

Complex-Valued Representation of Relevance and Semantic Relations for Claim-Evidence Pairs with PSO Boundary Calibration

Fadilah Nur Imani^{1*}, Arsyelina Husni Johan²

Department of Mathematics, Faculty of Science and Technology, Universitas Terbuka
**051961504@ecampus.ut.ac.id; fadilahnurimani@gmail.com*

ABSTRACT

This study constructs $f(c,e)$, a complex-valued function in a polar form as a measure of relevance and semantic orientation between claims and evidence, where the modulus component represents semantic relevance and the angle component represents semantic orientation, which reflects the direction of the relationship between claims and evidence. The function is proven to have four mathematical properties: well-definedness, boundedness, continuity, and symmetry. Particle Swarm Optimization (PSO) is used to determine the angular boundaries that partition the semantic orientation space into support, not enough information, and refute relation zones. This approach ensures that the resulting zone boundaries reflect the characteristics of the empirical data distribution. The angular boundary calibration experiment was conducted using two datasets, namely SNLI and FEVER, and with two embedding models, namely all-MiniLM-L6-v2 and mxbai-embed-large-v1. The F1-macro objective was used for PSO-based angular boundary calibration, and the calibrated boundaries were subsequently evaluated on held-out test subsets. The held-out F1-macro values ranged from 0.41350 to 0.59385, with the combination of SNLI and mxbai-embed-large-v1 producing the highest value. In this experiment, zone separation was descriptively clearer for SNLI than for FEVER, based on the observed semantic orientation distributions.

Keywords: complex numbers, cosine similarity, numerical optimization, particle swarm optimization, polar representation

1. INTRODUCTION

Claim and evidence are two fundamental, interconnected elements used in various contexts of information-based decision-making, such as claim verification of supplement benefits using scientific evidence in the health sector [1] and rubric-based automated essay scoring in the education sector [2]. This indicates that the assessment of the relationship between claim and evidence is an emerging need that can be leveraged across domains.

Under ideal conditions, this assessment relies on human labels as the gold standard. However, the process requires significant resources and is subject to subjectivity, meaning that disagreements among experts can still occur [3]. Therefore, various automated metrics have been developed to replicate that assessment in a measurable way.

In Natural Language Processing, this need is formulated through several tasks, including Natural Language Inference (NLI), fact verification, and evidence

extraction, all of which require a component capable of representing the relevance and semantic relationship between claim-evidence pairs. When labeled data is available, this component can be trained using a Natural Language Understanding (NLU) framework to produce assessments that closely approximate human annotations [3]. Existing approaches typically represent the relationship between claims and evidence by measuring the degree of similarity between the two texts [4], [5], most commonly using cosine similarity [6], [7].

In general, cosine similarity value is directly used as a scalar measure of semantic similarity. This value can be used as a measure of linguistic and semantic alignment, thus being able to capture implied meaning [8]. In claim-evidence analysis, cosine similarity between the vector representations of claim and evidence can capture the degree of semantic relation [9], which in turn can be used as a measure of evidence relevance to the claim [10]. However, the cosine similarity value derived from learned vector representations may be influenced by model regularization and therefore may not always reflect meaningful semantic similarity [11]. The use of cosine similarity as a scalar value also has geometric limitations. In addition to that, the use of cosine similarity as a single scalar measure also presents geometric limitations. Previous work has shown that cosine similarity has a saturation zone, where changes in the angle between vectors result in only minor changes in cosine values [12]. This potentially reduces the sensitivity of cosine similarity in distinguishing between types of semantic relations when represented as a single scalar value. This limitation is related to the nature of cosine similarity as a scalar measure, which does not directly provide a formal representation that simultaneously encodes magnitude and orientation angle within a single mathematical object. Therefore, the angle formed between two vectors becomes an important aspect to investigate, as it can provide more explicit geometric orientation information than the cosine similarity value itself.

A representation that can formally encode the cosine similarity value as a magnitude and the cosine angle as an orientation component within a single mathematical object is a complex-valued function represented in polar form. Polar representations have been used in text processing to encode different aspects of textual information in separate components [13]. In addition, rotation in complex space has been used to model various relation patterns by leveraging the geometric properties of complex numbers [14]. Previous work has shown that optimizing the angle between text-pair representations in complex space can better explore semantic similarity within cosine similarity saturation zones [12]. These findings suggest that angle components in complex-valued representations may provide an alternative basis for analyzing semantic relations beyond what is captured by cosine similarity as a scalar measure. However, the application of such angle-based characteristics to claim-evidence relation interpretation has received limited attention. This raises the question of whether different claim-evidence relation categories exhibit distinct angle distributions. If such differences exist, boundaries separating relation zones may be determined empirically from the observed data distribution. Therefore, an empirical optimization approach is required. Particle Swarm Optimization (PSO) is an effective optimization method for a wide range of nonlinear functions [15]. PSO has been used for data-driven threshold optimization [16] and in combination with cosine similarity [17], [18], [19].

Despite these findings, limited attention has been given in the literature to a complex-valued function that represents the semantic relevance and relation between claims and evidence, combining both magnitude and orientation angle into a single

mathematical object. Similarly, the boundaries that separate semantic relation zones within this angular space have not yet been widely explored through an empirical optimization procedure. This study attempts to address these two gaps. First, a complex-valued function representing claim-evidence relevance and semantic relation is constructed in the manner described above. Second, PSO is applied to determine the angular boundary values that divide the semantic orientation space into distinct claim-evidence relation zones based on empirical data. The resulting function may offer a useful basis for further development of methods for analyzing claim-evidence semantic relations.

