CHAPTER 6

6 SCALE TO COMPOSITION FAULT INCLINED (SCFI) & SCALE TO COMPOSITION HALENESS (SCH): DESIGN OF HEURISTIC METRICS TO ASSESS SERVICE COMPOSITION IS FAULT INCLINED OR HALE

6.1 OVERVIEW

QoS (Quality of Service) aware composition of web services is a continuous practice due to the different scope of development needs. Henceforth the current research is moving in a direction to find heuristic solutions based on soft computing techniques. But all these solutions are specific to one or two QoS factors mainly response time and reliability. According to the real-time practices, the QoS assessment by one or two factors is impractical. Moreover, these soft computing approaches are delivering the computational complexity as O(n2), which is due to the magnification of number evolution against the increment in the number of services available to choose. In this context here, we devised a two QoS metrics called Scale to Composition Fault Inclined and Scale to Composition Haleness, which enables to assess the services based on the multiple numbers of QoS metrics and should stabilize the computational complexity to O(n*log(n)). The experiment results are indicating the significance of the proposed model towards scalable and robust QoS- aware service composition.

6.2 PROLOGUE

Web service composition is a process of loosely connecting the existing web services and defines the sequence to execute them [66], which creates value-added composite web services to fulfill the desired task. The task of web service composition involves the search to select optimal services from the service repository; the optimality is a factor that varies from one context of service composition to other. The importance and priority given to the QoS factors of the web services such as Availability, Accessibility, cost, Integrity, throughput, roundtrip time, Reliability, Regulatory and Security, is the definition aspect of service optimality. The response time of an independent service can be optimal. However, the service may fail to fit in composition [66], [67], [18], [48], [49].

The availability of multiple services from different providers, which are compatible and meant to settle the same task, is common in the process of service composition. Choosing one of these services to fulfill the task completion is done by ranking them, which is based on their QoS properties. The prioritizing the specific QoS factor towards ranking the services meant for a task is purely based on the context of the composition. For instance, one composition context may demand the reliability of the service but no matter how much costly it would be, the other context of composition might have a barrier about the service cost rather than other QoS factors, or a context may expect the balance in QoS factors such as cost and reliability. Hence the search for an optimal web service from a given service repository is defined as complex and plays the vital role to determine the scalability and robustness of the service composition strategy. Therefore, the selecting optimal web services towards service composition under user constraints can define as the most challenging task.

The web service composition helps to define the dynamic solution for applications with the combination of independent tasks. Regarding this, the composition process should adapt suitable service among the available, which is based on the QoS factors and their priority order defined by the composition context [68].

6.3 ASSOCIATED WORK

Considerable research towards defining meta-heuristic models for QoS aware Service composition can find in recent literature. Yu et al. [48] used a greedy approach to incorporate QoS features for composite services and applied adaptive strategy to get efficient performance within the minimum search time to find the solution. Xiangbing et al. [49] proposed a web service modeling ontology-based (WSMO) web service composition method to solve QoS based service composition and applied the genetic algorithm for efficient performance within minimum search time to find the optimal solution. Li et al. [69] discussed a novel approach for optimal web service selection by chaos particle swarm optimization. Xiangwei et al. [79] presented discrete particle swarm optimization algorithms and color Petri nets (CPN) in the context of QoS based web service composition. In this approach, first CPN is modeling that describes the multi-attribute, multi-constraint relations between candidate services and applies discrete PSO for legal firing sequence and find the optimal solution [80]. Mao et al. [81] presented different meta-heuristic algorithms (particle swarm optimization, estimation of distribution algorithm, genetic algorithm) for efficient performance in web service composition. Zhao et al. [50] adopted an improved discrete immune optimization method based on PSO for QoS aware web service composition. In this approach, the author used an immune algorithm to improve the local best strategy and applied the particle swarm optimization algorithm to find the global optimization value and reduce the search capability and high scalability.

Parejo et al. [51] presented a hybrid approach GRASP and path re-linking algorithm to optimize the quality of service for QoS aware web service composition at runtime.

Though these models are significant regarding the context obtained, but all of these models are considering one or two QoS factors to assess the scalability of the compositions, which is impractical in reality. Moreover, the computational complexity of service composition is O(n2), since the increment in the number of services per each task and number of services required in a composition is magnifying the evaluation complexity. Hence, we proposed a meta-heuristic model that is also not specific to QoS factors and quantity of factors. The said model assesses the impact of each service by the measuring their Composition Fault Inclined Scope, which is based on multiple QoS factors. In contrast to all of the explored existing models, we devised the set of statistical assessment strategies in our earlier works called Web Service Composition Impact Scale towards Fault Proneness [72] and QoS metrics for robust service composition [73], which are aimed at prediction accuracy and scalability. The observations done through the experiments on these earlier models, we motivated to develop heuristic metrics called Scale to Composition Fault Inclined (SCFI) and Scale to Composition Haleness (SCH) here in this chapter.

6.4 DEFINING A HEURISTIC SCALE TO COMPOSITION FAULT INCLINED (SCFI) AND SCALE TO COMPOSITION HALENESS (SCH)

6.4.1 Dataset Preprocessing

The Dataset opted is of 14 attributes (see Table 6.1) with values of type continue and categorical. The detailed exploration of these attributes given in our earlier article [73]. The dataset opted is of the records, such that each record is of the 14 composition QOS representative attribute values. To facilitate the attribute optimization process devised here in this chapter, the values of the attributes in the given dataset should be numeric and categorical. Henceforth, initially, we convert all continuous values to categorical.

Table 6.1: Description of dataset attributes:

Attribu te ID	Attribute of Complete Record	Description	Value state of the Attribute
1	Associability	Services of same provider used for optimality	Ratio against expected
2	Cyclic	Number of services required to be cyclic	Services involved in composition with cyclic behavior
3	Dependent	No of services dependent of others	The count of service in composition dependent of other services
4	Parallel	No of services executes parallel	Count of services in composition with parallel execution
5	Repetitive	No of services invoked repeatedly due to failure	Count of services invoked repeatedly due to response failure

6	Uptime	Average of the services as composition uptime	Average of percentage of services uptime involved in composition
7	Services count	No of services in composition	Total Number of services involved in composition
8	Diversity	No of services of divergent providers or environment	Services that are not of same provider or same environment
9	Roundtrip time	The completion time of the composition	Composition completion time
10	Cost	composition cost	Total cost of the services as composition cost
11	Reliability	response accuracy	Percentage of response accuracy
12	response time	Composition response Time	Average response time of the services involved in composition
13	versioning ratio	Composition versioning count	No of times composition changed due to change of services, removing existing or adding new services
14	Status	Indicates composition is fault inclined or hale	1 represents fault inclined, 0 represents hale

