CHAPTER 5

5 SCALE TO COMPOSITION FAULT INCLINED (SCFI): HEURISTIC SCALE TO ASSESS THE IMPACT OF COMPOSITION TOWARDS FAULT INCLINED

5.1 OVERVIEW

Service composition in the service-oriented architecture is a significant activity. To achieve the quality of service and secured operations from the web service compositions, they need to verify their impact towards fault proneness before deploying that service composition. Henceforth, here in this chapter, we devised A heuristic metric to assess the service composition impact scale to composition fault inclined (SCFI). The designed model explores the higher and lower ranges of the SCFI, which is from the earlier compositions that are notified as either fault inclined or hale. The experimental results examined from the empirical study indicating that the devised metrics are significant towards estimating the state of given service composition is fault inclined or hale.

5.2 PROLOGUE

Service-Oriented Architecture (SOA) simplifies information technology-related operational tasks by consumption of ready-to-use services. Such SOA found to realize currently in e-commerce domains such as B2B, B2C, C2B, and C2C, in particular, the web services are one that considered serving under this SOA.

Web services are software components with native functionality that can be operable through the web. Another important factor about this web services is that more than one service can compose as one component by coupled loosely. The standard WSDL is web service descriptive language that let the self-exploration of the web services towards their functionality and UDDI is the registry that allows the devised web services to register and available to required functionality [61].

The composition of web services is the loosely interconnected set of Web service operations that act as a single component, which offers solutions for different tasks of an operation. Since the task of the composition is integrating different web services explored through different descriptors, it is the most fault-prone activity. The functionality of service composition includes the activities such as (i) identify the tasks involved in a given business operation, (ii) trace related web services to fulfill the need of each task, (iii) couple these services by exploring the order of that services usage, which is based on the expected information flow, (iv) and resolve the given operation by ordering the responses of the web services that coupled loosely as one component.

In order to achieve the quality of service and secure transactions in web service composition and usage, the impact of the composition should estimate before deploying those loosely coupled web services as one component.

The Web service compositions used earlier that could find in repositories and the services involved in those compositions helps to assess the impact of these web services towards fault-proneness.

The current composition strategies [62], [63], [64], [65], [57], [59], [22] are error-prone since these State-of-the-art techniques are not mature enough to guarantee the fault-free operations. However, finding these compositions as fault-prone after deployment is functionally costly and not significant towards end level solutions, also may lead to serious vulnerable. Hence the process of estimating the composition scope towards fault proneness is mandatory.

In this chapter, we propose a novel statistical approach to estimate the service composition is fault inclined or hale. Our method acts as an assessment strategy for any of existing web service composition approaches.

The chapter is structured as follows. Section 5.3 discusses related work. In section 5.4, the proposed statistical approach is explored, which followed by Section 5.5 that contains the results examined from the empirical study. The conclusion of the proposal and future research directions discussed in Section 5.6.

5.3 ASSOCIATED WORK

Service compositions with malfunctioned web services lead to form the highly fault-prone compositions. Henceforth the web service composition to serve as one component under SOA is complex and needs research domain attention to deliver effective strategies towards the QoS centric service. The model devised in [56] defined set of QoS factors to predict available services. Many of existing quality-aware service selection strategies aimed to select the best service among multiple services available. The model devised in [22] considering the linear programming to find the linear combination of availability, successful execution rate, response time, execution cost and reputation, which is regarding find the optimal service composition towards given business operation. The model devised in [57] is considering the temporal validity of the service factors. The authors in [58] modeled a mixed integer linear program that examines both local and global constraints.

The model devised in [59] is selecting services as a complex multi-choice multi-dimension rucksack problem that tends to define different quality levels to the services, which further taken into account towards service selection. All these solutions are depended strongly on the positive scores given by users to each parameter. However, it is not scalable to establish them in prospective order.

Though the QoS strategies defined are used in service composition the factor fault-proneness of the service composition as usual. In regard to this, a model devised in [60] explored a mechanism for fault proliferation and resurgence in dynamically connected service compositions. Dynamically coupled architecture outcomes in further complexness in need of fault proliferation between service groups of a composition accomplished by not depending on other service groups.