2. MATERIAL AND METHODS

This research combines two approaches aligned with its objectives: an analytical approach to construct a complex-valued function representing the semantic relevance and relation between claim-evidence pairs, and an empirical approach using PSO to calibrate the angular boundaries that divide this function's angular range into distinct relation types. The focus is on formulating the complex function and calibrating these boundaries, rather than on developing specific embedding models or datasets. The overall method is illustrated in Figure 1 below.

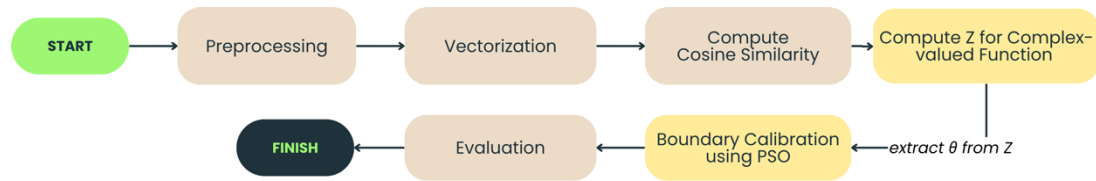


FIGURE 1. Overview of the proposed method, from preprocessing to evaluation.

2.1 VECTOR AND COSINE SIMILARITY

Let $\mathbf{c}, \mathbf{e} \in \mathbb{R}^n$ represent the vector representations of the claim-evidence texts, respectively. The cosine similarity between $\mathbf{c}$ and $\mathbf{e}$ is defined as follows [6]:

$$s = \text{sim}(\mathbf{c}, \mathbf{e}) = \cos \theta = \frac{\mathbf{c} \cdot \mathbf{e}}{\|\mathbf{c}\| \|\mathbf{e}\|} \quad (1)$$

Where $\|\mathbf{c}\|$ and $\|\mathbf{e}\|$ represent the Euclidean norms of $\mathbf{c}$ and $\mathbf{e}$, respectively, and $s \in [-1, 1]$ for all $\mathbf{c}, \mathbf{e} \neq 0$ [20].

2.2 TEXT REPRESENTATION TECHNIQUES

Text vectorization was performed using two sentence embedding models with different training approaches. First, all-MiniLM-L6-v2 [21], [22], which was trained contrastively on NLI data (entailment, neutral, contradiction). Second, mx-bai-embed-large-v1 [12], [23], which was optimized to distinguish angles between similar and dissimilar text pairs and is therefore theoretically more sensitive to subtle semantic

differences. Both models are configured to produce unit-norm vectors, allowing cosine similarity to be obtained directly from the dot product.

2.3 DATASETS

The experiments utilized test subsets from two datasets featuring distinct semantic relation schemes: SNLI [24] and FEVER [25], [26]. To avoid label distribution bias during PSO calibration, 3000 samples per label were drawn from each dataset, resulting in a total of 9000 pairs for each dataset.

SNLI consists of premise-hypothesis pairs labeled as entailment, neutral, and contradiction [24]. The dataset was selected because NLI is widely regarded as a fundamental framework for claim verification and has been used as an approach to claim verification when claims are evaluated against available evidence [1]. In this study, premises are treated as evidence and hypotheses as claims. Accordingly, entailment, neutral, and contradiction are interpreted as supports, not enough information (NEI), and refutes relations, respectively, following the use of NLI for claim verification [10]. These labels represent different semantic polarities, ranging from support to contradiction, with a neutral category in between, making SNLI a suitable proxy for evaluating semantic relation zones between claims and evidence.

FEVER, in contrast, directly annotates claim-evidence pairs with supports, not enough info (NEI), and refutes labels [25]. Although FEVER exhibits a similar three-category semantic polarity structure, it encodes factual claim-evidence relations rather than the logical premise-hypothesis relations represented in SNLI. The FEVER data used in this study were obtained from the NLI-style reformulation of FEVER [26]. In the reformulated dataset, the original FEVER claim is stored in the premise field, whereas the paired textual context is stored in the hypothesis field [26]. Several instances in the NLI-style FEVER data contain an empty hypothesis field, including instances labeled NEI. Since the present study analyzes semantic relations between claims and textual context, instances with an empty hypothesis were excluded, and only non-empty pairs were retained for the experiments. After this filtering step, 3000 pairs were sampled for each FEVER label to maintain a balanced class distribution.

2.4 COMPLEX-VALUED FUNCTION CONSTRUCTION

The embedding vectors were normalized during the vectorization process. Since $\|c\| = \|e\| = 1$, Equation (1) can be simplified as:

$$s = s(c, e) = c \cdot e \quad (2)$$

Based on (2), two components of the complex-valued function are defined as:

$$r = |s| \quad (3)$$

$$\theta = \arccos(s) \in [0, \pi] \text{ or } [0^\circ, 180^\circ] \quad (4)$$

Where r represents the semantic relevance between a claim and evidence as the intensity of geometric similarity, which is always non-negative. Previous studies have used cosine similarity as a measure of semantic similarity in claim verification and related semantic matching tasks [8], [10]. In this study, r in Equation (3) adopts the

same interpretation while preserving the non-negativity property required for a geometric magnitude. The value of θ represents semantic orientation, reflecting the direction of the relationship between a claim-evidence pair. Smaller values of θ indicate supporting relations, whereas larger values of θ indicate refuting relations.

