

Derivation of Optimized Threshold of Semantic Alignment Metrics for Interpretation of Interoperability and Reusability of Cross-enterprise Vehicle Service Interface Models

Sangita De^{1,2}, Přemek Brada², Juergen Mottok³

¹Volkswagen AG, Ingolstadt, Germany

²Department of Computer Science and Engineering, University of West Bohemia, Pilsen, Czech Republic

³Faculty of Electrical Engineering and Information Technology, Ostbayerische Technische Hochschule (OTH), Regensburg, Germany

E-mail: sangita.de@outlook.de, sangita.de@cariad.technology, brada@kiv.zcu.cz, juergen.mottok@oth-regensburg.de

Keywords: interface, automotive services, semantic, ontologies, metamodels, framework, optimized threshold, metrics

Received: March 13, 2023

Over the past decade, cars have turned gradually into real cyber-physical systems. The collaboration of services between the service-oriented, cross-enterprise vehicle application frameworks has increased to generate novel, smart and complicated vehicle services. Consequently, from an interoperability perspective, semantically mapping of vehicle service components' interface ontological models emerged as a significant research interest in the automotive application domain that manipulates several cross-enterprise synergy knowledge application frameworks. Also, several semantic quality metrics have been defined over time for the vehicle service interface ontological models. The empirically evaluated values of these metrics can be used to assess progress in cross-enterprise interoperability between the service and the clients' APIs ontological models in the vehicle domain. However, despite the potential benefits of semantic alignment quality metrics, the effective use of these metrics for vehicle service interface ontologies has proven elusive. Yes, such metrics can be used successfully for quantification, but they mostly fail to provide adequate annotations in subsequent decision-making in semantic interoperability and reusability. In fact, the absence of an effective and meaningful threshold for the semantic similarity measure between various vehicle service interface ontological models motivates this research to propose a novel design approach to an optimized threshold derivation for the semantic similarity metrics. This threshold is then applied to a set of defined semantic alignment metrics for vehicle service frameworks. This paper uses a real-world vehicle domain industrial case study to illustrate the design approach. Through the considered case study, this research highlights the significance of optimized semantic alignment metric thresholds in determining the degree of cross-enterprise semantic interoperability and reusability between heterogeneous vehicle service frameworks' interface ontological metamodels.

Povzetek: V avtomobilski domeni je potreba po semantičnem preslikavanju ontoloških modelov vmesnikov komponent vozil narasla. Obstoječe metrike semantične kakovosti komponent vozila so izboljšane z vpeljavo optimiziranih pragov pri določanju semantične interoperabilnosti.

1 Introduction

Applications in the automotive domain are implemented as multiple distributed service components, and those service components call each other's Application Program Interfaces (APIs) for the complete application to function. From a modeling perspective, to ensure semantic interoperability and meaningful data exchange between heterogeneous vehicle services' API models, it is substantial to link the framework-specific vehicle service's APIs data at the semantic level using a shared domain vocabulary. Domain-specific shared vocabulary motivates the use of ontologies [5]. Ontologies represent vehicle service Software Component (SWC) interface metamodels' specifications schematically. A critical characteristic of automotive service API ontologies is that they mainly change over time regarding structure and

semantics. Changes in vehicle applications are often caused over time due to changes in requirements within the domain. Transition in the API ontologies can be due to one of the following scenarios [1]:

- Changes in the domain w.r.t time, cost, and requirements.
- Changes in the conceptualization result in changes in the construct or the structure of the ontology schemas.
- Changes in the explicit specifications include changes in property, attributes knowledge representation language, and the service interface version specification.

In the transition mentioned above scenarios, it is substantial to preserve the quality of the service API ontological models w.r.t semantic representation of the domain concepts. Therefore, metrics are defined and

evaluated to empirically measure the semantic alignment quality between the various interoperating vehicle application frameworks' interface ontological models [5]. However, the quantified results of semantic quality metrics are insufficient to ensure better interoperability and reusability decision-making. Therefore, domain experts must determine meaningful thresholds for each semantic alignment metric [1]. However, these metrics thresholds should not be solely based on domain expert assumptions but also on analysis of the metric distribution datasets. Furthermore, the automotive industry is frequently evolving, so the automotive application domain is subject to frequent changes w.r.t concepts and requirements in the context of API models. Therefore, the derived metrics thresholds are not intended to be universally valid for all the semantic similarity metrics in all application contexts within the domain [3]. Nevertheless, for semantic interoperability, the derivation of the thresholds provides adequate annotations on the semantic similarity metrics variability between vehicle application frameworks. Furthermore, it helps focus on a reasonable percentage of semantically similar application framework-specific interface models [1].

1.1 Contribution

The semantic quality metrics can be used successfully to quantify interoperability; however, from a cost, effort, and time optimization perspective, it is also essential to elevate the reuse of vehicle service API ontological models in case of semantic synergies based on derived thresholds [2]. Based on the static analysis of semantic similarity metrics distributions for various vehicle application frameworks' interface ontological metamodels, thresholds for the metrics are derived using a proposed methodology. However, in the automotive domain, no sound methods and metrics currently can support the interpretation of the semantic similarity and particularity values to determine whether two interoperating vehicle application framework services' interfaces are semantically similar [3]. Interpretations in such cases are frequently based on an implicit threshold or

an arbitrary value determined by domain experts based on application context [3].

With several requirements in mind to avoid the problems associated with the earlier thresholds earlier approaches in other knowledge domains, this research proposes a methodology to determine the thresholds for a defined set of semantic alignment quality metrics (including semantic cohesion and coupling metrics). These semantic alignment metrics are defined for vehicle application frameworks' interface ontological models [6][1]. The proposed methodology basically adheres to the following points [1]:

- The methodology is not recommended to be driven or assumed only by domain expert opinion but also by static analysis of metric distribution datasets.
- The methodology should respect the metric scale and the distribution ranges. The methodology should also be flexible against deviations in metrics values, service interface versions, and the automotive cooperative application complexity.
- In the frequently changing vehicle application domain, the methodology to derive and optimize the threshold value for semantic quality metrics should be robust enough to be repeatable, applicable, transparent, and pragmatic when applied to a wide range of semantic quality metrics [1][16].

This research considers a real-world typical vehicle domain case study to study the derived threshold's stability. The proposed methodology is further applied on a set of defined, pre-evaluated semantic alignment quality metrics distribution datasets [2] for the vehicle application frameworks' interface models that are part of the case study. The semantic alignment quality metrics defined in this paper are pre-evaluated manually in terms of percentages in literature work [2]. OWL2 (Web Ontology Language version 2.0) is used as a metamodeling language to describe the vehicle services' interface ontological metamodels [1].

References of Related Work used in this Subsection	Thresholds derived from expert experiences	Thresholds derived from metric distribution analysis	Thresholds derived from Error models analysis	Thresholds for semantic quality metrics in automotive domain	Thresholds as a measure to indicate degree of interoperability & reusability
[10], [5]	☑				
[16],[4], [6],[7],[9],[6],[1],[8]	☑	☑			☑
[11],[12]			☑		
Author's Contributions to the State of the Art					
Current research work, [2], [5]	☑	☑		☑	☑

Table 1: Summary table of related research works.