In a gist, it can conclude that almost all the benchmarking service quality assessment models are attributed specific, user rating specific or both. Hence the importance of attributes is divergent from one composition requirement to other, and contextual factors influence the user ratings, and another key factor is all these benchmark models are assessing services based on their performance, but in practice, the functionality of one service may influence by the performance of another service. Henceforth here in this chapter, we devised a statistical approach that estimates the impact scale of service composition towards fault proneness, which is based on a devised metric called composition support of service compositions and service descriptors.

In contrast to all 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.

5.4 DEFINING SCALE TO SERVICE COMPOSITION FAULT INCLINED

The Dataset opted is of 14 attributes (see Table 5.1) 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. In regard 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 5.1: Description of dataset attributes:

Attribute 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

5.4.1 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 is 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 record set and refer that resultant attributes with values as a set $\overline{NRS}$.
- The canonical correlation (see section 5.4.2) 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).

5.4.2 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.

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

$$C_{yy}^{-1}C_{yx}C_{xx}^{-1}C_{xy}w_y = \rho^2 w_y \qquad \text{(Eq 5.1)}$$

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 5.1) becomes

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

$$C_{yx}C_{yx}^T w_y = \rho^2 w_y \qquad \text{(Eq 5.2)}$$

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

$$C_{xy} = U\Sigma V^T = \sum_{i=1}^{L} \rho_i u_i v_i^T \qquad \text{(Eq 5.3)}$$

There U and V are orthogonal square matrices $(U^T U = I,\ \mathrm{V}^T V = I)$ 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 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

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{(Eq 5.4)}$$

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 .

5.4.3 Defining Scale to Composition Fault Inclined

Initially, we apply the canonical correlation analysis (see section 5.4.2) on the processed dataset (see section 5.4.1). Then the compositions with selected optimal attribute values are used for the further process of the devised scale.

Let set of compositions $C = \{c_1, c_2, c_3, ..., c_n\}$ formed by the values of optimal attributes selected through canonical correlation process (see section 5.4.2).

Prepare a set $FS = \{a_1v_1, a_1v_2, ...a_1v_{m1}, a_2v_1, a_2v_2, ...a_2v_{m2}, ..., a_nv_1, a_nv_2, ...a_nv_{mn}\}$ that contains all the unique values of all optimal attributes. Here $\{a_1v_1, a_1v_2, ..., a_1v_{m1} \exists 0 < m1 \leq |C|\}$ represents the all possible unique values of the optimal attribute a_1 and $\{a_nv_1, a_nv_2, ..., a_nv_{mn} \exists 0 < mn <= |C|\}$ represents the all possible unique values of the optimal attribute a_n . Further, the values of the set FS can refer as features.

5.4.3.1 Process

Initially, we build a weighted graph WG such that values of FS as vertices and edges between these vertices under the constraints such as:

a) No edge is between two vertices if those two are values of the same attribute
b) An edge between two vertices that justifies the above condition is possible if those two vertices appear together in at least one given the composition.

Each edge weighted by the ratio of the given records contain the two vertices of the edge.

In the process of detecting the closeness of each feature of FS with compositions, further, we build a bipartite graph between compositions and all features.

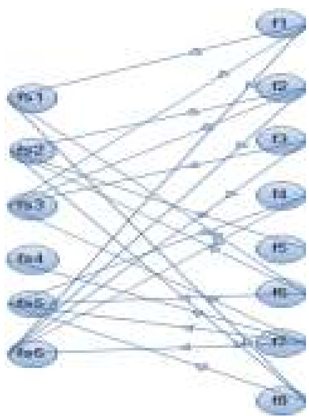

Figure 5.1: Bipartite graph between values of optimal attributes and compositions.

If a feature f_i exists in composition c_j then the weight of the connection between f_i and c_j will be the ratio of the weights of the edges between f_i and each feature of c_j that defined in weighted graph WG (see Eq5).

$$ew_{(c_i \leftrightarrow f_j)} = \frac{\sum_{k=1}^{k<|c_i|} \{ew_{f_k \leftrightarrow f_j} \forall f_k \in c_i \wedge k \neq j\}}{|c_i|} \qquad \text{(Eq 5.5)}$$

The graph representation (Figure 5.1) indicates the bipartite relation between features and compositions.