Based on these two components, the complex-valued function $f: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{C}$ is constructed as follows:

$$f(\mathbf{c}, \mathbf{e}) = r \cdot e^{i\theta} = |s| \cdot e^{i \cdot \arccos(s)} \quad (5)$$

Equation (5) represents the function in Euler exponential form, where r as modulus represents magnitude and θ represents the orientation angle in the complex plane [27]. By applying Euler's formula, $e^{i\theta} = \cos \theta + i \sin \theta$, Equation (5) can be expressed in Cartesian form through trigonometric substitution as follows:

$$\cos \theta = \cos(\arccos(s)) = s \quad (6)$$

Since $\sin^2 \theta + \cos^2 \theta = 1$, it follows that:

$$\sin \theta = \sin(\arccos(s)) = \sqrt{1 - s^2} \quad (7)$$

Substituting (3), (4), (6), and (7) into Equation (5) yields the polar form in (8) and the Cartesian form in (9), respectively:

$$\begin{aligned} f(\mathbf{c}, \mathbf{e}) &= r \cdot e^{i\theta} = r \cdot (\cos \theta + i \sin \theta) \\ &= |s| \cdot (\cos(\arccos(s)) + i \sin(\arccos(s))) \end{aligned} \quad (8)$$

$$= |s| \cdot (s + i\sqrt{1 - s^2}) = |s| \cdot s + i \cdot |s| \cdot \sqrt{1 - s^2} \quad (9)$$

Equation (8) represents the polar form of the function, which is visualized in the complex plane in the Results section, whereas Equation (9) represents the Cartesian form used for numerical computation.

The value of θ provides a continuous measure of semantic orientation between claim-evidence pairs based on their embedding geometry. It does not directly determine the semantic relation label of a claim-evidence pair, such as supports, NEI, or refutes. To partition the semantic orientation space into these three relation zones, two angular boundaries, θ_s and θ_r , with $0^\circ < \theta_s < \theta_r < 180^\circ$, are required. These boundaries determine how the continuous semantic orientation space is partitioned into three relation zones, which are then used to assign the supports, NEI, and refutes labels to claim-evidence pairs. The values of θ_s and θ_r are determined empirically in the next section.

2.5 BOUNDARY CALIBRATION USING PSO

The angular boundaries θ_s and θ_r were calibrated empirically using PSO. A PSO variant with inertia weight was employed, where w as the inertia weight that controls

the balance between global exploration and local exploitation [28]. The optimization variables consist of the pair (θ_s, θ_r) , subject to the constraint $0^\circ < \theta_s < \theta_r < 180^\circ$. Particle positions were randomly initialized with $\theta_s \sim U(0^\circ, 70^\circ)$ and $\theta_r \sim U(60^\circ, 100^\circ)$, chosen based on the angle ranges relevant to the experimental data distribution, and updated iteratively according to:

$$v_{id} = w \cdot v_{id} + c_1 \cdot r_1 \cdot (p_{id} - x_{id}) + c_2 \cdot r_2 \cdot (p_{gd} - x_{id}) \quad (10)$$

$$x_{id} = x_{id} + v_{id} \quad (11)$$

The resulting semantic relation zones are defined as follows:

$$Zona(\theta) = \begin{cases} Supports, & \theta \in [0^\circ, \theta_s] \\ NEI, & \theta \in (\theta_s, \theta_r) \\ Refutes, & \theta \in [\theta_r, 180^\circ] \end{cases} \quad (12)$$

The objective function is the $F1_{macro}$ score over the semantic relation labels:

$$F1_{macro} = \frac{1}{3} \sum_{k=1}^3 \frac{2 \cdot precision_k \cdot recall_k}{precision_k + recall_k} \quad (13)$$

where $k \in \{Supports, NEI, Refutes\}$ represents the semantic relation zone index.

In the implementation, the optimization objective is expressed as the negative $F1_{macro}$ score, so minimizing the objective is equivalent to maximizing Equation (13):

$$minimize -F1_{macro}(\theta_s, \theta_r) \quad (14)$$

Calibration was performed separately for each dataset-embedding combination, resulting in four sets of (θ_s, θ_r) values. The PSO configuration consisted of a swarm size of $(n_{partikel}) = 30$, a maximum of $(n_i) = 200$ iterations, inertia weight $(w) = 0,9$, and coefficients $(c_1) = 2,0$ and $(c_2) = 2,0$.

2.6 EXPERIMENTAL PROCEDURES

The experiments were conducted separately for each dataset-embedding combination. The data for each dataset were divided into a calibration subset and a held-out test subset using an 80:20 stratified. The calibration subset was used exclusively for PSO-based angular boundary calibration, while the remaining 20% was reserved as the held-out test subset and was not used during boundary calibration.

For each dataset-embedding combination, the claim and evidence texts were encoded using the corresponding embedding model. The resulting embeddings were normalized before cosine similarity was calculated. For each claim-evidence pair, the cosine similarity s was used to obtain the modulus $r = |s|$ and the semantic orientation angle $\theta = \arccos(s)$, which were then used to construct the complex-valued representation.