1.2 State of the art

This subsection briefly overviews previous research attempts to define metric thresholds, as illustrated in Table 1 [1].

Discussions. Various research works from different knowledge domains have defined metric thresholds based on researchers' experiences. For example, for the McCabe metric, the threshold value was defined as 10 [10], and for the NPATH metric, the threshold value was defined as 200 [5]. However, the threshold values derived from experience lack adequate scientific evidence to be reproduced or generalized. This research proposes a threshold derivation methodology that combines vehicle domain experts' experiences with real-world scientific assumptions and evidence, making the method robust and flexible to be refactored and generalized.

Research works also propose methodologies to derive optimum thresholds based on quality metric models distribution data analysis. For example, Erni and French et al. [16][4] proposed a multi-metrics methodology and use of mean (μ) and standard deviation (σ) methods to derive a threshold T based on object-oriented data [4]. The threshold T was evaluated as $T = \mu + \sigma$ or $T = \mu - \sigma$ where the high or the low metric values indicated potential problems, respectively. The disadvantage of this methodology is the assumption of normally distributed metrics or data normality. The assumption of only normalized metric distribution methodology does not consider open-world assumptions like, for example, the structural evolution of objects over time that might result in deviation in metric values, limiting the usage of this methodology. In fact, the methodology described in this work addresses open-world assumption w.r.t semantic alignment quality metrics for vehicle service API models.

The research work [7] defines absolute and relative thresholds for filtering metric data sets of values [6]. Statistics-based thresholds are derived from statistical analysis of metric values from a population sample. Much research has also proposed power laws as the preferred methodology for representing metric distributions in graphs to analyze relationships between classes and objects in an object-oriented system [9][6]. Similarly, Baxter et al. [8] also concluded that some analyzed software metrics dataset values follow a power law distribution. The study proposed that in-degree subclasses are a power law distribution and out-degree fields are not power law distributions. However, all these research works fall short of concluding how to use these complex distributions and the coefficients of these distributions to establish baseline values to judge systems. In contrast, this research work is focused on defining thresholds with direct applicability to differentiate framework-specific API models, judge semantic quality, and pinpoint problems.

The thresholds for quality metrics can also be derived using error model analysis. Shatnawi et al. [11] proposed using Receiver-Operating Characteristic (ROC) methodology to explore thresholds to predict the bugs in different error categories. However, there are two significant drawbacks to their derived results. Firstly, the

method to derive the threshold is not monotonic. Secondly, for every different release of Eclipse, different thresholds were derived, which implies weak stability of the derived threshold. The studies of Benlarbi et al. [12] show no empirical evidence for a defined threshold model that can be used to predict faults or errors. However, these results apply only to the specific error prediction model that the authors have used. In contrast to these works, this research paper's proposed monotonic methodology for the semantic similarity metric's threshold derivation ensures stability and flexibility against changes in vehicle service interface versions, SWCs' interface concepts, and the automotive cooperative system's complexity and sizes.

2 Design approach to optimum similarity threshold

This section provides an overview of the proposed design methodology to derive an optimized threshold for semantic alignment metrics. An automotive domain real-world industrial *case study* has also been used to demonstrate the proposed methodology [5][2].

2.1 Methodology to derive optimum similarity threshold

The design methodology to derive the optimum threshold is based on the static analysis of semantic alignment quality metric distributions datasets and is composed of three fundamental steps [3][17]:

1. In the first step, define two groups of vehicle service API ontological metamodel sources: *inter* and *intra-group*. Then, within each *intra-group*, define pairs of APIs ontological metamodel sources. In each *intra-group*, a framework-specific vehicle service API ontology is paired and semantically compared with a platform-agnostic, vehicle domain-specific *mediator* interface ontological metamodel. The mediator ontology is a framework-independent, generic vehicle service API ontology source; hence, it is a more abstract ontological metamodel compared to the framework-specific API ontological metamodel sources specifications. Therefore, this further implies that the ontology source pairs within each *intra-group* share less or a few semantic commonalities between their API traits. However, due to more concrete specifications, the framework-specific API ontological sources from different *intra-groups* share more synergies in their API semantic concepts or traits (also called *inter-group* similarities) when compared to the semantic similarity within each *intra-group*.
2. In the second step, for the above-defined *intra-group* sources, compute the semantic similarities between each pair of APIs ontological metamodel sources (i.e., the *intra-group* similarities) using the defined semantic similarity metrics. Then, aggregate all the metrics' lowest probability semantic similarity results to obtain an **IN** distribution representing less similar API semantic traits. Additionally, compute the *inter-*

group semantic similarities between each combination of an API ontology source from the first *intra-group* and an API ontology source from the second *intra-group* using the same set of semantic similarity metrics. Also, aggregate all the highest probability semantic similarity results to obtain an **IS** distribution representing more similar API semantic traits [3].

3. In the third step, compare the **IS** and **IN** similarity distributions. If, in the case, **IS** and **IN** distributions have no overlap between their acceptable data ranges (min, max) or interval, define the threshold τ_{opt} using any value between τ_{SL} (the lowest value of **IS**) and τ_{NH} (the highest value of **IN**). Else, in case of overlap in similarity distributions **IS** and **IN**, there are some false negatives (**FN_D**) and some false positives (**FP_D**) data values in the distributions, that is [3][17]:
 - a. Compute the proportion **FN_D** in the **IS** distribution for all samples of the similarity threshold between τ_{NH} to τ_{SL} . In this step, consider every value below the similarity threshold, τ_{SL} as **FN_D**.
 - b. Compute the proportion **FP_D** in the **IN** distribution for all samples of the similarity threshold between τ_{NH} to τ_{SL} . Consider every value above the similarity threshold, τ_{NH} , in this step as an **FP_D**.
 - c. Compute the average sum of the **FP_D** and **FN_D** proportions obtained in steps 3a and 3b. Each possible similarity threshold value, τ_{opt} , for each semantic alignment quality metric is considered at the point in the acceptable distribution range, where the average sum of **FP_D** and **FN_D** proportions is observed to be minimum.

Additionally, in the case where there is an overlap observed between the **IS** and **IN** metric distributions, there cannot be any **FN_D**s below τ_{SL} , but there will be some **FP_D**s above τ_{NH} as the distributions overlap. However, the derived values of the threshold by the given methodology are subject to dynamic changes based on the changes in the values of the semantic alignment quality metrics, which depend on the considered *case study* source's

interface specifications[1][16]. The workflow model for the above-proposed methodology on threshold derivation is illustrated in Figure 1[3].

2.2 Comparison of semantic similarity metric distributions for threshold derivation

The design methodology for the optimum threshold is based on semantic similarity metrics distributions, **IS** and **IN**. Figure 2. illustrates a conceptual, ideal example or an usecase for the semantic similar metric distribution ranges and corresponding threshold derivation, where the minimum value of the **S** distribution, τ_{SL} , is greater than τ_{NH} the maximum value of **IN** distribution[3]. In this ideal case, the highest value of all **IN** distributions is τ_{NH} , the value above which the two compared genes are similar, and the lowest value of all **IS** distributions is τ_{SL} , the value under which the two compared genes are non-similar [3].