6.4.2 Attribute Optimization for Defining Scale to Composition Fault Inclined

Let partition the preprocessed set of records based on their labels, such that the records labeled as hale are one set, records labeled as fault inclined is another set. Consider the unique values of each attribute values set $f_i v(NRS)$ in the resultant records-set (NRS) with records labeled as hale and their coverage percentage as $f_i v = \{f_i(v_1,c_1), f_i(v_2,c_2), f_i(v_3,c_3), f_i(v_4,c_4),......, f_i(v_j,c_j)\}$. Further, the attribute optimization for fault inclined records is done as follows:

- Let consider the records set $rs(NRS)$ contains records those labeled as normal.
- Let $f_i(FPS)$ be the attribute f_i of FPS and $f_i(FPS)_{vs}$ be the set of values assigned to that attribute in FPS
- Create an empty set $\overline{f_i(NRS)_{vs}}$ of size $|f_i(FPS)_{vs}|$, then fill it with values from $f_i v(NRS)$ according to their coverage percentage such that $|f_i(FPS)_{vs}| \cong |\overline{f_i(NRS)_{vs}}|$.
- This process is opted to prepare the attribute values vector $\overline{f_i(NRS)_{vs}}$ of each attribute f_i of the NRS,
- This process should be applied to all attributes of the recordset and refer that resultant attributes with values as a set $\overline{NRS}$.
- The canonical correlation (see section 6.4.4) will be done further, which is between each attribute values set $f_i(FPS)_{vs}$ and $\overline{f_i(NRS)_{vs}}$ of FPS and $\overline{NRS}$ respectively.
- Further, the attributes of the FPS can consider as optimal, which are having the canonical correlation is less than given threshold or zero. Further we form a record set $OFPS$, which is having records with values of only attributes that are assessed as optimal through canonical correlation, and this record set $OFPS$ is used further to define the scale to Composition Fault Inclined (SCFI).

6.4.3 Attribute Optimization for Defining Scale to Composition Haleness

As like as the process explored in section 6.4.2, consider the unique values of each attribute values set $f_i v(FPS)$ in the resultant records-set (FPS) with records labeled as normal and their coverage percentage as $f_i v(FPS) = \{f_i(v_1,c_1), f_i(v_2,c_2), f_i(v_3,c_3), f_i(v_4,c_4),......, f_i(v_j,c_j)\}$. Further, the attribute optimization for Fault Inclined records is done as follows:

- Let consider the records set $rs(FPS)$ contains records those labeled as fault inclined.

- Let $f_i(NRS)$ be the attribute f_i of NRS and $f_i(NRS)_{vs}$ be the set of values assigned to that attribute in NRS
- Create an empty set $\overline{f_i(FPS)_{vs}}$ of size $|f_i(NRS)_{vs}|$, then fill it with values from $f_i v(FPS)$ according to their coverage percentage such that $|f_i(NRS)_{vs}| \cong |\overline{f_i(FPS)_{vs}}|$.
- This process is opted to prepare the attribute values vector $\overline{f_i(FPS)_{vs}}$ of each attribute f_i the FPS,
- This process should be applied to all attributes of the recordset and refer that resultant attributes with values as a set $\overline{FPS}$.
- The canonical correlation analysis (see section 6.4.4) will be done further, which is between each attribute values set $f_i(NRS)_{vs}$ and $\overline{f_i(FPS)_{vs}}$ of NRS and $\overline{FPS}$ respectively.

Further, the attributes of the NRS can be considered as optimal, which are having the canonical correlation is less than given threshold or zero. Further we form a record set $ONRS$, which is having records with values of only attributes that are assessed as optimal through canonical correlation, and this record set $ONRS$ is used further to define the Scale to Composition Haleness (SCH).

6.4.4 Canonical Correlation Analysis

Canonical correlation analysis (CCA) [74], [75] is an old statistical technique which has during the last decade become popular in various signal processing and data analysis applications because it often provides in practice quite excellent and meaningful results. Standard CCA measures the linear relationships between two multidimensional datasets X and Y using their second-order statistics, auto covariances and cross-covariances. It finds two bases, one for both X and Y, in which the cross-correlation matrix between the data sets X and Y becomes diagonal and the correlations of the diagonal are maximized.

In CCA, the dimensions of the data vectors $x \in X$ and $y \in Y$ can be different, but they are assumed to have zero means. The number of the data vectors in X and Y must be the same. The exact conditions required for the canonical correlations and the problem solution are discussed in [74], [75]. It turns out these canonical correlations can compute by solving the eigenvector equations.

$$C_{xx}^{-1}C_{xy}C_{yy}^{-1}C_{yx}w_x = \rho^2 w_x$$

$$C_{yy}^{-1}C_{yx}C_{xx}^{-1}C_{xy}w_y = \rho^2 w_y \qquad \text{(Eq 6.1)}$$

Here $C_{yx} = E\{yx^T\}$. The Eigenvalues ρ^2 are squared canonical correlations and the eigenvectors w_x and w_x are normalized CCA basis vectors. Only non-zero solutions to these equations are usually of interest, and their number is equal to the smaller of the dimensions of the vectors x and y.

The solution (1) can simplify if the data vectors x and y are pre-whitened [76], which is the usual practice in many ICA and BSS methods. After pre-whitening, both C_{xx} and C_{yy} become unit matrices, and noting that $C_{yx} = C_{xy}^T$ (Eq 6.1) becomes

$$C_{xy}C_{xy}^T w_x = \rho^2 w_x$$

$$C_{yx}C_{yx}^T w_y = \rho^2 w_y \qquad \text{(Eq 6.2)}$$

But these are just the defining equations for the singular value decomposition (SVD) [77] of the cross-covariance matrix C_{xy} :

$$C_{xy} = U\Sigma V^T = \sum_{i=1}^{L} \rho_i u_i v_i^T \qquad \text{(Eq 6.3)}$$

There U and V are orthogonal square matrices $(U^T U = I,\ V^T V = I)$ containing the singular vectors u_i and v_i . In our case, these singular vectors are the basis vectors w_{xi} and w_{yi} providing canonical correlations. In general, the dimensionalities of the matrices U

and V and consequently the singular vectors u_i and v_i are different corresponding to different dimensions of the data vectors x and y . The pseudo diagonal matrix

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{(Eq 6.4)}$$

Consists of a diagonal matrix D containing the non-zero singular values appended with zero matrices so that the matrix Σ is compatible with the different dimensions of x and y . These non-zero singular values are just the nonzero canonical correlations. If the cross-covariance matrix C_{xy} has full rank, their number is the smaller one of the dimensions of the data vectors x and y .