PSO was applied only to the calibration subset to determine the angular boundaries. For each dataset-embedding combination, calibration was performed separately, resulting in four sets of angular boundary values. The PSO configuration consisted of a swarm size of 30 particles, a maximum of 200 iterations, an inertia weight of $(w) = 0,9$, and coefficients $(c_1) = 2,0$ and $(c_2) = 2,0$. The angular boundaries obtained from the calibration subset were then fixed and applied to the held-out test subset without further optimization.

3. RESULTS AND DISCUSSION

3.1 PROOF OF PROPERTIES OF COMPLEX-VALUED FUNCTION

Four mathematical properties of the complex-valued function $f(\mathbf{c}, \mathbf{e})$ constructed in the previous section are formally proven here: well-definedness, boundedness, continuity, and symmetry.

Proposition 1

The function $f(\mathbf{c}, \mathbf{e})$ is defined for all $\mathbf{c}, \mathbf{e} \in \mathbb{R}^n$ with $\mathbf{c}, \mathbf{e} \neq 0$.

Proof. Since $\mathbf{c}, \mathbf{e} \neq 0$ and the vector representations are normalized, $\|\mathbf{c}\| = \|\mathbf{e}\| = 1$, giving Equation (2): $s = \mathbf{c} \cdot \mathbf{e}$ with $s \in [-1, 1]$. Based on Equation (9), $f(\mathbf{c}, \mathbf{e}) = |s| \cdot s + i \cdot |s| \cdot \sqrt{1 - s^2}$. Since both the real component $|s| \cdot s$ and the imaginary component $|s| \cdot \sqrt{1 - s^2}$ are defined for $s \in [-1, 1]$, the function $f(\mathbf{c}, \mathbf{e})$ produces exactly one complex value for every $\mathbf{c}, \mathbf{e} \neq (0, 0)$ pair. ■

Proposition 2

For all $\mathbf{c}, \mathbf{e} \neq 0$, $|f(\mathbf{c}, \mathbf{e})| \in [0, 1]$.

Proof. Based on Equation (9):

$$|f(\mathbf{c}, \mathbf{e})| = \sqrt{(|s| \cdot s)^2 + (|s| \cdot \sqrt{1 - s^2})^2} = \sqrt{|s|^2 \cdot (s^2 + (1 - s^2))} = \sqrt{|s|^2} = |s|$$

Since $s \in [-1, 1]$, it follows that $|s| \in [0, 1]$, and therefore $|f(\mathbf{c}, \mathbf{e})| \in [0, 1]$. ■

Proposition 3

The function $f(\mathbf{c}, \mathbf{e})$ is continuous on $\{\mathbf{c}, \mathbf{e} \mid \mathbf{c}, \mathbf{e} \in \mathbb{R}^n; \mathbf{c}, \mathbf{e} \neq 0\}$.

Proof. Equation (2), $s = \mathbf{c} \cdot \mathbf{e}$ is continuous on $\{\mathbf{c}, \mathbf{e} \mid \mathbf{c}, \mathbf{e} \neq 0\}$ since $\|\mathbf{c}\| = \|\mathbf{e}\| = 1$. The real component in Equation (9), $|s| \cdot s$ is continuous as the product of continuous functions. The function $\sqrt{1 - s^2}$ is continuous for $s \in [-1, 1]$, so the imaginary component, $|s| \cdot \sqrt{1 - s^2}$, is likewise continuous as the product of continuous functions. Since both components are continuous, $f(\mathbf{c}, \mathbf{e})$ is continuous on its domain. ■

Proposition 4

For all $\mathbf{c}, \mathbf{e} \neq 0$, $f(\mathbf{c}, \mathbf{e}) = f(\mathbf{e}, \mathbf{c})$.

Proof. Since the dot product is commutative, $\mathbf{c} \cdot \mathbf{e} = \mathbf{e} \cdot \mathbf{c}$, it follows that $s(\mathbf{c}, \mathbf{e}) = s(\mathbf{e}, \mathbf{c})$. Based on Equation (9), $f(\mathbf{c}, \mathbf{e}) = |s| \cdot s + i \cdot |s| \cdot \sqrt{1 - s^2}$ depends on the value of s . Substituting $s(\mathbf{c}, \mathbf{e}) = s(\mathbf{e}, \mathbf{c})$ directly gives $f(\mathbf{c}, \mathbf{e}) = f(\mathbf{e}, \mathbf{c})$. ■

Together, these propositions support the use of $f(\mathbf{c}, \mathbf{e})$ as a representation of claim-evidence relevance and semantic relations.

3.2 ANGULAR BOUNDARY CALIBRATION RESULTS

The PSO calibration results are summarized in Table 1, which presents the calibrated angular boundaries (θ_s, θ_r) and the corresponding $F1_{macro}$ scores for each dataset-embedding combination. Since calibration was performed separately for each combination, the resulting boundaries reflect the semantic orientation distribution characteristics of the corresponding dataset and embedding model. The calibrated angular boundaries were then fixed and applied to the held-out test subset, with the resulting test performance presented in Table 2.