A semantic similarity metric value greater than τ_{SL} implies that the service API semantic traits are similar for a given inter-group ontology source pair and that the source pairs are interoperable. Similarly, a semantic similarity value lower than τ_{NH} means that for a given *intra-group* ontology source pairs, the API semantic traits are non-similar. A semantic similarity metric value between τ_{SL} and τ_{NH} acceptable data interval means that the API semantic traits for the source pairs (*inter* or *intra*) are nearly similar, and thus, in this case, any value in the acceptable dataset interval range can be selected as the threshold, or it might also require expert opinion to interpret the optimum threshold in case of special usecases.

Figure 3. illustrates a conceptual, non-ideal usecase where the **IS** and **IN** distributions overlap, meaning that there are some **FP_D**s (that is, the metric values for the *intra-group* pair of sources from the **IN** distribution that are non-similar, but that have a similarity value greater than τ_{SL}) and **FN_D**s (that is the metric values for the *inter-group* source pairs from **IS** distribution that are similar but have a similarity value lower than τ_{NH}) [3]. In this case, a semantic similarity metric value lower than τ_{SL} means that the vehicle service API semantic traits for *inter-group* source pairs compared are non-similar.

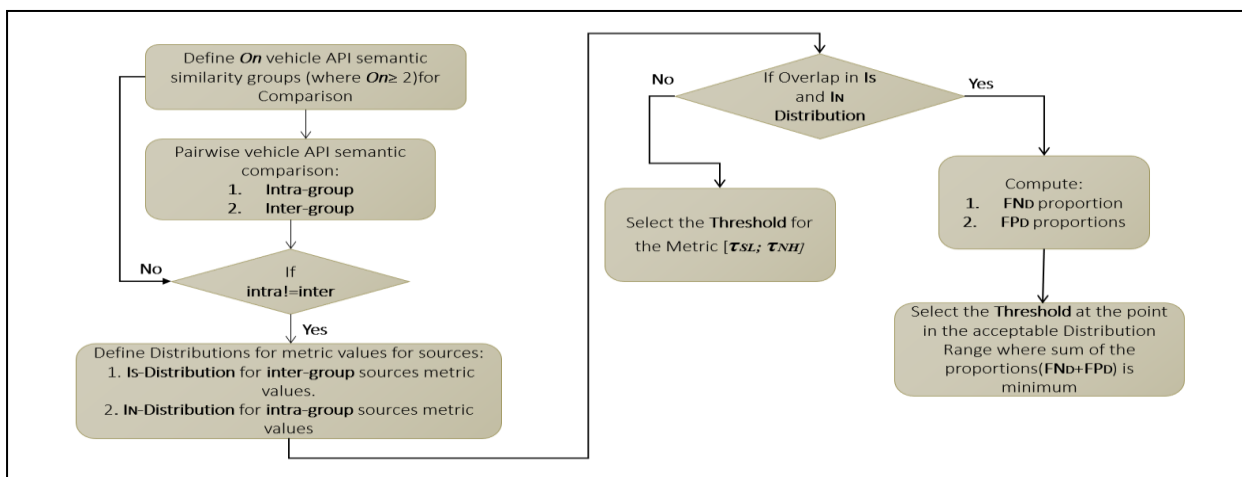


Figure 1: Workflow of optimum semantic similarity threshold derivation for metrics.

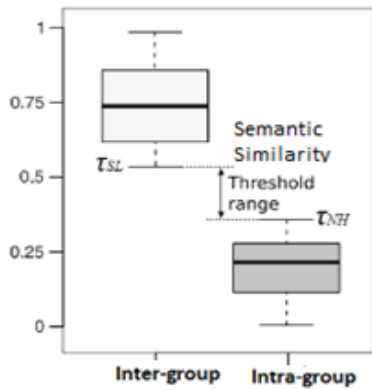


Figure 2: Ideal case for semantic similarity threshold.

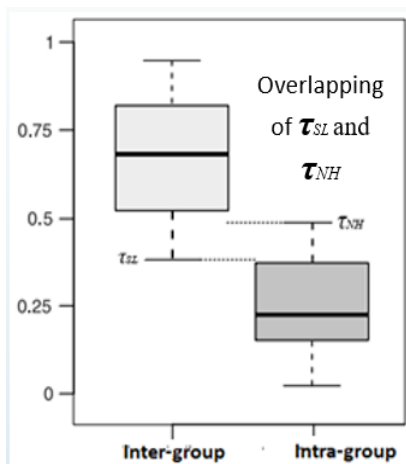


Figure 3: Non-Ideal case for semantic similarity threshold derivation.

Additionally, in this case a similarity value greater than τ_{NH} means that the service API semantic traits for *intra-group* source pair compared are non-similar. Also, as implied from Figure 3., if $\tau_{SL} < \tau_{NH}$, then as τ_{opt} gets

closer to τ_{SL} in the distribution acceptable dataset range, there will be more **FNDs** and fewer **FPDs**, and as τ_{opt} gets closer to τ_{NH} in the distribution's acceptable dataset range, there will be more **FPDs** and fewer **FNDs** [3].

2.3 An industrial case study

This subsection describes a typical automotive industrial *case study* to demonstrate the proposed design approach on optimized threshold derivation for semantic similarity metrics is *Keyless Vehicle Entry*, as illustrated in Figure 4. In this *case study*, the owner of a car wants to give the vehicle access to someone just by using his mobile phone, and the owner of the car is geographically located far away from his car [5][2]. This *case study* involves service collaborations from third-party, cross-knowledge domain platforms such as Robotics, Telematics, Infotainment, Cloud, etc.

To simplify the illustration of the given *case study*, we consider the three most used cross-enterprise vehicle application component frameworks that are used as service collaborators to realize this complex *case study*. They are namely, AUTOSAR Adaptive, Franca (from Genivi), and ROS2. To explore the cross-enterprise semantic alignment, the service component's interface models of the three frameworks mentioned above are modelled as ontological metamodel schemas for semantic mapping[13][15], as illustrated in Figure 5.

To bridge the semantic gap between the heterogeneous vehicle service API's ontological metamodel sources of the given *case study*, a vehicle domain-specific, platform-agnostic, *mediator* API ontological metamodel, namely, **DM**, is used in the current scope[13][5][15]. The degree of semantic alignments based on *TBox* axioms (*asserted* and *inferred*) between the application framework-specific interface ontological metamodel's concepts, relations (e.g., *is-a*), and properties are empirically evaluated using metrics.