6.4.5 Defining the Scale to Composition Fault Inclined (SCFI)

Let consider the compositions set $OFPS$ that formed due to canonical correlation analysis (see section 6.4.4).

Further, form a set $F(OFPS)$ such that

$$\begin{aligned} F(OFPS) = \{ & f_1(OFPS) = \{v_{11}, v_{12}, v_{13}, ...v_{1a}\}, \\ & f_2(OFPS) = \{v_{21}, v_{22}, v_{23}, ...v_{2b}\}, \\ & . \\ & . \\ & . \\ & . \\ & ., \\ & f_i(OFPS) = \{v_{i1}, v_{i2}, v_{i3}, ...v_{ic}\}\} \end{aligned}$$

Here in the above description $f_i(OFPS) = \{v_{i1}, v_{i2}, v_{i3}, ...\} \forall i = 1..n$ represents the optimal attribute f_i and the unique values $\{v_{i1}, v_{i2}, v_{i3}, ...\}$ assigned to that attribute of all the records in the set $OFPS$.

Rank each value v_{ij} of optimal attribute f_i , which is based on their coverage in the $f_i(OFPS)$.

Further, represent each composition $\{r_i \forall i = 1...|OFPS| \wedge r_i \in OFPS\}$ as a set $rs(r_i)$ with the respective rank of the value of each optimal attribute as follows:

$$r_i = \{f_1(v_j \forall j \in \{1..|f_1|\}), f_2(v_j \forall j \in \{1..|f_2|\}), f_3(v_j \forall j \in \{1..|f_3|\}),......., f_i(v_j \forall j \in \{1..|f_i|\})\}$$

$$rs(r_i) = \{r(f_1(v_j \forall j \in \{1..|f_1|\})), r(f_2(v_j \forall j \in \{1..|f_2|\})), r(f_3(v_j \forall j \in \{1..|f_3|\})),......., r(f_i(v_j \forall j \in \{1..|f_i|\}))\}$$

Here in this description r_i is a record that belongs to the $OFPS$, which is representing the set of respective values of the optimal attributes. The representation $f_i(v_j \forall j \in \{1..|f_i|\})$ is the value v_j of optimal attribute f_i, and $|f_i|$ represents the size of all possible values to the attribute f_i. And the set $rs(r_i)$ is representing composition r_i by the respective ranks of the values of the optimal attributes. The representation $r(f_i(v_j \forall j \in \{1..|f_i|\}))$ is the rank of the value v_j of the attribute f_i.

Further, for each $rs(r_i)$, find the aggregate rank $ar(r_i)$ as follows, which is an average of ranks representing the respective values of the optimal attributes of the composition r_i

$$ar(r_i) = \frac{\sum_{i=1}^{n} r(f_i(v_j \forall j \in \{1...|f_i|\})}{n} \qquad \text{(Eq 6.5)}$$

The standards defined by ANOVA [71],

(i) The measured average reflects the centrality of the distribution, but not significant to consider it alone as representation of the distribution, since it is not considering the uniform distribution.

(ii) The standard deviation of these ranks represents the how they deviated from each other, which is also not confirming the distribution status.

(iii) The kurtosis represents the state of uniform distribution. If kurtosis found to be platykurtic (kurtosis value less than three), then it is representing the uniform distribution.

(iv) Henceforth, the distribution with platykurtic value is significant to consider as uniform distribution.

Henceforth, we measure the kurtosis of each distribution and order them by their kurtosis from minimal to maximal. The kurtosis of the ranks of each composition t_i is measured as follows:

$$\sigma_{ar(r_i)} = \sqrt{\frac{\sum_{i=1}^{n}(r(f_i(v_j \forall j \in \{1...|f_i|\}) - ar(r_i))^2}{n}} \qquad \text{(Eq 6.6)}$$

$$m4 = \frac{\sum_{i=1}^{n}(r(f_i(v_j \forall j \in \{1...|f_i|\}) - ar(r_i))^4}{n} \qquad \text{(Eq 6.7)}$$

$$g_{(r_i)} = \frac{m4}{\sigma_{ar(r_i)}} \qquad \text{(Eq 6.8)}$$

Here in these equations $\sigma_{ar(r_i)}$ represents the variation observed between ranks of optimal attributes of a composition r_t and $g_{(r_i)}$ represents the kurtosis observed between the ranks of the optimal attributes of the composition r_i

Further, we consider the compositions with the platykurtic distribution of the ranks, and then mean of the ranks of these records will consider as a scale to assess the composition fault Inclined.

$$\mu(OFPS) = \frac{\sum_{i=1}^{n} ar(r_i)}{n} \qquad \text{(Eq 6.9)}$$

Here $\mu(OFPS)$ represents the mean of the aggregate ranks of n compositions of $OFPS$

$$SCFI = \frac{\sum_{i=1}^{m} ar(r_i)}{m} \qquad \text{(Eq 6.10)}$$

Here in the above equation $SCFI$ represents the scale composition fault Inclined, m represents the number of records with platykurtic rank distribution ($g_{(r_i)} < 3$) and having the rank higher than $\mu(OFPS)$.

The lower and upper bounds of the scale will assess as follows:

$$stdv_{OFPS} = \sqrt{\frac{\sum_{i=1}^{m}\left(ar(r_i) - SCFI\right)^2}{m-1}} \qquad \text{(Eq 6.11)}$$

Here in above equation the standard deviation of the aggregate ranks of all record in $ODRS$ is measured

$$SCFI_{low} = SCFI - stdv_{OFPS}$$

$$SCFI_{upr} = SCFI + stdv_{OFPS}$$

6.4.6 Scale to Composition Haleness (SCH).

The scale that devised here in this section is aimed to assess the Haleness state of the composition. The Strategy that explored on compositions set $OFPS$ to define SCFI (see section 6.4.5) is also applied on compositions set $ONRS$ to devise Scale to Composition Haleness (SCH). The process applied on $ONRS$ is briefed here:

Form a set $F(ONRS)$ such that

$$\begin{aligned} F(ONRS) = \{ & f_1(ONRS) = \{v_{11}, v_{12}, v_{13}, \ldots\ldots v_{1a}\}, \\ & f_2(ONRS) = \{v_{21}, v_{22}, v_{23}, \ldots\ldots v_{2b}\}, \\ & . \\ & . \\ & . \\ & . \\ & ., \\ & f_i(ONRS) = \{v_{i1}, v_{i2}, v_{i3}, \ldots\ldots v_{ic}\}\} \end{aligned}$$

Here in the above description $f_i(ONRS) = \{v_{i1}, v_{i2}, v_{i3}, \ldots\ldots\} \forall i = 1..n$ represents the optimal attribute f_i and the unique values $\{v_{i1}, v_{i2}, v_{i3}, \ldots\ldots\}$ assigned to that attribute of all the records in set $ONRS$.