TABLE 1.
Experimental Results of Angular Boundary Calibration using PSO

Dataset	Embedding	θ_s	θ_r	$F1_{macro}$	$F1_s$	$F1_{NEI}$	$F1_r$
SNLI	all-MiniLM-L6-v2	52,34°	68,07°	0,57586	0,63075	0,46108	0,63574
	mxlbai-embed-large-v1	39,69°	53,06°	0,59871	0,63957	0,46477	0,69180
FEVER	all-MiniLM-L6-v2	41,25°	50,00°	0,41399	0,47553	0,34795	0,41848
	mxlbai-embed-large-v1	36,42°	50,00°	0,41331	0,57729	0,47549	0,18713

TABLE 2.
Experimental Results of Held-Out Test Evaluation Using Calibrated Boundaries

Dataset	Embedding	$F1_{macro}$	$F1_s$	$F1_{NEI}$	$F1_r$
SNLI	all-MiniLM-L6-v2	0,57928	0,60927	0,46923	0,65934
	mxlbai-embed-large-v1	0,59385	0,62420	0,45425	0,70310
FEVER	all-MiniLM-L6-v2	0,41350	0,48945	0,32455	0,42649
	mxlbai-embed-large-v1	0,43347	0,58389	0,49641	0,22011

To further examine the separation patterns reflected by the calibrated boundaries, Table 3 presents the distribution of semantic orientation angles θ for each claim-evidence relation across the various dataset and embedding model combinations. These distributions can be examined through two key characteristics, namely the width of the NEI zone (θ_s, θ_r) and the distance between the *Mean* θ values of the three claim-evidence relations. A wider NEI zone indicates greater separation among the relation distributions in the semantic orientation space.

TABLE 3.
Distribution of Semantic Orientation Angles Across Claim-Evidence Relation

Dataset	Embedding	Relation	<i>Mean</i> θ	<i>Std.</i> θ	<i>Min.</i> θ	<i>Max.</i> θ
SNLI	all-MiniLM-L6-v2	Supports	47,4°	12,7°	3,8°	90,5°
		NEI	57,4°	12,5°	12,4°	92,5°
		Refutes	69,7°	13,0°	12,4°	101,9°
	mxlbai-embed-large-v1	Supports	34,7°	10,5°	1,8°	63,8°
		NEI	43,2°	10,3°	6,2°	71,4°
		Refutes	56,4°	11,2°	12,7°	85,2°
FEVER	all-MiniLM-L6-v2	Supports	43,7°	13,2°	3,1°	94,3°
		NEI	53,8°	17,1°	13,4°	95,2°

mxbai-embed-large-v1	Refutes	50,8°	13,7°	17,0°	98,7°
	Supports	36,5°	9,1°	4,4°	73,4°
	NEI	45,5°	12,3°	15,6°	75,2°
	Refutes	44,0°	9,3°	13,9°	76,1°

Figure 2 visualizes the distribution of θ of $f(\mathbf{c}, \mathbf{e})$ in polar coordinates, where point colors indicate relation zones, radial lines denote the calibrated angular boundaries, and star markers represent the *Mean θ* of each relation zone. The figure provides a visual representation of the distributions summarized in Table 3, showing the overlap and separation among relation zones.

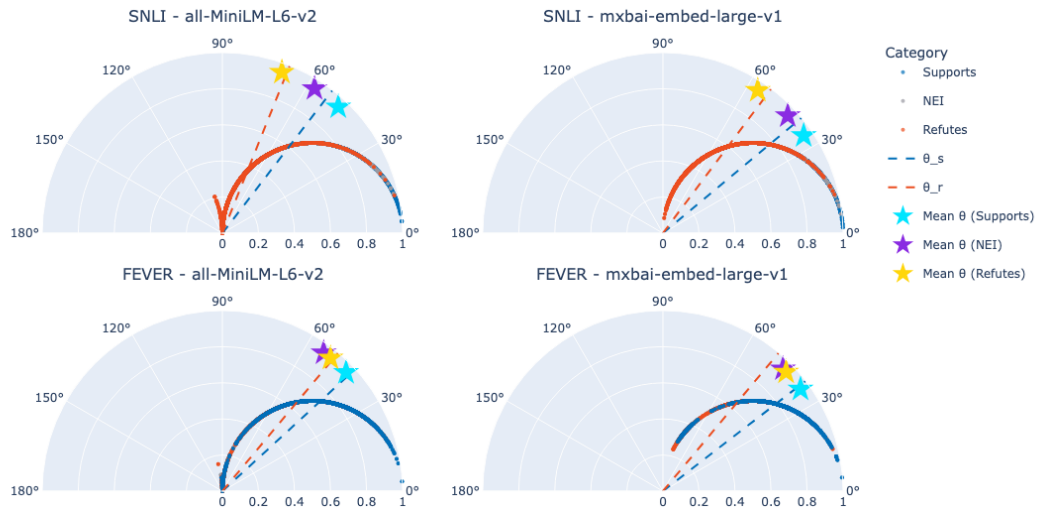


FIGURE 2. Angular Boundaries Across Dataset-Embedding Combinations

In addition, the PSO convergence curves in Figure 3 show that the $F1_{macro}$ score increased from the initial particle positions and stabilized within 200 iterations across all combinations, indicating improvement over the initial solutions.