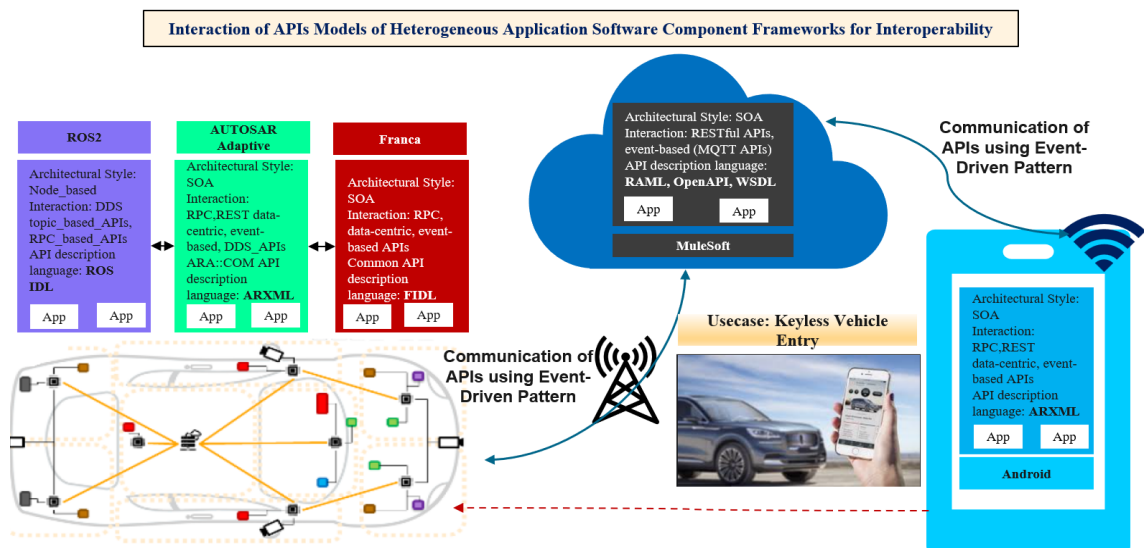


Figure 4: An industrial automotive domain case study.

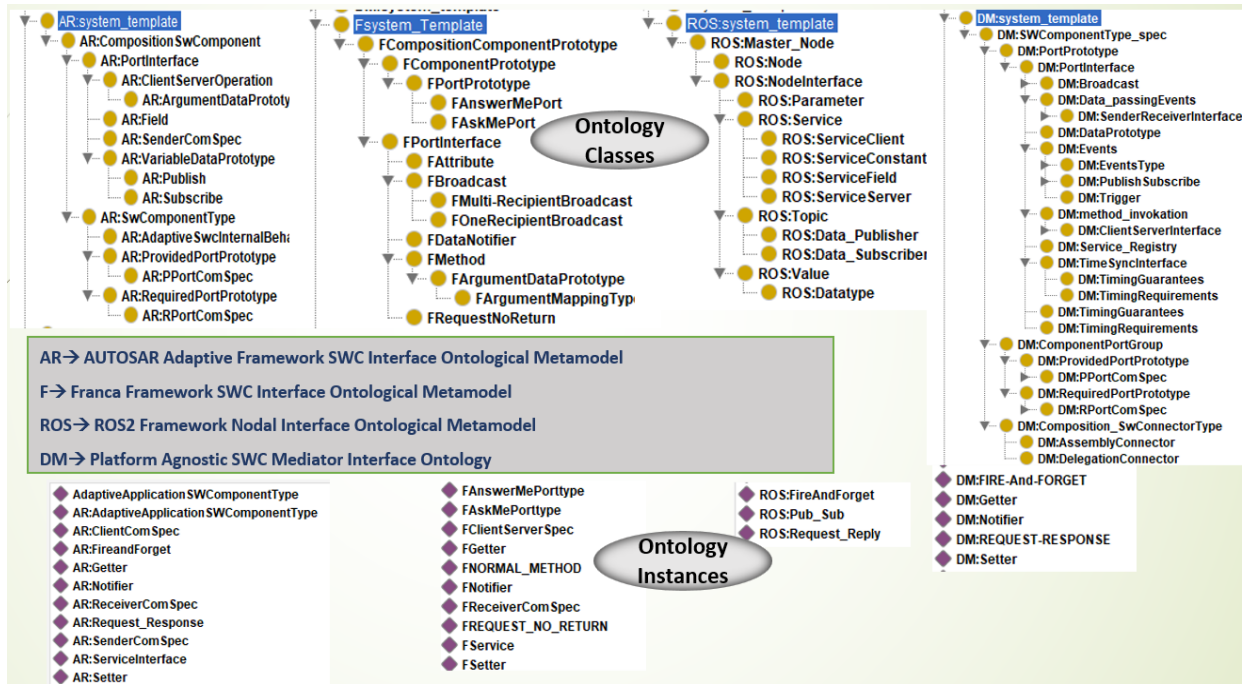


Figure 5: Abstract representation of case study using conceptual vehicle interface ontological metamodels.

Additionally, the degree of semantic similarity based on *ABox* axioms (*asserted* and *inferred*) between the interface ontological metamodels' class instances are also empirically evaluated at the instance or knowledgebase level using semantic similarity metrics [2][5][13].

Regarding the considered case study, each framework-specific interface ontological metamodels of AUTOSAR Adaptive, Franca, and ROS2 application frameworks are firstly paired with the *mediator* interface ontology, **DM**, as *intra-group* source pairs. Later, each of the above framework-specific interface ontological metamodels is paired with each other as *inter-group* source pairs [2][5][3].

However, this *case study* also includes service contributions from service-providing SWCs of Android and MuleSoft (for Amazon Web Services) application frameworks, which are not considered in the current scope.

3 Application of methodology to semantic similarity metrics for determination of optimum threshold

Based on an earlier literature work on the evaluation of semantic similarity metrics [2], this subsection briefly defines and provides an overview of a set of three semantic similarity quality metrics, namely **SSS**, **IRR**, **CIC** used for the evaluation of semantic alignment depth between interoperating vehicle application frameworks service API's ontological metamodels[13][14][17].

The proposed methodology on threshold derivation is also applied to these three given pre-evaluated semantic similarity quality metrics. The application of the methodology to derive a threshold can be dynamically

adapted based on changes in application framework interface model concepts and platform specifics[1][2].

As mentioned in the earlier subsection 2.3, the framework-specific interface ontological metamodels as part of the *case study* are paired as *inter* and *intra-group* source pairs. That is, for the *intra-group* source pairs, **Source 1** includes AUTOSAR Adaptive framework-specific SWC interface ontology, and the *mediator* ontology, **DM**, **Source 2** includes Franca framework-specific SWC interface ontology and the *mediator* ontology, **DM**, **Source 3** includes ROS2 framework-specific nodal interface ontology and the *mediator* ontology, **DM** [2].

Similarly, each framework-specific service API ontology within the scope of the *case study* is then paired with each other as *inter-groups* source pairs to identify the possible semantic similarities between each source pair[3].

That is, *inter-group Source 1* includes AUTOSAR Adaptive framework-specific SWC interface ontology and the Franca framework-specific SWC interface ontology, **Source 2** includes Franca framework-specific SWC interface ontology and the ROS2 framework-specific nodal interface ontology, **Source 3** includes AUTOSAR Adaptive framework-specific SWC interface ontology and the ROS2 framework-specific nodal interface ontology[3][2].