The devised process of identifying feature weights using bipartite graph is explored below:

Initially, we form a matrix that contains the edge weights of a bipartite graph, such that each feature weight towards each composition.

Table 5.2 is the matrix format of the weight of the edge between composition and feature in the bipartite graph. The weight indicates the ratio of the edge weight between feature and all other features of the target composition, which will consider from the weighted graph.

The graph representation (see Figure 5.1) of the set of compositions gives us the idea of applying link-based ranking models for the evaluation of connected sets. In this bipartite graph, the closeness support of composition c is proportional to the degree of all feature weights. However, it is crucial to have different closeness weights for different compositions to reflect their different importance. The evaluation of composition c influence should derive from these weights. Here comes the question of how to acquire weights in a set of compositions. Intuitively, a composition with high closeness weights should contain many of the features. The reinforcing relationship of compositions and feature sets is just like the relationship between hubs and authorities in the HITS model [78].

Regarding, the compositions as "pure" hubs and the features as "pure" authorities, we can apply HITS to this bipartite graph. The following explored the process:

Table 5.2: matrix A as follows that represents the connection weights between a feature and each composition

	$a_1 v_1$	$a_1 v_2$	.	.	.	$a_1 v_{m1}$	$a_2 v_1$	$a_2 v_2$	.	.	.	$a_2 v_{m2}$	.	.	.	$a_n v_{mn}$
c_1	$ew_{(c_1 \leftrightarrow a_1 v_1}$	$ew_{(c_1 \leftrightarrow a_1 v_2}$	.	.	.	$ew_{(c_1 \leftrightarrow a_1 v_{ml}}$	$ew_{(c_1 \leftrightarrow a_2 v}$	$ew_{(c_1 \leftrightarrow a_2 v_2}$	.	.	.	$ew_{(c_1 \leftrightarrow a_1 v_{m1}}$	.	.	.	$ew_{(c_1 \leftrightarrow a_1 v_m}$
c_2	$ew_{(c_2 \leftrightarrow a_1 v}$	$ew_{(c_2 \leftrightarrow a_1 v}$	.	.	.	$ew_{(c_2 \leftrightarrow a_1 v_m}$	$ew_{(c_2 \leftrightarrow a_2 v}$	$ew_{(c_2 \leftrightarrow a_2 v_2}$	.	.		$ew_{(c_2 \leftrightarrow a_1 v_{m1}}$	.	.	.	$ew_{(c_2 \leftrightarrow a_1 v_m}$
.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
$c_{\|C\|}$	$ew_{(c_{\|C\|} \leftrightarrow a_1 v}$	$ew_{(c_{\|C\|} \leftrightarrow a_1 v}$	.	.	.	$ew_{(c_{\|C\|} \leftrightarrow a_1 v_n}$	$ew_{(c_{\|C\|} \leftrightarrow a_2}$	$ew_{(c_{\|C\|} \leftrightarrow a_2 v}$	.	.	.	$ew_{(c_{\|C\|} \leftrightarrow a_1 v_n}$				$ew_{(c_{\|C\|} \leftrightarrow a_1 v_n}$

Let matrix representation of compositions and features of the set FS as a matrix 'A'(see Table 5.2). The value represents the edge weight between composition and features that calculated by using (Eq 5.7).

Find Authority weights by transposing the matrix A and summing up the columns of each row of matrix transposed that represents edge weights between a feature-set and all possible compositions that gives a matrix v, which is representing the authority weights

Now find the hub weights through matrix multiplication between matrix A and matrix v.

$$u = A \times v$$

Then the closeness support of each feature f can measure as follows

$$cs(f) = \frac{\sum_{i=1}^{|C|} \{u(c_i) : (f \to c_i) \neq 0\}}{\sum_{i=1}^{|C|} u(c_i)} \qquad \text{(Eq 5.6)}$$

