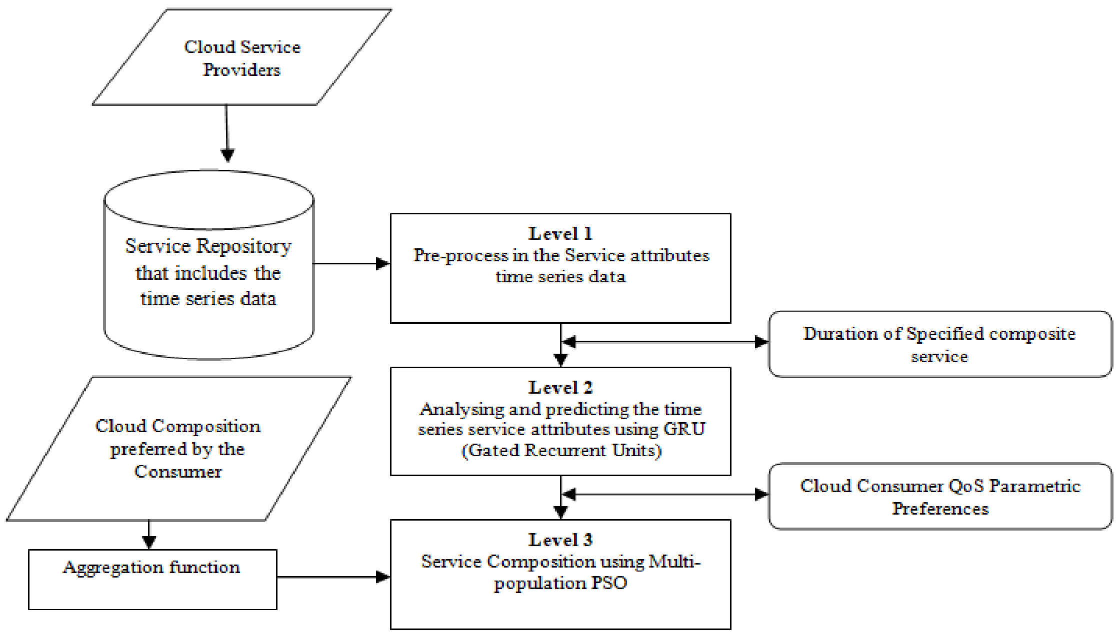

Fig. 2 Proposed Mechanism

4.1 Preprocessing the Service attribute Time Series Data

Time series data set is defined as a data obtained in the chronological order at regular time frequencies in the context of service composition. In general, the data obtained from the service repository is decomposed into three components that include trend, level and noisy data. Trend of the data points indicate the positive and negative trajectory of the data over a specific period of time. Level indicates the mean of the series data points. The variations in the trajectory are indicated through the noise level of the data.
Normalization of the obtained data points is required as they are obtained from dynamic environment in the service repository rescaling od the QoS attributes is considered as the mandatory process. The rescaling of the QoS attributes canbe done using Eq. 9

$$N = \frac{M - min}{max - min} \qquad (9)$$

in the above equation max and min indicates the maximum and minimum values that are obtained from the time series data. ‘N’ and ‘M’ indicates the observations obtained from the newly scaled vector.

4.2 Gated Recurrent Unit

Gated recurrent Unit (GRU) was developed by cho et. al.[ ] and it is widely discriminated as a variant of LSTM (Long Short Term Memory). Initially the GRU is equipped with two different gates update gate and the reset gate as shown in figure 3

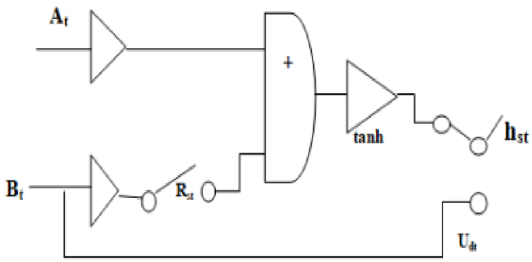

Fig. 3 Gated recurrent unit

Several recurrent neural network models have evidenced that it is very difficult to learn the dependencies that are long term in nature such that as such dependencies may result in the gradient decays over the specific period of time. Henceforth, GRU is considered to be an effective solution as it uses the memory units with reset and update gates that may result in the prediction of hidden layer. The main applications of these gates are that it will control the interaction among the memory units that are adjacent and units within the structure.

During the initial phase the GRU is applied on the lagged values of every QoS Parameters such that the prediction of the noteworthy QoS attributes will be flexible to the process. Then to avoid the context of over fitting the concept of the dropout is applied on the prediction model GRU. Finally fur look ahead (M) and four lagged values (N) of the QoS attributes are utilized.

4.3 Service Composition using Enhanced PSO

The modeling of the composite service is done by using Particle Swarm Optimization technique such that exploration of the best QoS parameters to maximize the optimization level of the composition process could be done by using PSO as shown in below figure 4.