The greater the number of *inter* and *intra-group* pairs of APIs ontological metamodel sources used for semantic similarity metrics evaluation and the corresponding threshold derivation, the greater is the stability of the derived threshold. Consequently, the reliability of using such a derived threshold as a benchmark is greater in defining semantic interoperability and reusability of the vehicle service API ontological metamodel sources [3][1].

3.1 Derivation of optimized threshold for semantic similarity schema (SSS) metric

Definition: For a given automotive SWC API ontology schema, say, O_i , the Semantic Similarity Schema (SSS) metric is represented as the percentage of the fraction of the number of semantically *equivalent* (*is-a* relationships) schema classes (EQV) that is *equivalent* to the classes of other SWC APIs ontology schema when semantically compared to the total number (TRC) of the schema classes in O_i . TRC may include inheritance classes (IHCL), noninheritance classes (INHCL), as well as EQV in the given schema O_i [2][5][14].

$$SSS = EQV / TRC \tag{1}$$

Where $TRC = IHCL + INHCL + EQV$ in (1).

With the proposed methodology on threshold derivation, the threshold for the Semantic Similarity metric, **SSS** can be derived based on the evaluated metric values for the different vehicle service API ontology source pairs of the given *case study*.

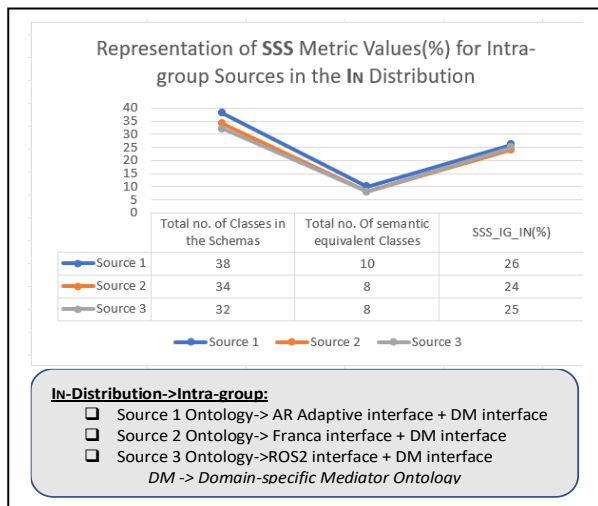


Figure 6: Overview of IN distribution for SSS metric.

For the *intra-group* source pairs with the mediator Ontology, **DM**, the evaluated values of the metric are represented in the respective **IN** distribution, following step 2 of the proposed methodology. As observed from Figure 6., for the *intra-group* sources, the maximum value of **SSS** metric in the **IN** distribution, τ_{NH} , is 26 on a scale of 100. Similarly, for the *inter-group* source pairs, the evaluated values of the **SSS** metric are represented in the respective **IS** distribution, also as seen in Figure 7. In Figure 7., it can also be observed that for the *inter-group* sources, the minimum value of **SSS** metric in the **IS** distribution $\tau_{SL} = 27$. The **SSS** metric can be used as a cohesion interface semantic similarity metric[2][5].

As illustrated in Figure 6. and Figure 7. shows no overlap between the **IS** and the **IN** distributions for the **SSS** metric dataset distribution ranges (min, max). Therefore, as specified in step 3 of the proposed methodology, the

optimum threshold, τ_{opt} , for the **SSS** metric can be considered as any value between τ_{SL} and τ_{NH} .

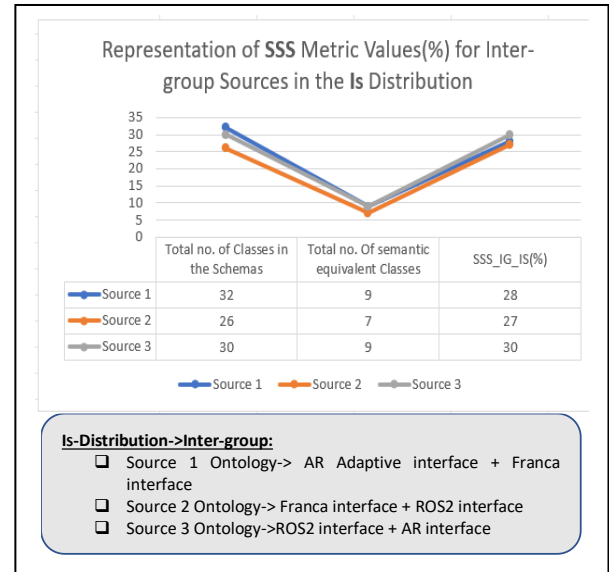


Figure 7: Overview of IS distribution for SSS metric.

Hence, based on the static analysis of the distribution acceptable dataset ranges, the center value between τ_{SL} and τ_{NH} is selected as the optimum threshold, that is τ_{opt} for **SSS** Metric for all the given *inter* and *intra-group* source pairs is selected as 26.5 [3].

3.2 Derivation of optimized threshold for instance relationship richness metric

Definition: For a given automotive SWC API ontology schema, say, O_i , the Instance Relationship Richness Metric (**IRR**) metric represents the depth of the knowledgebase[14] and is represented as the percentage of the fraction of the total number of *sameAs* instances (ISA) of schema classes in O_i compared to the total number of individuals of the given schema classes (TR_{Inst}) existing in O_i [2]. TR_{Inst} may include *sameAs* and *differentFrom* individuals (IDF) of the ontological schema classes. **IRR** is a cohesion semantic similarity metric [5].

$$IRR = ISA / TR_{Inst} \tag{2}$$

Where $TR_{Inst} = ISA + IDF$ in (2).

Likewise, the **SSS** metric for the *intra-group* sources (**Source 1**, **Source 2**, and **Source 3**), the evaluated values of the **IRR** metric for the ontology sources of the given *case study* are represented in the respective **IN** distribution with $\tau_{NH} = 33.33$ on a scale of 100, as seen in Figure 8.

Also, for the *inter-group* sources, the evaluated values of the **IRR** metric are represented for the source pairs in the respective **IS** distribution, following step 2 of the proposed methodology[3][5].

As seen in Figure. 9, the minimum value of **IS** distribution for **IRR** metric, $\tau_{SL} = 35$ on a scale of 100. As seen in Figure 8. and Figure 9., the metric values for the source pairs (*inter* and *intra-group*) for the **IRR** metric

have an overlap between the maximum and the minimum values of the distributions, **IS** and **IN**.

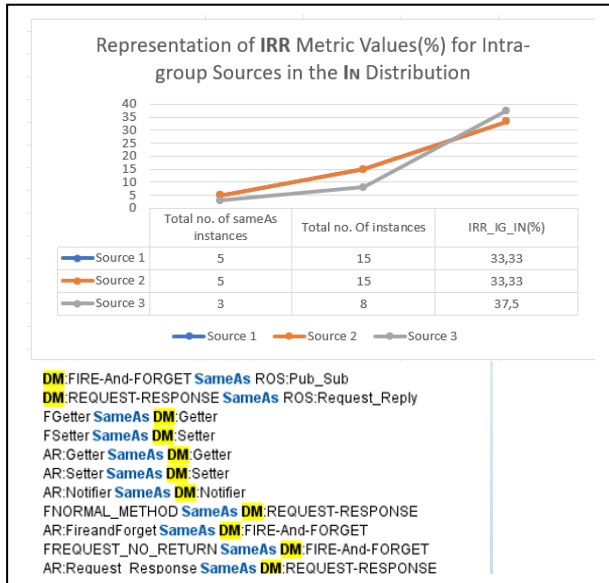


Figure 8: Overview of **IN** distribution and instance semantic similarity relation examples for **IRR** metric.