And then the fault-proneness of each composition of composition set C can measure as follows

$$cfp(c_i) = 1 - \frac{\sum_{j=1}^{|FS|} \{cs(f_j) : (f_j \to c_i) \neq 0\}}{ec(c_i)} \qquad \text{(Eq 5.7)}$$

Here in the above equation $ec(c_i)$ represents the total number of edges connected to composition c_i

Then the Scale of Composition Fault Inclined $SCFI$ can find as follows:

$$SCFI = \frac{\sum_{i=1}^{|C|} cfp(c_i)}{|C|} \qquad \text{(Eq 5.8)}$$

Here in the above equation $|C|$ indicates the total number of compositions

Further, find the standard deviation of the "cfp" of compositions from $SCFI$ as follows:

$$dv_{SCFI} = \sqrt{\frac{\left(\sum_{i=1}^{|C|} \left(cfp(c_i) - SCFI\right)^2\right)}{(|C| - 1)}} \qquad \text{(Eq 5.9)}$$

Here in the above equation dv_{SCFI} represents the standard deviation of the fault-proneness of compositions from $SCFI$

The lower and upper bounds of the scale to composition fault inclined $SCFI$ can explore as follows

Lower bound of $SCFI$ is

$$SCFI_l = SCFI - dv_{SCFI}$$

Upper Bound of sdp is

$$SCFI_h = SCFI + dv_{SCFI}$$

Further, this scale can use to diagnose the composition is fault inclined or hale as follows:

i) Composition c can be said as safe if and only if $cfp(c) < SCFI_l$
ii) Composition c can be said as fault inclined risk is low if and only if $cfp(c) \geq SCFI_l \,\&\&\, cfp(c) < SCFI$
iii) Composition c can be said as highly fault inclined if and only if $cfp(c) \geq SCFI \,\&\&\, cfp(c) < SCFI_h$
iv) Composition c surely diagnosed as fault inclined if $cfp(c) \geq SCFI_h$

5.5 EMPIRICAL ANALYSIS OF THE PROPOSED MODEL

We explored the credibility of the proposed model on a compositions dataset of 303 records, such that each record represents the values of the QoS attributes, which are explored in section 5.4.1. The 70% of these records were used to devise the scale t composition fault inclined $SCFI$. Further, we used the rest 30% records to predict the fault inclined scope using this $SCFI$. Interestingly, the empirical study delivered promising results. The statistics explored in Table 5.3.

Table 5.3: Statistics of the experiment results

Total Number of Records	303
Range of fields count in a record	14 (training), 13 (testing)
Total number of Features Found in dataset	112 (cc<0.04), 154 (cc<0.05) and 265 (cc<0.051)

Number of bipartite edges found	1247 (cc<0.051), 830 (cc<0.05) and 691 (cc<0.04)
Scale to Composition Fault Inclined	0.46942
Scale to Composition Fault Inclined Threshold Upper Bound	0.51137
Scale to Composition Fault Inclined Threshold Lower Bound	0.42747

The attributes of the compositions selected as optimal under different canonical correlation thresholds are defined in Table 5.4, Table 5.5 and Table 5.6, and the same is visualized in Figure 5.2.

Table 5.4: Canonical correlation of all 13 features of fault inclined records and hale records

Attribute ID	CC value
1	0.000398
2	0.107357
3	0.050546
4	0.044147
5	0.050081
6	0.103026
7	0.021013
8	0.029505

9	0.03851
10	0.098274
11	0.042488
12	0.021347
13	0.099685

Table 5.5: Selected features of the fault inclined records with canonical correlation threshold <0.04

Attribute ID	CC value
1	0.000398
7	0.021013
8	0.029505
9	0.03851
12	0.021347

Table 5.6: Selected features of the fault inclined records with canonical correlation threshold <0.05

Attribute ID	CC value
1	0.000398
4	0.044147
7	0.021013
8	0.029505

9	0.03851
11	0.042488
12	0.021347

Table 5.7: Selected features of the fault inclined records with canonical correlation threshold <0.051 (mean of the Canonical correlations of all features)

Attribute ID	CC value
1	0.000398
3	0.050546
4	0.044147
5	0.050081
7	0.021013
8	0.029505
9	0.03851
11	0.042488
12	0.021347