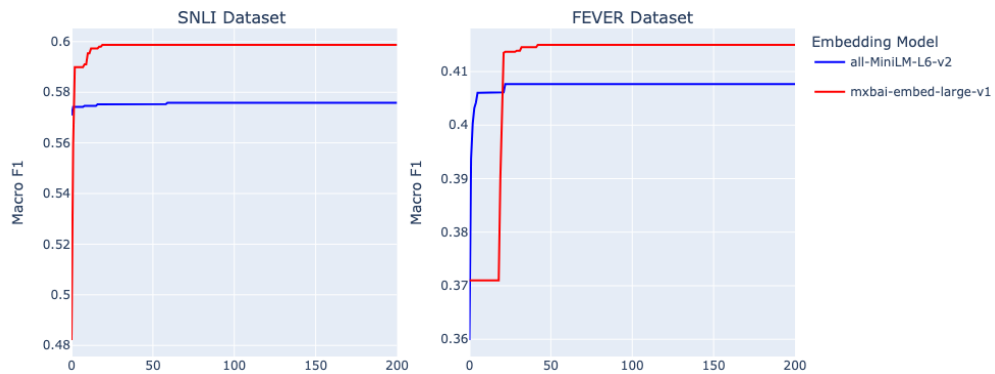


FIGURE 3. Comparison of F1-macro Across Dataset-Embedding Combinations

3.3 INTERPRETATION OF SEMANTIC RELATION ZONES

Based on the PSO calibration results in Table 1, all dataset-embedding combinations yielded angular boundaries satisfying constraint $0^\circ < \theta_s < \theta_r < 180^\circ$,

with supports, NEI, and refutes occupying the lower, intermediate, and higher orientation-angle regions, respectively. The highest $F1_{macro}$ value (59,871%) was obtained for SNLI dataset with mxbai-embed-large-v1 embedding, with calibrated boundaries $(\theta_s, \theta_r) = (39,69^\circ; 53,06^\circ)$, while the lowest value (41,331%) for FEVER dataset with mxbai-embed-large-v1 embedding, with $(\theta_s, \theta_r) = (36,42^\circ; 50,00^\circ)$. In three of the four combinations, the NEI relation yielded the lowest $F1$ score. The only exception was FEVER with mxbai-embed-large-v1, for which the lowest $F1$ score was observed for the refutes relation. To complement the calibration results, the calibrated angular boundaries were then fixed and applied to the held-out test subsets, with the resulting performance reported in Table 2. Since the held-out pairs were not used during boundary calibration, the test results provide an independent evaluation of the calibrated relation zones on unseen data. The same pattern across the three semantic relations was observed in the held-out results, with NEI yielding the lowest $F1$ score in three combinations and refutes yielding the lowest $F1$ score for FEVER with mxbai-embed-large-v1. These results suggest that distinguishing the NEI and refutes relation zones remains the most challenging aspect of the resulting partition.

To better understand this pattern, the distributions of θ prior to calibration are presented in Table 3 and Figure 2. For the SNLI dataset, the *Mean* θ values follow the expected ordering of the relation zones, namely supports, NEI, and refutes. In contrast, for the FEVER dataset, the *Mean* θ of the NEI relation exceeds that of refutes and is accompanied by a larger standard deviation. In addition, for the FEVER and mxbai-embed-large-v1 combination, the NEI and refutes relations show similar *Mean* θ values, while the minimum θ value of NEI exceeds that of refutes. Together, these observations suggest overlap between the two distributions. This pattern is consistent with the relatively low $F1$ scores obtained for the two relations in Table 1.

As shown in Figure 2, the calibrated angular boundaries obtained by PSO do not always align with the *Mean* θ values of the relation distributions. In some dataset-embedding combinations, θ_s and θ_r are located between the centers of adjacent relation distributions, whereas in others they are shifted away from those centers. This observation suggests that *Mean* θ alone may not be sufficient to explain the location of the calibrated angular boundaries, as the relation distributions differ in both spread and overlap.

The width of the NEI zone, defined by the difference between the support and refute boundaries, varies across the dataset-embedding combinations. For SNLI, the NEI zone widths are $15,73^\circ$ for all-MiniLM-L6-v2 and $13,37^\circ$ for mxbai-embed-large-v1, while for FEVER they are $7,48^\circ$ and $13,58^\circ$, respectively. FEVER therefore has a narrower NEI zone than SNLI for all-MiniLM-L6-v2, but a slightly wider NEI zone for mxbai-embed-large-v1. Although the SNLI combinations obtained higher $F1_{macro}$ scores than the FEVER combinations, the zone widths do not show the same consistent ordering across the two embedding models.

The differences between SNLI dan FEVER datasets should also be considered when interpreting the observed results. FEVER more directly represents the claim-evidence setting considered in this study, whereas SNLI represents premise-hypothesis relations and was included as an NLI-based proxy for claim verification [10], [24], [25]. Therefore, the higher $F1_{macro}$ observed for SNLI should not be interpreted as evidence that SNLI is more representative of the claim-evidence context motivating the proposed complex-valued representation. The observed difference in

zone separation should also be considered in relation to the NLI task formulation and annotation process.

Previous work has shown that SNLI contains annotation artifacts, where linguistic patterns in the hypothesis are associated with the annotated relations. A classifier using only the hypothesis was able to predict SNLI labels above the majority baseline, and patterns related to lexical choice and sentence length were found to be associated with the labels [29]. Such patterns may be reflected in the embeddings of SNLI hypotheses. If so, these patterns could be one factor associated with the observed separation between the angular distributions of the SNLI relations and the higher $F1_{macro}$ observed for SNLI. However, this possible relationship was not isolated or directly evaluated in the present study, so its contribution to the observed performance cannot be determined from the current experiments.