Due to the overlap in **IRR** metric distributions, **IS** and **IN**, the optimum threshold, τ_{opt} , for the **IRR** metric can be derived at the point in the acceptable dataset range in the distributions that minimize the average sum of **FNDs** and **FPDs** proportions[3][17].

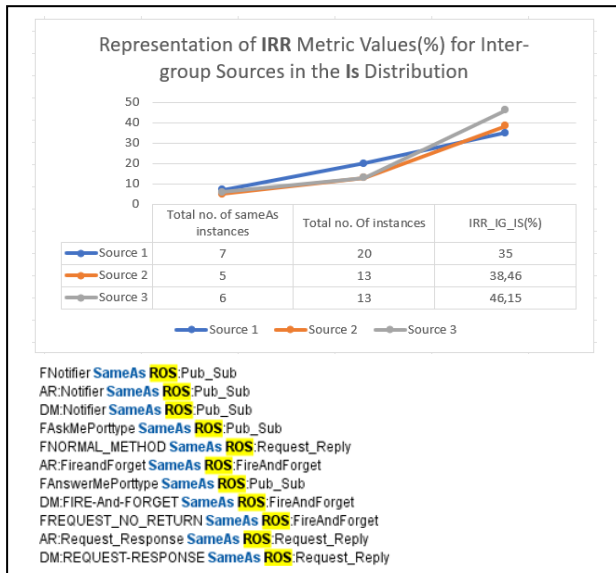


Figure 9: Overview of **IS** distribution and instance relationship examples for **IRR** metric.

Therefore, as specified in steps 3a and 3b of the proposed methodology and as illustrated in Figure 10., for each possible value within the acceptable distribution dataset range of the **IRR** metric, the sum of the **FPD** and **FND** proportions is calculated [3]. In the **FPD** proportion, **Fp**, is calculated for all the **IRR** metric values in the respective **IN** distribution for the *intra-group* sources, where **FPD** is expressed as absolute floating numbers.

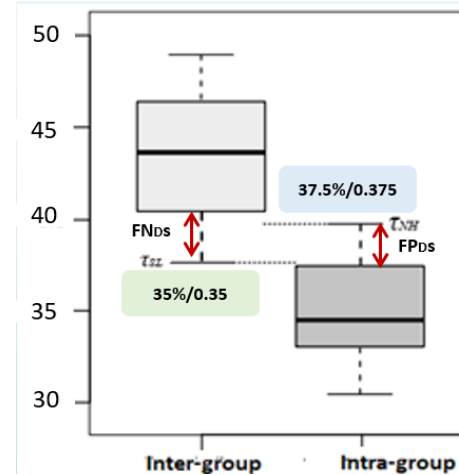


Figure 10: Overview of overlap between **IS** and **IN** distributions of **IRR** metric.

Similarly, in the **FND** proportion, **F_n** is also calculated for all the **IRR** metric values in the **IS** distribution for the *inter-group* sources and is expressed as absolute floating numbers, as illustrated in Table 2. The absolute (abs) value of τ_{opt} , for the **IRR** metric can also be derived by an equation as seen in equation (3).

$$\tau_{opt} \text{ (abs)} = \text{round} ((\text{float} (\mathbf{F}_n) + \text{float} (\mathbf{F}_p)) / 2) \quad (3)$$

F_n (absolute floating Value)	F_p (absolute floating Value)	$\tau_{opt} = \text{round}((\mathbf{F}_n + \mathbf{F}_p) / 2)$
0.35	0.375	0.36
0.36	0.365	0.363
0.365	0.37	0.37
0.355	0.355	0.355

Table 2: Calculation of **F_n** and **F_p** from **FND** and **FPD** Proportions of **IS** and **IN** distributions.

It is also important to understand here that if the selected τ_{opt} for **IRR** metric gets closer to τ_{SL} then there are more **FNDs** and fewer **FPDs**, and as τ_{opt} gets closer to τ_{NH} there are more **FPDs** and fewer **FNDs** [3]. Therefore, for the **IRR** metric based on the least average sum of **F_n** and **F_p**, the τ_{opt} is selected as 0.363, an absolute value for the threshold representing 36.3 % on a scale of 100, also highlighted in gray in Table 2 above.

3.3 Derivation of optimized threshold for class instance connectivity metric

Definition: For a given automotive SWC API ontology schema, say, O_i , the Class Instance Connectivity (CIC) metric is represented as the percentage of the fraction of the number of *sameAs* individuals of semantically equivalence schema classes (EQC) existing