Rank each value v_{ij} of optimal attribute f_i, which is based on their coverage in the $f_i(ONRS)$.

Further, represent each composition $\{r_i \forall i = 1 \ldots | ONRS | \wedge r_i \in ONRS\}$ as a set $rs(r_i)$ with the respective rank of the value of each optimal attribute as follows:

$$r_i = \{f_1(v_j \forall j \in \{1..|f_1|\}), f_2(v_j \forall j \in \{1..|f_2|\}), f_3(v_j \forall j \in \{1..|f_3|\}), \ldots\ldots., f_i(v_j \forall j \in \{1..|f_i|\})\}$$

$$rs(r_i) = \{r(f_1(v_j \forall j \in \{1..|f_1|\})), r(f_2(v_j \forall j \in \{1..|f_2|\})), r(f_3(v_j \forall j \in \{1..|f_3|\})), ..., r(f_i(v_j \forall j \in \{1..|f_i|\}))\}$$

For each $rs(r_i)$, find the aggregate rank $ar(r_i)$ as follows, which is an average of ranks representing the respective values of the optimal attributes of the composition r_i

$$ar(r_i) = \frac{\sum_{i=1}^{n} r(f_i(v_j \forall j \in \{1...|f_i|\})}{n} \quad \text{(Eq 6.12)}$$

According to the ANOVA [71] standards (explored in section 6.4.5), we measure the kurtosis of each distribution and order them by their kurtosis from minimal to maximal. The kurtosis of the ranks of each composition t_i is measured as follows:

$$\sigma_{ar(r_i)} = \sqrt{\frac{\sum_{i=1}^{n} (r(f_i(v_j \forall j \in \{1...|f_i|\}) - ar(r_i))^2}{n}} \quad \text{(Eq 6.13)}$$

$$m4 = \frac{\sum_{i=1}^{n} (r(f_i(v_j \forall j \in \{1...|f_i|\}) - ar(r_i))^4}{n} \quad \text{(Eq 6.14)}$$

$$g_{(r_i)} = \frac{m4}{\sigma_{ar(r_i)}} \quad \text{(Eq 6.15)}$$

Here in these equations $\sigma_{ar(r_i)}$ represents the variation observed between ranks of optimal attributes of a composition r_t and $g_{(r_i)}$ represents the kurtosis observed between the ranks of the optimal attributes of the composition r_i.

Further, we consider the compositions with platykurtic distribution of the ranks, and then mean of the ranks of these records will consider as a scale to assess the Composition Fault Inclined.

$$\mu(ONRS) = \frac{\sum_{i=1}^{n} ar(r_i)}{n} \quad \text{(Eq 6.16)}$$

Here $\mu(ONRS)$ represents the mean of the aggregate ranks of n compositions of $ONRS$

$$SCH = \frac{\sum_{i=1}^{m} ar(r_i)}{m} \quad \text{(Eq 6.17)}$$

Here in the above equation SCH represents the scale to Composition haleness, m represents the number of records with platykurtic rank distribution ($g_{(r_i)} < 3$) and having the aggregate rank greater than $\mu(ONRS)$.

The lower and upper bounds of the scale will assess as follows:

$$\overline{\mu} = \frac{\sum_{i=1}^{m} \{ar(r_i) \exists g_{(r_i)} < 3\}}{m} \qquad \text{(Eq 6.18)}$$

The above equation is finding the mean $\overline{\mu}$ of the aggregate rank of the records with platykurtic feature rank distribution.

$$stdv_{ONRS} = \sqrt{\frac{\sum_{i=1}^{m} \left(ar(r_i) - \overline{\mu}\right)^2}{m-1}} \qquad \text{(Eq 6.19)}$$

Here in above equation the standard deviation of the aggregate ranks of all record in $ONRS$ is measured

$$SCH_{low} = SCH - stdv_{ONRS}$$

$$SCH_{upr} = SCH + stdv_{ONRS}$$

The Scale to Composition Fault Inclined ($SCFI$) and Scale to Composition Haleness (SCH) that are assessed from the given compositions set for training will be used further to assess the scope of Composition Fault Inclined or haleness of a given composition.

For a given composition mr to be tested, as explored in section 6.4.5,

- Preprocess the given composition and reform it as record mr_{SCFI} with only values of optimal attributes of the $SCFI$ and similarly reform it as record mr_{SCH} with only values of optimal attributes of the SCH.
- To assess the Fault Inclined of the composition mr_{SCFI}, form $rs(mr_{SCFI})$ with the respective rank of the value of each optimal attribute of $SCFI$.

- To assess the Composition Haleness of the composition mr_{SCH}, form $rs(mr_{SCH})$ with the respective rank of the value of each optimal attribute of SCH.
- Find the aggregate rank of the mr_{SCFI} as $ar(mr_{SCFI})$ and also find the aggregate rank of the mr_{SCH} as $ar(mr_{SCH})$

Further, the state of given composition mr is assessed as follows:

Table 6.2: The Scale of Composition Fault Inclined, Haleness, and assessing strategy

$ar(mr_{SCFI}) \leq SCFI_{low}$ && $ar(mr_{SCH}) \geq SCH_{upr}$	Haleness
$ar(mr_{SCFI}) \geq SCFI_{upr}$ && $ar(mr_{SCH}) \geq SCH_{upr}$	Inclined
$ar(mr_{SCFI}) \geq SCFI$ && $ar(mr_{SCH}) \geq SCH_{upr}$	Inclined
$ar(mr_{SCFI}) \leq SCFI_{low}$ && $ar(mr_{SCH}) \geq SCH$	Haleness
$ar(mr_{SCFI}) \leq SCFI$ && $ar(mr_{SCH}) \geq SCH$	Haleness
$ar(mr_{SCFI}) \geq SCFI_{upr}$ && $ar(mr_{SCH}) \geq SCH$	Inclined
$ar(mr_{SCFI}) \geq SCFI$ && $ar(mr_{SCH}) \geq SCH$	Inclined
$ar(mr_{SCFI}) \leq SCFI_{low}$ && $ar(mr_{SCH}) \leq SCH_{low}$	Inclined
$ar(mr_{SCFI}) \leq SCFI$ && $ar(mr_{SCH}) \leq SCH_{low}$	Inclined
$ar(mr_{SCFI}) \geq SCFI_{upr}$ && $ar(mr_{SCH}) \leq SCH_{low}$	Inclined
$ar(mr_{SCFI}) \geq SCFI$ && $ar(mr_{SCH}) \leq SCH_{low}$	Inclined
$ar(mr_{SCFI}) \leq SCFI_{low}$ && $ar(mr_{SCH}) < SCH$	Haleness
$ar(mr_{SCFI}) \leq SCFI$ && $ar(mr_{SCH}) < SCH$	Inclined
$ar(mr_{SCFI}) \geq SCFI_{upr}$ && $ar(mr_{SCH}) < SCH$	Inclined