At a more general level, the distribution of θ across all experiments shows that the observed claim-evidence pairs are concentrated within a limited portion of the theoretical $0^\circ - 180^\circ$ semantic orientation space. The maximum observed θ value is $101,9^\circ$, whereas all calibrated angular boundaries lie below 90° . Furthermore, the mean values of θ_s and θ_r across the four dataset-embedding combinations are $42,7425^\circ$ and $55,2825^\circ$, respectively. These observations suggest that, for the examined datasets and embedding models, the derived semantic orientation values are concentrated within a relatively narrow angular range. This pattern is specific to the datasets and embedding models examined in this study, and further evaluation across other datasets and embedding models is needed to determine whether similar angular distributions are observed.

For the evaluated datasets and embedding models, the calibration results provide an empirical interpretation of the semantic orientation space constructed by the proposed complex-valued function. The magnitude and angular components represent semantic relevance and semantic orientation, respectively, while the calibrated angular boundaries provide an empirical partition of this continuous representation into semantic relation zones.

4. CONCLUSION

This study introduced a complex-valued function that represents claim-evidence relevance and semantic orientation within a single mathematical object. The proven mathematical properties of the function indicate that the proposed representation behaves consistently on its domain, providing a formal basis for describing claim-evidence relations through both magnitude and angular components. This construction enables semantic relevance and semantic orientation to be represented simultaneously within a single mathematical object. In addition, the semantic orientation angle derived from the embedding geometry provides the continuous coordinate on which empirically calibrated boundaries can be used to partition claim-evidence relations into supports, NEI, and refutes zones.

The calibrated angular boundaries do not always align with the *Mean θ* values of the respective relations, suggesting that *Mean θ* alone is insufficient to explain the resulting boundaries. The spread and overlap of the relation distributions also appear to contribute to the quality of the resulting partitioning. The calibrated boundaries were subsequently evaluated on held-out test subsets, providing an independent

evaluation of the resulting relation zones. Nevertheless, the separability between relation zones in the evaluated datasets remains limited. Future research could examine the stability of calibrated angular boundaries across different embedding models and claim-evidence domains, as well as investigate the influence of dataset-specific annotation characteristics and explore alternative complex-valued function constructions with more independent real and imaginary components.

DECLARATION OF GENERATIVE AI USE

During the preparation of this manuscript, the authors used Claude (Anthropic) and ChatGPT (OpenAI) to assist in improving the clarity, structure, or readability of the text. After using this tool, the authors thoroughly reviewed, edited, and verified the entire content to ensure it accurately represents their own ideas and interpretations. The authors take full responsibility for the integrity and originality of the published work.

REFERENCES

- [1] N. Karkera, S. Ghosh, G. Escames, and S. K. Palaniappan, “MelAnalyze: fact-checking melatonin claims using large language models and natural language inference,” *BMC Med. Inform. Decis. Mak.*, vol. 26, no. 1, 2026, doi: 10.1186/s12911-025-03291-2.
- [2] Y. Wang, Z. Ding, X. Wu, S. Sun, N. Liu, and X. Zhai, “Autoscore: Enhancing automated scoring with multi-agent large language models via structured component recognition,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2026, pp. 40898–40906.
- [3] Y. Wang, S. Tao, N. Xie, H. Yang, T. Baldwin, and K. Verspoor, “Collective Human Opinions in Semantic Textual Similarity,” *Trans. Assoc. Comput. Linguist.*, vol. 11, pp. 997–1013, Aug. 2023, doi: 10.1162/tacl_a_00584.
- [4] X. Hu, Z. Guo, G. Wu, A. Liu, L. Wen, and P. S. Yu, “CHEF: A pilot Chinese dataset for evidence-based fact-checking,” in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2022, pp. 3362–3376. Accessed: Jun. 14, 2026. [Online]. Available: <https://aclanthology.org/2022.naacl-main.246/>
- [5] X. Hu, Z. Hong, Z. Guo, L. Wen, and P. Yu, “Read it twice: Towards faithfully interpretable fact verification by revisiting evidence,” in *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023, pp. 2319–2323. Accessed: Jun. 14, 2026. [Online]. Available: <https://dl.acm.org/doi/10.1145/3539618.3592049>
- [6] J. Han, M. Kamber, and J. Pei, “2 - Getting to Know Your Data,” in *Data Mining: Concepts and Techniques (Third Edition)*, Boston: Morgan Kaufmann, 2012, pp. 39–82. doi: <https://doi.org/10.1016/B978-0-12-381479-1.00002-2>.
- [7] X. Peng, H. Jiang, J. Chen, M. Liu, and X. Chen, “Leveraging large language models and embedding representations for enhanced word similarity computation,” *Sci. Rep.*, vol. 16, no. 1, 2026, doi: 10.1038/s41598-025-31102-1.