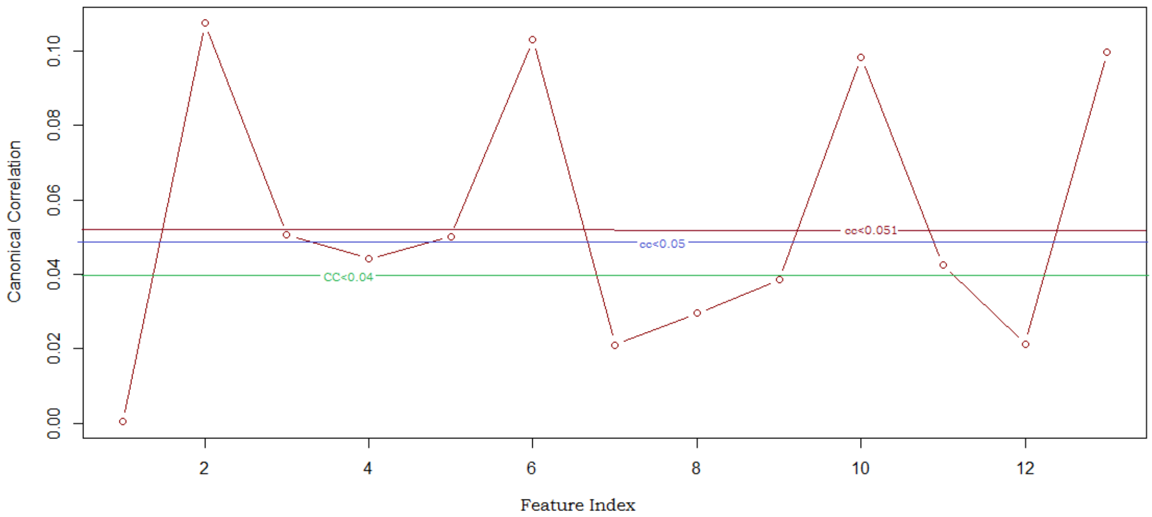

Figure 5.2: A line chart that representing the attributes scope under different canonical correlation thresholds

5.5.1 Performance Analysis

Total records Tested 30% (90 records) (70 fault inclined records and 20 hale records) cc threshold is less than 0.04.

Total number of records found with SCFI less than lower bound 21 (false negative 4 (fault inclined records claimed as hale) and true negative 17 (claimed records hale that are labeled hale)).

Total number of records found with SCFI greater than lower bound is 69 (true positives 66 and false positives 3).

As per the results explored, the proposed model is accurate to the level of 92.3%. The failure percentage is 4.3%.

The experiments also conducted on the same dataset under canonical correlation threshold <0.05 and 0.051 and the results are as follows:

Total records Tested 30% (90 records) (70 fault inclined records and 20 hale records) cc threshold is less than 0.05.

Total number of records found with SCFI less than lower bound 18 (false negative 2(fault inclined record claimed as normal) and true negative 18 (claimed records normal that are normal)).

Total number of records found with SCFI greater than lower bound is 72 (true positives 68 and false positives 2).

As per these results, the accuracy of the SDP under canonical correlation threshold of 0.05 is 95.6%.

Total records Tested 30% (90 records) (70 fault inclined records and 20 hale records) cc threshold is less than 0.051.

Total number of records found with SCFI less than lower bound 20 (false negative 2(fault inclined record claimed as hale) and true negative 19 (claimed records hale that are hale)).

Total number of records found with SCFI greater than lower bound is 72 (true positives 68 and false positives 1).

As per these results, the accuracy of the SCFI under canonical correlation threshold of 0.05 is 96.5%.

The observed time complexity is scalable since the completion time is incrementing with the same ratio against the increase in features count due to higher CC threshold (see Figure 5.3).