$ar(mr_{SCFI}) \geq SCFI \ \&\& \ ar(mr_{SCH}) < SCH$	Inclined

Here in Table 6.2, all possible combinations of SCFI and SCH and the impact of those combinations explored. Regardless of the mr_{SCH}, if mr_{SCFI} is greater than $SCFI$ then the record confirmed to be Composition Fault Inclined. But in contrast, the Composition Haleness is dependent of $SCFI$, which is indicating that though the composition mr_{SCH} is more than the value of SCH, it's mr_{SCFI} must be less than the $SCFI$ to conclude that the given composition mr is scaled as Haleness. This may lead to slight increase in false positives in prediction but strictly avoids false negatives, which is an accuracy measurement of fault Inclined.

6.5 EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

The experiments were carried out on dataset that explored in section 6.4.1. Initially partitioned the processed dataset into normal and fault inclined records and then the optimal attributes of fault inclined compositions and normal compositions were traced out, which is by using the process explored (see section 6.4.2, 6.4.3, and 6.4.4). Further, the scale Composition Fault Inclined (SCFI) and Scale to Composition Haleness (SCH) was devised through the process explored (see section 6.4.5 and 6.4.6). The exploration of the input data and results shown in Table 6.3, Table 6.4 and Table 6.5. The visualization of the optimal scope of attributes for fault inclined and hale compositions can find in Figure 6.1 and Figure 6.2.

Table 6.3: Statistics of the experiment results

Total Number of Compositions	303
Range of QoS attributes of a composition	13
Compositions used for defining scale	80% (242 records)

Compositions used for performance analysis	20% (61 records)
Scale of Composition Fault Inclined $SCFI$ observed	7.11324
$SCFI_{low}$ observed	5.575204
$SCFI_{upr}$ observed	8.651277
Scale of Composition Haleness SCH observed	2.982372
SCH_{low} observed	2.35923
SCH_{upr} observed	3.605513

Table 6.4: Selected QoS attributes of the compositions labeled as fault inclined under different canonical correlation threshold

<0.04		<0.05		<0.051(0.06) (mean of the CC		<0.1		<0.2	
Attribute ID	CC value	ID	CC	ID	CC	ID	CC	ID	CC
1	0.000398	1	0.000398	1	0.000398	1	0.000398	1	0.000398
7	0.021013	4	0.044147	3	0.050546	2	0.107357	2	0.107357

8	0.029505	7	0.021013	4	0.044147	3	0.050546	3	0.050546
9	0.03851	8	0.029505	5	0.050081	4	0.044147	4	0.044147
12	0.021347	9	0.03851	7	0.021013	5	0.050081	5	0.050081
		11	0.042488	8	0.029505	7	0.021013	6	0.103026
		12	0.021347	9	0.03851	8	0.029505	7	0.021013
				11	0.042488	9	0.03851	8	0.029505
				12	0.021347	10	0.098274	9	0.03851
						11	0.042488	10	0.098274
						12	0.021347	11	0.042488
						13	0.099685	12	0.021347
								13	0.099685

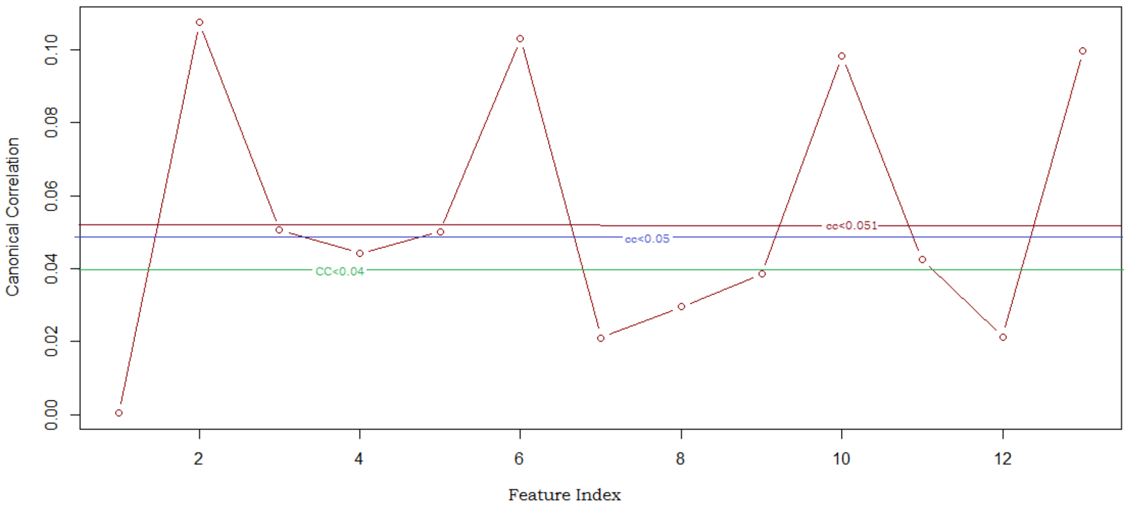

Figure 6.1: QoS Attributes of the compositions labeled as fault inclined and their optimality under divergent CC thresholds

Table 6.5: Selected QoS attributes of the compositions labeled as hale under different canonical correlation thresholds

<0.04		<0.06		<0.07585903 (mean of the cc vales) (0.08)		<0.1		<0.2 (yes, It's all)	
1	0.011762	1	0.011762	1	0.011762	1	0.011762	1	0.011762
5	0.004778	5	0.004778	2	0.069674	2	0.069674	2	0.069674
8	0.039421	8	0.039421	5	0.004778	3	0.082918	3	0.082918
11	0.021706	9	0.043843	8	0.039421	5	0.004778	4	0.190604

	11	0.021706	9	0.043843	8	0.039421	5	0.004778
	13	0.043684	11	0.021706	9	0.043843	6	0.177771
			12	0.066248	10	0.089583	7	0.144176
			13	0.043684	11	0.021706	8	0.039421
					12	0.066248	9	0.043843
					13	0.043684	10	0.089583
							11	0.021706
							12	0.066248
							13	0.043684