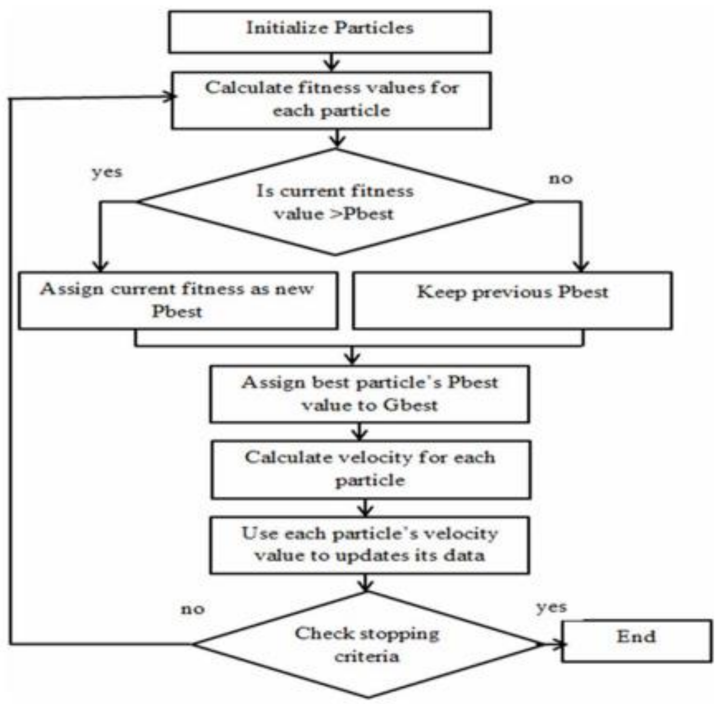

Fig. 4 Enhanced PSO for Service Composition

4.3.1 Deep learning inspired service Composition

Algorithm: Intelligent Cloud Service Composition

i. Initialize the system parameters relative to the QoS attributes of the cloud Services that included the population size, upper and lower bound variables iterations and factors guided by the acceleration parameters

ii. Initial population is generated using Gated recurrent unit through which the candidate services that are required for the service composition are obtained.

a. Set Current_rank = 1
b. M:=(Qos Parametrised values of candidate services)
c. N := M
d. While M!= 0 (/* Preprocess the QoS parameters of the cloud services */)
e. for i=1 to n{
f. if Sc_i is non dominated {
g. rank(Sc_i) = Current_rank
h. }
i. }
j. for i = 1 to n{
k. if rank(Sc_i) = Current_rank{
l. remove Sc_i from the enlisted candidate services
m. M = M-1
n. }
o. }
p. Current rank := current_rank + 1
q. Evaluate the fitness

iii. Evaluate the fitness value of the candidate cloud services using Eq. 8

Fitness Function $F_s(p) = \int_1^M F_s(m)\, dm$

iv. Evaluate the performance of the every candidate service and compare it with the best performance attribute.

5. Performance Evaluation

In this research study an deep learning inspired service composition mechanism is developed to handle the variance in the QoS parameters of the functionally similar clod services in a dynamic multi-cloud environment. The performance of the proposed method is evaluated based on various existing mechanisms in a context of multi object optimization.All the algorithms are developed and implemented in MATLAB R2013a using QWS dataset obtained from the service repository.
The following box plots illustrate the Set Coverage box-plots for 2-objective problem with 50 candidate services having 25, 50, 75, and 100 abstract services. In Fig. 5, each rectangular box represents a set of Set Coverage values for PSO.

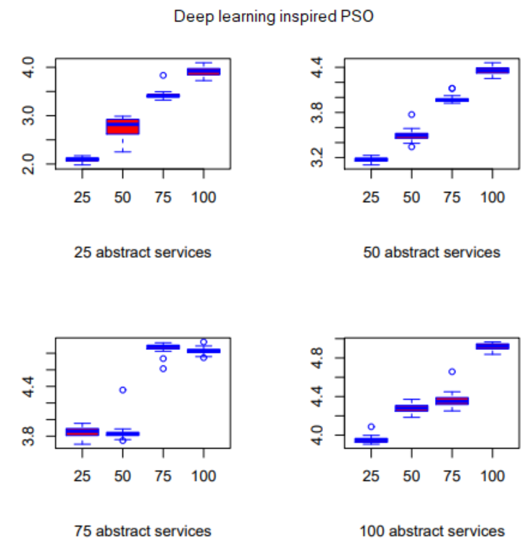

Fig. 5 Box plots for 2-objective optimization problem

Hypervolume box-plots for a 2-objective problem (30 independent runs). A scenario with 25, 50, 75, and 100 abstract services versus 25, 50, 75, and 100 candidate services is considered for Extended PSO algorithm.