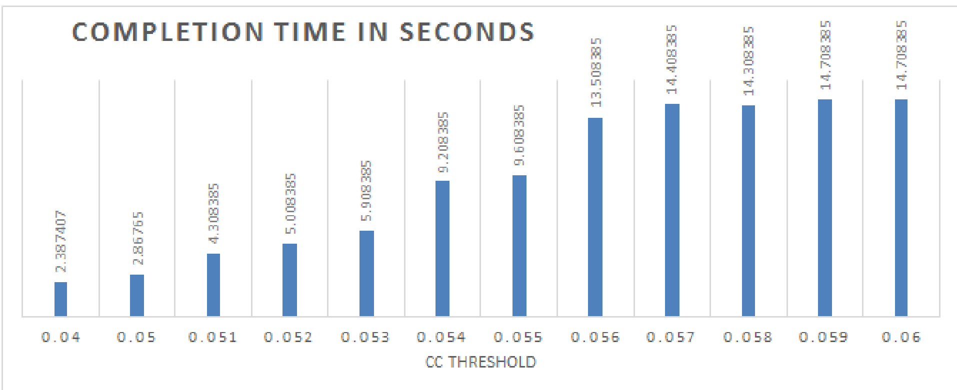

Figure 5.3: The completion time of defining Scale to Composition Fault Inclined (SCFI) under divergent canonical correlation thresholds

Hence it is obvious to conclude that the applying canonical correlation towards optimized attribute selection is significant to boost our earlier models [72], [73].

We used Composition Fault Inclined Prediction accuracy (the percentage of valid predictions by the proposed) as the primary performance measure. In addition to measuring accuracy, the precision, recall, and F-measure were used to measure the performance; these are defined using following equations.

$$pr = \frac{t_+}{t_+ + f_+} \quad \text{(Eq 5.10)}$$

Here in above Equation the pr indicates the precision, t_+ shows the true positives and f_+ indicates the false positive

$$rc = \frac{t_+}{t_+ + f_-} \quad \text{(Eq 5.11)}$$

Here in above Equation, the 'rc' indicates the recall, 'f_-' indicates the false negative.

$$F = \frac{2 * pr * rc}{pr + rc} \quad \text{(Eq 5.12)}$$

Here in the above Equation, 'F' indicates the F-measure.

Table 5.8: Precision, recall and F-measure values found from the results of the empirical analysis.

	Precision	recall	f-measure
<0.04	0.956521739	0.985074627	0.970588235
<0.05	0.971428571	0.971428571	0.971428571
<0.051	0.985507246	0.971428571	0.978417266

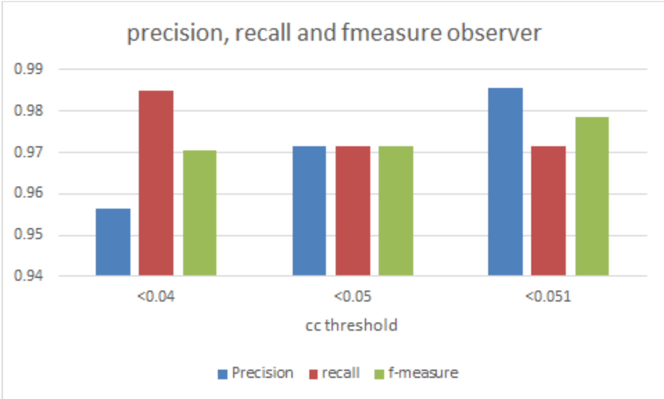

Figure 5.4: performance analysis of the prediction accuracy through Scale to composition fault inclined under divergent CC thresholds given.

5.6 CHAPTER SUMMARY

Canonical Correlation analysis of the composition's QoS attributes is devised here in this chapter, which is promising the selection of optimal attributes to simplify the process of defining Scale to Composition Fault Inclined. Our earlier research articles devised models [72], [73], which are amplified towards minimizing the process complexity by the model designed here in this chapter. The exploration of the results concluding that the canonical correlation analysis is promising and significant to select optimal attributes of the composition's QoS attribute dataset. Further, these optimal attributes are used to define the Scale Composition Fault Inclined (SCFI), which is observed to be robust, and is with minimal process complexity and retains the maximal prediction accuracy. In future, a scale to Hale Composition can devise, which is based on the QoS attributes of the compositions found to be hale.