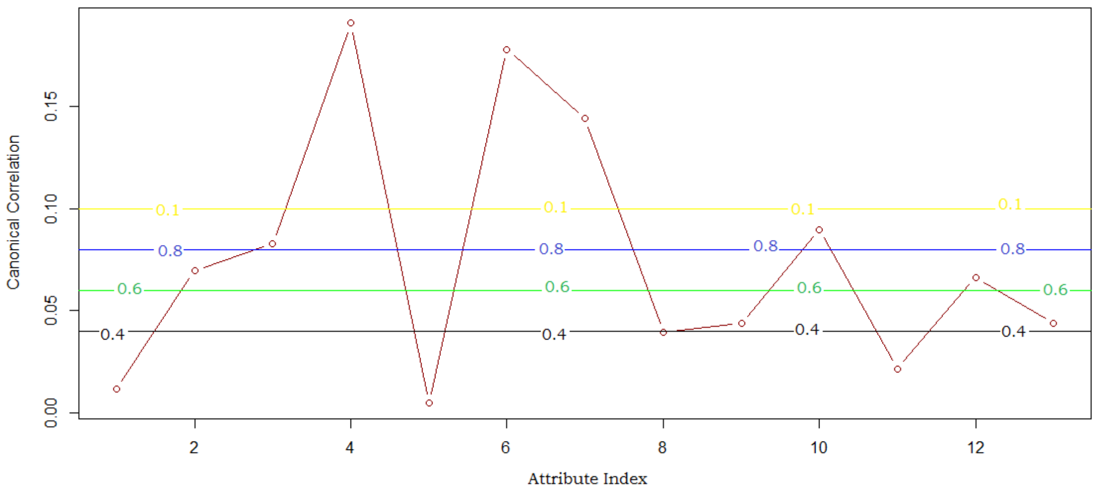

Figure 6.2: QoS Attributes of the Compositions labeled as hale and their optimality under divergent CC thresholds.

6.5.1 Performance Analysis

The robustness and prediction accuracy of the scales *SCFI* and *SCH* are assessed through 62 records, which are of the combination of 40 faults inclined and 22 hale compositions.

The prediction statistics are as follow:

The count of true positives are (records predicted as truly fault inclined) 40. The count of true negatives are (records predicted as truly hale) 20. The count of false positives are (records predicted as falsely fault inclined) 2. And the count of false negatives are (the records predicted as falsely hale) 0.

Since the experimental results indicating that fault inclined composition prediction is 40 out of given 40 compositions, hence prediction error towards fault inclined compositions are 0.0. The prediction accuracy of normal compositions is observed as 20 among the given 22 compositions, hence the prediction error ratio is approx. 0.09, which is to be negligible as in the track of sensitive service composition needs, the composition fault inclined should not be diagnosed as hale, in contrast, a hale composition can be suspected falsely as fault inclined and may recommend further assessment strategies.

Figure 6.3: Process completion time of SCFI and SCH under divergent optimal attributes selected through various canonical correlation thresholds.

The prediction accuracy of the model devised here in this chapter is explored through statistical assessment metrics called precision, recall and f-measure (see Table 6.5). The value obtained for metric recall indicating that the devised model is highly robust and scalable towards assessing the fault-prone scope, and the precision also indicating that the prediction accuracy of the model is high and approximately it is 97%.

Table 6.6: The precision, recall and the f-measure of the predictions

Precision	0.952381
Recall	1
F-Measure	0.97561

The process time of the application is stable since the increase in number of optimal attributes is not influencing the process complexity (see Figure 6.3).

6.6 COMPARATIVE STUDY

The performance analysis of the proposed model done using the dataset prepared from the services and their impact information synthesized. The reason to not opting the existing datasets such as [82], [83] is that these datasets are not considering the maximum number of QoS metrics to assess the service quality, they limited to QoS metrics response time, throughput. Nevertheless, service selection influenced by other QoS metrics such as Availability, Accessibility, cost, Integrity, throughput, roundtrip time, Reliability, Regulatory and another critical factor Security. The dataset that used to assess the performance of the proposal is explored in Table 6.7.

Table 6.7: Description of the dataset used for experiments

Range of tasks involved in compositions	7-25

Range of services available for each task	7-16

Java implementation of the proposed model did the empirical analysis. Further to obtain the statistical measures such as skewness, kurtosis of the discovered compositions was done by using the expression language called R. The experiments were done with different count of compositions range from 7 (sparse) to 25 (dense) as input.

The results observed from experiments indicating that the proposed Orthogenesis AGPA based Optimal Fitness Aware Service Compositions Discovery (SCFI&SCH) are scalable and robust that compared to other AGPA based QoS aware web service composition approaches called GRASP [51], DIO [50], AGPA [48] WSMO [49], MHS [81], and CQPSO [69]. The evolutions completion time of SCFI&SCH is considerably low and scalable that outperformed the other models compared, the evolution complexity observed at proposed SCFI&SCH is linear (see Figure 6.4). The skewness observed for QoS fitness distribution over the top n (here in experiments $n = 10$) resultant compositions from the SCFI&SCH is low and optimal (see Figure 6.5).

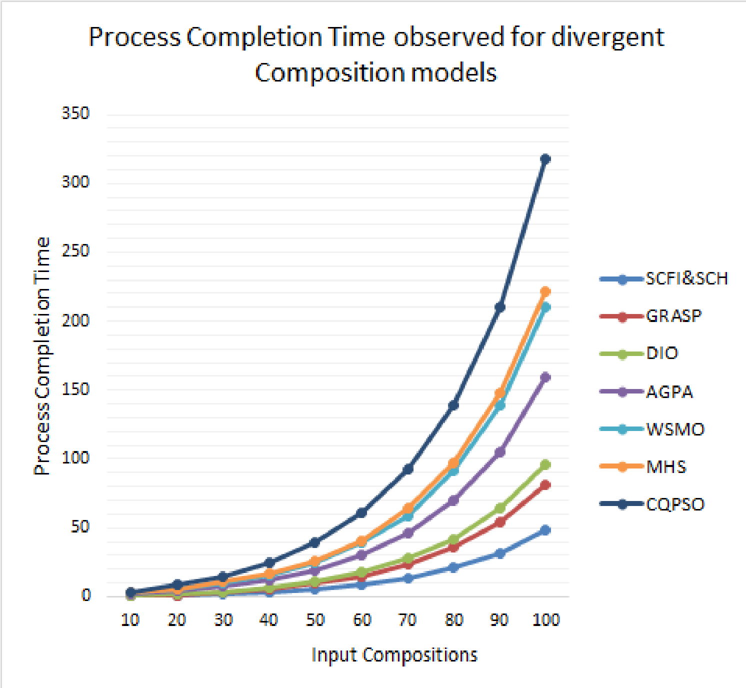

Figure 6.4: The completion time observed for contemporary QoS aware Service compositions [51], [50], [48], [49], [81], [69] and SCFI&SCH