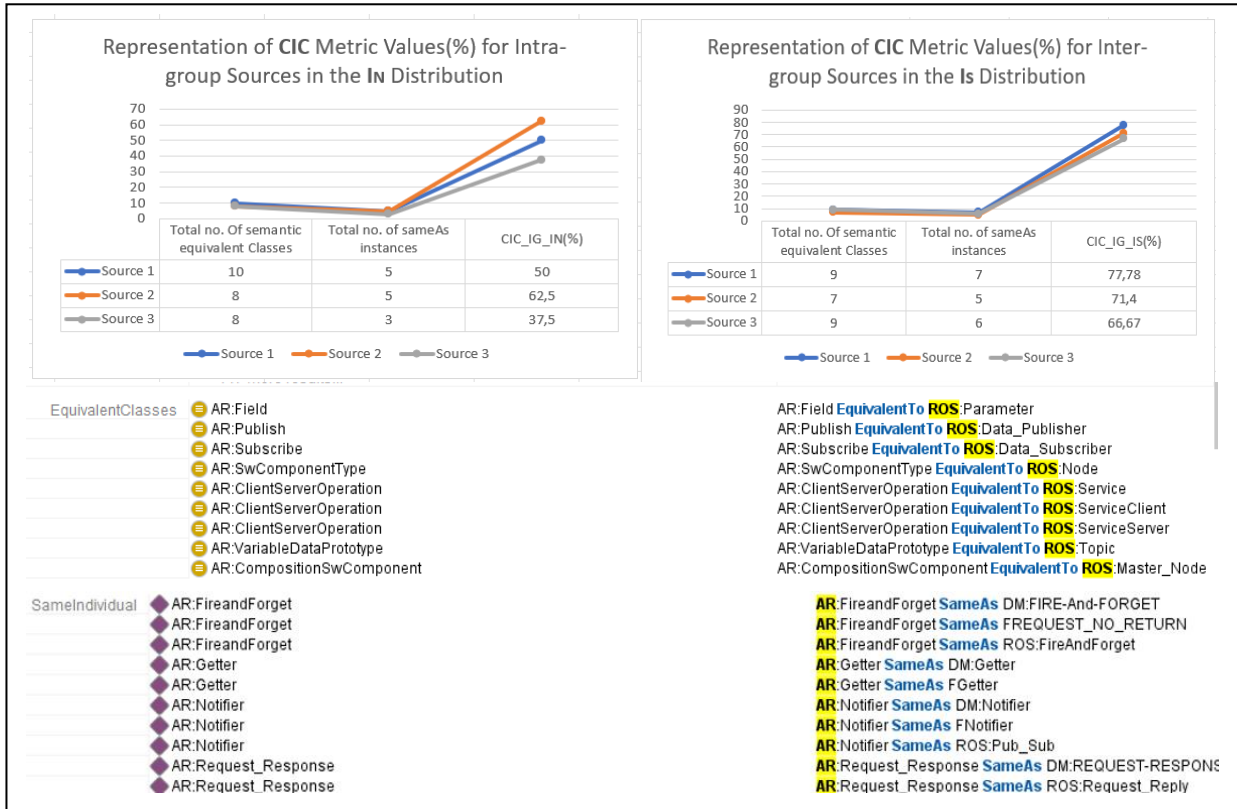


Figure 11: Overview of Is and IN distribution and class instance connectivity relationship examples for CIC metric.

in O_i compared to the total number of semantically equivalent schema classes (TRQ) existing in O_i . The CIC metric fundamentally indicates the fraction of schema class equivalence relationships that are being utilized at the individual level of the schema [2][5][14].

$$CIC = EQC / TRQ \tag{4}$$

Where $EQC \propto (sameAs\ INDIVIDUAL(O_i))$ in (4).

Like the SSS and IRR metrics, the threshold for the CIC metric can be derived based on the above-specified methodology for the different service API ontological metamodel sources of the given case study[5][2]. For the intra-group source pairs [3], the evaluated values of the CIC metric for the source pairs with mediator Ontology, DM, were represented in the respective IN distribution, as seen in Figure 10. Also, for the inter-group sources [3], the evaluated values of the CIC metric were represented for the source pairs of the given case study w.r.t each other in the Is distribution, as illustrated in Figure 11.

Figure 11 shows that for the intra-group sources, the maximum value of CIC metric in the IN distribution $\tau_{NH} = 62.5$. Similarly, it can also be observed that for the intra-group sources[3], the minimum value of CIC metric in the Is distribution, $\tau_{SL}=66.67$ on a scale of 100. Like IRR and SSS metrics, CIC metric can be used to measure interface semantic similarity in semantic cohesion and coupling scopes.

Hence, as illustrated in the above Figure 11., for the CIC metric, there is no overlap between Is and the IN distribution data ranges (min, max). Therefore, following

the step 3 of the proposed methodology[3][17], the optimum threshold, τ_{opt} , for the CIC Metric can be considered as any value between τ_{SL} and τ_{NH} , an ideal case for threshold derivation, as illustrated also in Figure 2. Hence, based on the static analysis of the Is and the IN distribution acceptable dataset ranges, the centre value in the acceptable dataset range that is between τ_{SL} and τ_{NH} is selected as the optimum threshold. The τ_{opt} for CIC metric for all the given sources (inter and intra-group) is selected as 65.

3.4 Results and discussions

To understand better the effect of the derived optimum semantic similarity threshold on reusability and interoperability between the vehicle service API ontological metamodel sources, this subsection provides comparative analysis of the results. For the SSS metric, based on manual static analysis of metric values distributions, the derived optimum threshold is $\tau_{opt} = 26.5$, which further means that the respective metric values for all the inter-group sources, namely, Source 1, Source 2 and Source 3 of the given case study are above the derived τ_{opt} value[1][3]. The derived threshold value for the SSS metric implies that all the given inter-group sources are semantically similar at ontology schema level. Moreover, in the cases of semantic synergies in their service interface concepts, properties and attributes, the interface ontological metamodel source pairs can be reused [16] instead of each other in processes like semantic integration for developing complex automotive applications. This would reduce cost, effort and avoid redundancy of data.

For the *intra-group* source pairs, based on the derived optimum threshold value for **SSS** metric, it can be considered that except for **Source 1**, the other two sources, namely **Source 2** and **Source 3**, are not semantically interoperable with each other at the ontology schema level. Based on manual static analysis of metric distributions, namely, **Is** and **In**, the optimum threshold, τ_{opt} , for the **IRR** metric is derived as 36.3, which means that the respective metric value for the *intra-group* sources, namely, **Source 3**, is above the derived threshold, τ_{opt} value and for **Source 1** and **Source 2**, the metric values are below the derived threshold value [1][3][16]. This further implies that the **Source 3** interface ontological metamodel source pairs within the *intra-group* **Source 3** are semantically similar to each other, whereas the interface ontological metamodel source pair within the *intra-group* **Sources 2** and **1** are not semantically similar to each other.

Due to an overlap in the **Is** and the **In** distributions data ranges (min, max) of the **IRR** metric, for the *inter-group* sources, namely, **Source 2** and **Source 3**, the metric values are above the derived threshold, τ_{opt} value whereas, for the **Source 1** the metric value is below the derived threshold value. This further implies that the interface ontological metamodel source pairs within the *inter-group* **Sources 2** and **3** are semantically similar and interoperable with each other and can be reused for each other in processes like semantic integrations for developing automotive applications [16][14][13].

The optimum threshold, τ_{opt} , for the **CIC** metric based on static analysis of metric distributions is derived as 65, which means that the respective metric values for all the *inter-group* sources, namely, **Source 1**, **Source 2**, and **Source 3**, are above the derived threshold, τ_{opt} , value. This further implies all the interface ontological metamodel source pairs within each of the given *inter-group* source pairs have strong semantic similarity between their conceptual classes at the schematic level and semantic similarity between their individuals at the instance level when compared with each other [3].

No overlap was observed between the **Is** and the **In** metric distribution data ranges (min, max) for the **CIC** metric. Hence, for all the given *intra-group* sources, the metric values are below the derived *threshold*, τ_{opt} , a value that implies all the interface ontological metamodel source pairs within each of the given *intra-group* sources have poor semantic similarity between their conceptual classes and individuals at ontology schema and instance levels when compared with each other.

Limitations. The threshold determination for a semantic similarity measure is limited by the number of semantic similarity metrics, vehicle application frameworks, and corresponding annotations that are available to study the threshold stability. In fact, there can be several possible variations in the quantity, granularity, and reliability between different semantic similarity metrics distributions. Consequently, it becomes difficult to determine a stable and universal threshold for all the semantic similarity quality metrics within the vehicle application domain [3]. Therefore, although the thresholds that are derived in this research work are application

agnostic, however, in an ever-evolving vehicle domain, it is preferable to recompute the threshold on a new set of semantic similarity metrics based on the application context and annotations that are available [1][6].