- [8] B. G. Teferra *et al.*, “Alignment of Large Language Model Responses With Human Therapists in Motivational Interviewing,” *JAMA Netw. Open*, vol. 9, no. 3, pp. e262750–e262750, Mar. 2026, doi: 10.1001/jamanetworkopen.2026.2750.
- [9] H. Gusdevi, A. Setyanto, K. Kusrini, and E. Utami, “Cosine Similarity-Based Evidences Selection for Fact Verification Using SBERT on the FEVER Dataset,” *CogITO Smart Journal*, vol. 11, no. 1, pp. 52–66, 2025.
- [10] F. Yang, E. Dragut, and A. Mukherjee, “Improving Evidence Retrieval with Claim-Evidence Entailment,” in *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, R. Mitkov and G. Angelova, Eds., Held Online: INCOMA Ltd., Sep. 2021, pp. 1553–1558. [Online]. Available: <https://aclanthology.org/2021.ranlp-1.174/>
- [11] H. Steck, C. Ekanadham, and N. Kallus, “Is cosine-similarity of embeddings really about similarity?,” in *Companion Proceedings of the ACM Web Conference 2024*, 2024, pp. 887–890. [Online]. Available: <https://dl.acm.org/doi/abs/10.1145/3589335.3651526>
- [12] X. Li and J. Li, “AoE: Angle-optimized Embeddings for Semantic Textual Similarity,” in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, L.-W. Ku, A. Martins, and V. Srikumar, Eds., Bangkok, Thailand: Association for Computational Linguistics, Aug. 2024, pp. 1825–1839. doi: 10.18653/v1/2024.acl-long.101.
- [13] X. Wen, C. Yan, H. Huang, and H. Shen, “Embedding word positions with polar coordinates,” *Knowl. Based. Syst.*, p. 113903, 2025, Accessed: Jun. 14, 2026. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0950705125009499>
- [14] Z. Sun, Z.-H. Deng, J.-Y. Nie, and J. Tang, “Rotate: Knowledge graph embedding by relational rotation in complex space,” in *7th International Conference on Learning Representations, ICLR 2019*, 2019. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85083949932&partnerID=40&md5=6ee341486d831f7ed1bbc6c55821f082>
- [15] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in *Proceedings of ICNN'95-international conference on neural networks*, IEEE, 1995, pp. 1942–1948.
- [16] E. A. K. Zaman, A. Ahmad, and A. Mohamed, “Adaptive threshold optimisation for online feature selection using dynamic particle swarm optimisation in determining feature relevancy and redundancy,” *Appl. Soft Comput.*, vol. 156, p. 111477, 2024, doi: <https://doi.org/10.1016/j.asoc.2024.111477>.
- [17] Y. Liu *et al.*, “A random particle swarm optimization based on cosine similarity for global optimization and classification problems,” *Biomimetics*, vol. 9, no. 4, p. 204, 2024, doi: 10.3390/biomimetics9040204.
- [18] X.-Y. Wang, J.-Y. Li, Z.-H. Zhan, and J. Zhang, “A Directional Scale-based Cosine Similarity Indicator for Adaptive Particle Swarm Optimization,” in *2025 15th International Conference on Information Science and Technology (ICIST)*, IEEE, 2025, pp. 60–65. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/11306619>

- [19] S. Miles, L. Yao, W. Meng, C. M. Black, and Z. Ben Miled, “Comparing PSO-based clustering over contextual vector embeddings to modern topic modeling,” *Inf. Process. Manag.*, vol. 59, no. 3, p. 102921, 2022, doi: <https://doi.org/10.1016/j.ipm.2022.102921>.
- [20] G. Strang, *Introduction to linear algebra*, 6th ed. Wellesley-Cambridge Press, 2022.
- [21] N. Reimers and I. Gurevych, “Sentence-bert: Sentence embeddings using siamese bert-networks,” in *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, 2019, pp. 3982–3992.
- [22] N. Reimers and I. Gurevych, “Making monolingual sentence embeddings multilingual using knowledge distillation,” in *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, 2020, pp. 4512–4525.
- [23] S. Lee, A. Shakir, D. Koenig, and J. Lipp, “Open source strikes bread-new fluffy embeddings model,” URL <https://www.mixedbread.ai/blog/mxbai-embed-large-v1>, 2024.
- [24] S. Bowman, G. Angeli, C. Potts, and C. D. Manning, “A large annotated corpus for learning natural language inference,” in *Proceedings of the 2015 conference on empirical methods in natural language processing*, 2015, pp. 632–642.
- [25] J. Thorne, A. Vlachos, C. Christodoulopoulos, and A. Mittal, “FEVER: a large-scale dataset for fact extraction and VERification,” in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 2018, pp. 809–819.
- [26] Pietro Lesci, “pietrolesci/nli_fever · Datasets at Hugging Face.” Accessed: May 12, 2026. [Online]. Available: https://huggingface.co/datasets/pietrolesci/nli_fever
- [27] J. W. Brown and R. V Churchill, *Complex variables and applications*. McGraw-Hill, 2009.
- [28] Y. Shi and R. Eberhart, “A modified particle swarm optimizer,” in *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360)*, 1998, pp. 69–73. doi: 10.1109/ICEC.1998.699146.
- [29] S. Gururangan, S. Swayamdipta, O. Levy, R. Schwartz, S. R. Bowman, and N. A. Smith, “Annotation Artifacts in Natural Language Inference Data,” in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, M. Walker, H. Ji, and A. Stent, Eds., New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 107–112. doi: 10.18653/v1/N18-2017.