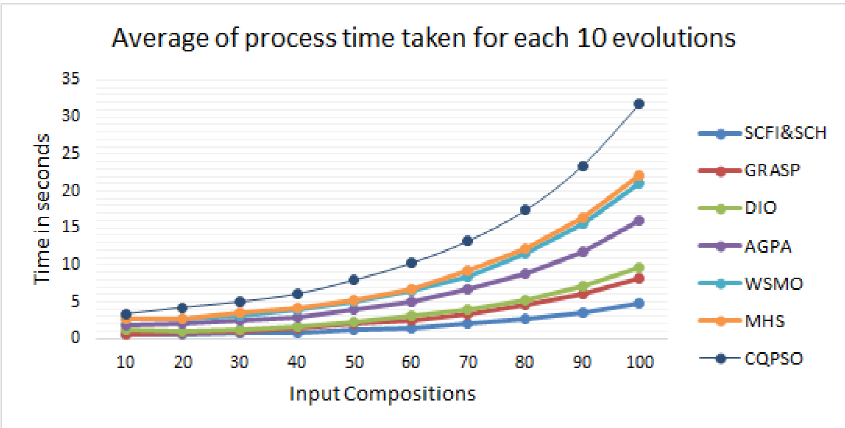

Figure 6.5: The computational complexity observed for evolutions contemporary QoS aware Service compositions [51], [50], [48], [49], [81], [69] and SCFI&SCH

Fitness observed for top 10 resultant compositions from divergent GA based ervice composition algorithms

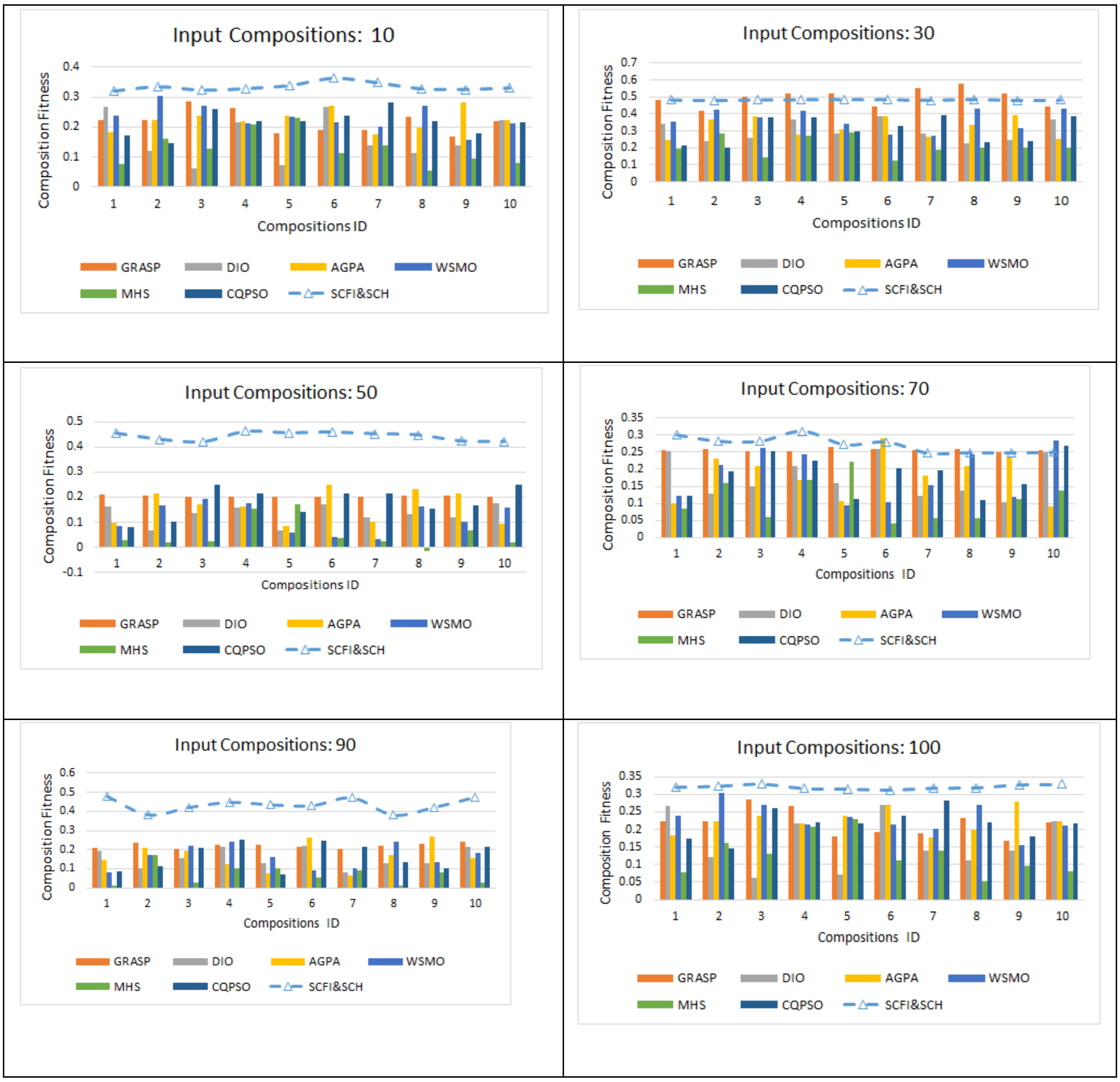

Figure 6.6: Fitness distribution over best n resultant compositions from divergent composition approaches

The significance of average fitness spanned over test cases conducted with different input compositions assessed using statistical tests called t-test, Wilcoxon signed rank test. The results obtained from these tests, and their corresponding degree of probabilities depicted in Table 6.8, Table 6.9, Table 6.10 and Table 6.11 respectively.

The t-scores obtained for SCFI&SCH corresponding to other methods depicting that in all test conditions the average fitness is significantly high that compared to the average fitness observed under all test conditions from other traditional methods. This is since, the t-scores found for SCFI&SCH corresponding to other models are positive (see Table 6.8) with degree of probability less than the threshold value 0.05 (see Table 6.9). The average fitness of the other benchmark models spanned over different test cases are evincing the similarity, this is since, their degree of probability is higher than the p-value threshold (0.05) considered (see Table 6.9).