Also, with the proposed methodology, the appropriate choice of **Is** and **In** source pair groups and the corresponding distributions' analysis, which is crucial in the threshold determination process, requires some degree of domain expertise. In the absence of a proper degree of knowledge in the vehicle domain, it is difficult to interpret meaningful information using the threshold. Hence, in such cases, the thresholds are not accurate for semantic similarity measures [3][1].

4 Conclusion

The concept of semantic interoperability in automotive software engineering is notably elusive. This research proposes a methodology for determining a threshold for interpreting meaningful information based on the dataset values in the semantic alignment metric distributions. To study the stability of the derived optimized threshold, the proposed methodology was applied to a set of pre-evaluated interface semantic alignment cohesion and coupling metrics that are defined for the vehicle application framework's interface ontological metamodels [2][5].

The interface semantic alignment metric considered in this research work has been empirically evaluated in earlier literature [2] to measure the semantic alignment quality between vehicle application framework interface ontological metamodels. This work mainly focuses on the derivation of the threshold for the given set of semantic similarity metrics applicable to ontological metamodels at the instance and schema levels. This work derives a semantic similarity threshold for metrics based on comparing distributions of semantic similarity values of pairs of service interface ontological metamodels and non-similar service interface ontological metamodels. A real-world vehicle domain *case study* was considered in the research scope to demonstrate the proposed methodology on threshold derivation [2][5]. The *case study* also reveals the impact of the derived threshold on the vehicle services' semantic interoperability and reusability.

Due to frequently evolving vehicle domain applications' SWCs requirements, corresponding changes in interface ontological metamodel concepts and semantic alignment metrics evaluation methods and metric distributions, it can be claimed that although the proposed methodology on threshold derivation for semantic similarity measure is monotonic, however, the derived threshold values are not universal and are subject to changes in the future. In fact, in such cases, before using the proposed methodology, every service user should first check whether the original metric values distribution of the given semantic alignment metrics is still relevant in their own application's context or not [3].

This work derives the threshold for all the given sets of semantic similarity metrics as a percentage on a scale of 100. The values of the derived thresholds vary in each semantic similarity metric context. With this work, we

also propose even an optimum threshold derivation in the worst-case scenarios by minimizing the proportions of false-positive and false-negative similarity matches when there is an overlap in similar and non-similar distributions for a given interface semantic similarity metric.

From a time, cost, and effort estimation perspective, the results of this work can help automotive software engineers evaluate and interpret vehicle services' interfaces' semantic interoperability and reusability in a more meaningful way through meaningful information interpretation. Nevertheless, as a future work avenue, it would be interesting to apply the proposed methodology on threshold derivation to suites of semantic similarity metrics proposed by other vehicle domain experts or to metrics for other types of software models used in the vehicle domain for semantic similarity measure, for example, ecore models, etc.

References

- [1] T. L. Alves, C. Ypma and J. Visser, "Deriving metric thresholds from benchmark data," *2010 IEEE International Conference on Software Maintenance*, Timisoara, Romania, 2010, pp. 1-10, doi: 10.1109/ICSM.2010.5609747.
- [2] S. De, J. Mottok, P. Brada (2023), "Evaluation of Semantic Interoperability of Automotive Service API Models Based on Metamodels Similarity Metrics Using a Semi-automated Approach," In: Miraz, M.H., Southall, G., Ali, M., Ware, A. (eds) *Emerging Technologies in Computing. iCETiC 2022. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol 463. Springer, Cham, https://doi.org/10.1007/978-3-031-25161-0_2.
- [3] C. Bettembourg, C. Diot, O. Dameron, "Optimal Threshold Determination for Interpreting Semantic Similarity and Particularity: Application to the Comparison of Gene Sets and Metabolic Pathways Using GO and ChEBI." *PLOS ONE* 10(7): e0133579, 2015, <https://doi.org/10.1371/journal.pone.0133579>.
- [4] V. A. French, "Establishing software metric thresholds," *International Workshop on Software Measurement (IWSM'99)*, 1999.
- [5] S. De, P. Brada, J. Mottok & M. Niklas, (2021), "The Empirical Evaluation of Semantic Alignment Quality Metrics for Vehicle Domain Component Frameworks Interface Ontologies," In: *The International FLAIRS Conference Proceedings*, 34. <https://doi.org/10.32473/flairs.v34i1.128512>.
- [6] K. A. Ferreira, M. Bigonha, R. D. Bigonha, L.F. Mendes & H.C. Almeida (2012). Identifying thresholds for object-oriented software metrics, In: *The Journal of Systems and Software*, 85, 244-257.
- [7] M. Lanza, R. Marinescu (2006). *Object-Oriented Metrics in Practice: Using Software Metrics to Characterize, Evaluate, and Improve the Design of Object-Oriented Systems*. Springer-Verlag, Germany.
- [8] Baxter, G., Frean, M., Noble, J., Rickerby, M., Smith, H., Visser, M., Melton, H., Tempero, E., 2006. Understanding the shape of java software. In: *OOPSLA'06, OR, Portland, USA*.
- [9] R. Wheeldon, & S. Counsell, (2003), "Power law distributions in class relationships", In: *Proceedings Third IEEE International Workshop on Source Code Analysis and Manipulation*, pp. 45-54.
- [10] T. J. McCabe, "A Complexity Measure," in *IEEE Transactions on Software Engineering*, vol. SE-2, no. 4, pp. 308-320, Dec. 1976, doi: 10.1109/TSE.1976.233837.
- [11] R. Shatnawi, W. Li, J. Swain, and T. Newman, "Finding software metrics threshold values using ROC curves," In: *Journal of Software Maintenance and Evolution: Research and Practice*, 2009.
- [12] S. Benlarbi, K. E. Emam, N. Goel, and S. Rai, "Thresholds for object-oriented measures," in *ISSRE '00: Proc. of the 11th International Symposium on Software Reliability Engineering. IEEE*, 2000, p. 24.
- [13] L. Yonglin, Z. Zhi, L. Qun, "An ontological metamodeling framework for semantic simulation model engineering", *Journal of Systems Engineering and Electronics*, vol. 31, no. 3, pp. 527-538 (2020).
- [14] S. Tartir, I. B. Arpinar, and A. Sheth, (2010), "Ontological Evaluation and Validation", In: book: *Theory and Applications of Ontology: Computer Applications*: 115-130.
- [15] T. Poulain, N. Cullot, K. Yé tongnon, "Ontology mapping specification in description logics for cooperative systems", In: *Journal des sciences pour l'ingénieur (JSPI)*, vol. 7, pp.64-71. hal-00722532f (2006).
- [16] K. Erni and C. Lewerentz, "Applying design-metrics to object-oriented frameworks," In: *METRICS '96: Proceedings of the 3rd International Symposium on Software Metrics*. Washington, DC, USA: IEEE Computer Society, 1996, p. 64.
- [17] J.L. Sevilla, V. Segura, A. Podhorski, E. Guruceaga, J.M. Mato, L.A. Martinez-Cruz, "Correlation between gene expression and GO semantic similarity", In: *IEEE/ACM Trans Computational Biology and Bioinformatics*. 2005; 2(4):330–8. doi: 10.1109/TCBB.2005.50 PMID: 17044170.