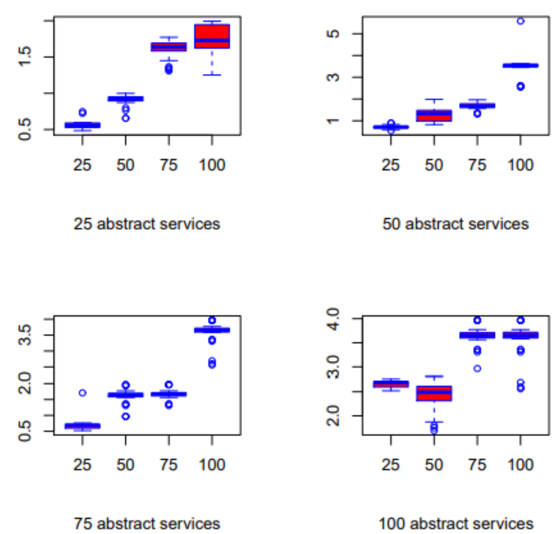

Fig. 6 Box plots for 2-objective optimization problem

The average Hypervolume box-plots for a 3-objective problem (30 independent runs). A scenario with 25, 50, 75, and 100 abstract services versus 25, 50, 75, and 100 candidate services is considered for PSO algorithm.For each algorithm, the blue line inside the box represents median and the box height represent dispersion. Blue circles are the outliers. The algorithm with high median value finds a good approximation set. A low dispersion value, small box and a low number of outliers represent a stable/ robustness algorithm.

6. Conclusion and Future Work

In this research study a deep learning inspired meta-heuristic approach is devised that with stand the continuous demands from different consumers in the context of selecting optimal composition pattern in dynamic cloud environment. The experimental studies indicate that the integration of deep learning mechanism with meta-heuristic optimization algorithm depicts better performance results than the existing meta-heuristic approaches of cloud service composition. As a part of future work Various other Computational Intelligence techniques can be applied for web services, cloud service, and Big service composition. Computational Intelligence techniques can be applied to energy aware service composition in multiple cloud environments and cyber-physical-social systems.

References

[1] Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., Savarese, S.: Social lstm: Human trajectory prediction in crowded spaces. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 961–971 (2016)

[2] Ardagna, D., Casale, G., Ciavotta, M., P´erez, J.F., Wang, W.: Quality-of-service in cloud computing: modeling techniques and their applications. J. Internet Services and Applications 5(1), 11:1–11:17 (2014)

[3] Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation 9(8), 1735–1780 (1997)

[4] Kritikos, K., Pernici, B., Plebani, P., Cappiello, C., Comuzzi, M., Benbernou, S., I, B., Kert´esz, A., Parkin, M., Carro, M.: A survey on service quality description. ACM Comput. Surv. 46(1), 1:1–1:58 (2013)

[5] Lemos, A.L., Daniel, F., Benatallah, B.: Web service composition: A survey of techniques and tools. ACM Comput. Surv. 48(3), 33:1–33:41 (2016)

[6] Liu, S., Wei, Y., Tang, K., Qin, A.K., Yao, X.: Qos-aware long-term based service composition in cloud computing. In: Evolutionary Computation (CEC), 2015 IEEE Congress on. pp. 3362–3369 (2015)

[7] Medjahed, B., Benatallah, B., Bouguettaya, A., Ngu, A.H.H., Elmagarmid, A.K.: Business-to-business interactions: issues and enabling technologies. VLDB J. 12(1), 59–85 (2003)

[8] Zeng L, Benatallah B, Ngu AHH et al (2004) QoS-aware middleware for web services composition. IEEE Trans Softw Eng 30:311–327

[9] Jaeger MC, Muhl G, Golze S (2005) QoS-aware composition of web services: a look at selection algorithms. In: IEEE international conference on web services (ICWS'05), IEEECanfora G,

[10] Di Penta Esposito R, Villani ML (2005) An approach for QoS-aware service composition based on genetic algorithms. In: Proceedings of the 7th annual conference on genetic and evolutionary computation, ACM, pp 1069–1075

[11] Lewis PR, Faniyi F, Bahsoon R, Yao X (2013) Markets and clouds: adaptive and resilient computational resource allocation inspired by economics. In: Suri N, Cabri G (eds) Adaptive, dynamic, and resilient systems. CRC Press, Boca Raton, pp 285–318

[12] Ye Z, Bouguettaya A, Zhou X (2012) QoS-aware cloud service composition based on economic models. In: International conference on service-oriented computing, Springer, pp 111–12.

[13] Haytamy SS, Kholidy HA, Omara FA (2018) ICSD: integrated cloud services dataset. In: Yang A et al (eds) Lecture notes in computer science (including subseries lecture notes in artifcial intelligence and lecture notes in bioinformatics). Springer

[14] Ye Z, Mistry S, Bouguettaya A, Dong H (2016) Long-term QoS-aware cloud service composition using multivariate time series analysis. IEEE Trans Serv Comput 9:382–393

[15] Zheng Z, Wu X, Zhang Y et al (2013) QoS ranking prediction for cloud services. IEEE TransParallel Distrib Syst 24:1213–1222.