Table 6.8: The Average fitness observed from SCFI&SCH and contemporary QoS aware Service compositions [51], [50], [48], [49], [81], [69] for each two best cases of divergent input compositions

Mean of the fitness							
No of compositions↓	SCFI&SCH	GRASP	DIO	AGPA	WSMO	MHS	CQPSO
10	0.5138	0.5875	0.3261	0.327	0.3333	0.323	0.3241
10	0.4899	0.3357	0.3156	0.3166	0.3159	0.3071	0.3229
30	0.4788	0.6079	0.34	0.3314	0.3691	0.2384	0.3064
30	0.4784	0.3128	0.2901	0.3297	0.3451	0.2053	0.2874
50	0.4603	0.1915	0.1474	0.1538	0.1109	0.0738	0.1618
50	0.4114	0.1859	0.1475	0.1556	0.0993	0.0678	0.1724
70	0.2087	0.2336	0.1852	0.1797	0.1809	0.1209	0.1816
70	0.3597	0.2673	0.2007	0.2027	0.2034	0.1333	0.1998
90	0.472	0.1864	0.1581	0.1559	0.1487	0.0769	0.152
90	0.375	0.2401	0.1745	0.172	0.1554	0.0922	0.1712
100	0.3436	0.1543	0.1387	0.1674	0.1662	0.1245	0.1698
100	0.2041	0.3158	0.2037	0.249	0.247	0.1761	0.2423

Table 6.9: one-tailed t-scores observed between average fitness values of each pair of AGPA based approaches considered and proposed

t-score							
	SCFI&SCH	GRASP	DIO	AGPA	WSMO	MHS	CQPSO
SCFI&SCH		1.85282	1.85282	4.54245	4.29244	5.96342	4.83606
GRASP	-1.8528		1.69982	1.50505	1.53129	2.77954	1.62635
DIO	-1.8528	-1.6998		-0.3013	-0.112	1.68943	-0.1807
AGPA	-4.5425	-1.5051	0.30125		-0.112	1.68943	-0.1807
WSMO	-4.2924	-1.5313	0.11201	0.11201		1.62266	-0.04056
MHS	-5.9634	-2.7795	-1.6894	-1.6894	-1.6894		-1.93679
CQPSO	-4.8361	-1.6264	0.1807	0.1807	0.04056	1.93679	

Table 6.10: The degree of probability (p-value) of one-tailed t-scores observed between average fitness values of each pair of AGPA based approaches considered and proposed

p-value/t-score							
	SCFI&SCH	GRASP	DIO	AGPA	WSMO	MHS	CQPSO
SCFI&SCH		0.03869	0.03869	0.00008	0.00015	0.00001	0.000039
GRASP	0.03869		0.05163	0.07327	0.06998	0.00547	0.059058
DIO	0.03869	0.05163		0.38303	0.45592	0.05263	0.42913

AGPA	0.00008	0.07327	0.07327		0.45592	.052631	0.42913
WSMO	0.00015	0.06998	0.06998	0.45592		0.05945	0.484008
MHS	0.00001	0.00547	0.00547	0.05263	.052631		0.032859
CQPSO	3.9E-05	0.05906	0.05906	0.42913	0.484008	0.032859	

The z-scores and their corresponding degree of probability obtained from Wilcoxon signed rank test depicted in Table 6.10 and Table 6.11 respectively. Concerning the observations of these scores, it is apparent to conclude that the average fitness spanned over test cases conducted with different input compositions under SCFI&SCH is outperformed the other benchmarking models. This is since the degree of probability observed for the z-scores obtained from the proposed model SCFI & SCH corresponding to other benchmarking models are less than the p-value threshold 0.01 (see Table 6.11 and Table 6.12). The similar impact evinced for the modes GRASP [51] and MHS [81] corresponding to other benchmarking models (see Table 6.11 and Table 6.12).

Table 6.11: one-tailed z-scores of Wilcoxon signed rank test observed between average fitness values of each pair of AGPA based approaches considered and proposed

Z-Score							
	SCFI&SCH	GRASP	DIO	AGPA	WSMO	MHS	CQPSO
SCFI&SCH		-2.1181	-3.0594	-2.9025	-2.9025	-3.0594	-2.9025
GRASP	-2.1181		-3.0594	-2.8241	-2.7456	-3.0594	-2.8241
DIO	-3.0594	-3.0594		-1.1767	-0.3922	-3.0594	-3.0594
AGPA	-2.9025	-2.8241	-1.1767		-0.7452	-3.0594	-0.6276
WSMO	-2.9025	-2.7456	-0.3922	-0.7452		-3.0594	-0.2353
MHS	-3.0594	-3.0594	-3.0594	-3.0594	-3.0594		-3.0594
CQPSO	-2.9025	-2.8241	-3.0594	-0.6276	-0.2353	-3.0594	

Table 6.12: The degree of probability (p-value) of one-tailed z-scores of Wilcoxon signed rank test observed between average fitness values of each pair of AGPA based approaches considered and proposed

p-value/z-score							
	SCFI&SCH	GRASP	DIO	AGPA	WSMO	MHS	CQPSO
SCFI&SCH		0.017	0.00111	0.00187	0.00187	0.00111	0.00187
GRASP	0.017		0.00111	0.0024	0.00298	0.00111	0.0024
DIO	0.00111	0.00111		0.119	0.34827	0.00111	0.00111
AGPA	0.00187	0.0024	0.119		0.22663	0.00111	0.26435
WSMO	0.00187	0.00298	0.34827	0.22663		0.00111	0.40517
MHS	0.00111	0.00111	0.00111	0.00111	0.00111		0.00111
CQPSO	0.00187	0.0024	0.00111	0.26435	0.40517	0.00111	

6.7 CHAPTER SUMMARY

This chapter introduced a novel heuristic scale to assess the fault Inclined scope of the given service composition. Regarding this, two heuristic metrics called Scale to Composition Fault Inclined (SCFI) and Scale to Composition Haleness (SCH) is devised. In contrast to the existing benchmarking models, the proposed metrics are assessing the composition fault Inclined scope, and haleness referred as SCFI and SCH respectively. Further, the combinations of these SCFI and SCH values of the given service compositions are used to assess the state of that composition. The process opted to devise these metrics is initially finding the optimal attributes of the given fault inclined and hale composition records that represented by QoS attributes, which is done through the canonical correlation analysis. Further, the service compositions of fault inclined and hale with optimal attributes are used to assess the metrics SCFI and SCH. The experimental results are optimistic and concluding the prediction accuracy and robustness. The future work can be the definition of fuzzy model to estimate the combination of SCFI and SCH values.