Technical Report

1 Properties of *Similarity*

In this part we prove the three properties of the two proposed similarity calculation methods. Under the given axiom translation rules and certain embedding method, the semantic vector which a certain axiom is transformed into is unique. In the following part, we denote $Emb(\alpha)$ as v_1, $Emb(\beta)$ as v_2.

Proof of the properties of Similarity:

Range: It is easy to find $Similarity_{Cos}(v_1, v_2) \in [0, 1]$ and $Similarity_{Euc}(v_1, v_2) \in (0, 1]$ according to the range of Cosine Distance and Euclidean Distance.$Similarity_{Cos}$ is monotonically decreasing with respect to the included angle $\gamma \in [0, \pi)$ while $Similarity_{Euc}$ is monotone concerning the Euclid Distance of the two vectors. So the closer the similarity calculation value is to 1, the higher the similarity between the two vectors is. In turn, the closer the similarity calculation value is to 0, the lower the similarity between the two vectors is.

Grammatical Reflexivity: Because of the uniqueness of the transformation under the certain rules given in this paper, the included angle between v_1 and v_1 is 0, in this way $CosineDistance(v_1, v_1) = 1$ and we conclude that $Similarity_{Cos}(v_1, v_1) = \frac{1}{2}(1+CosineDistance(v_1, v_1)) = 1$. Similarly, $Similarity_{Euc}(v_1, v_1) = 0$, then $Sim(\alpha, \beta) = 1/(1 + Similarity(v_1, v_1)) = 1$. Thus we prove that $Sim(\phi, \phi) = 1$ for any ϕ.

Symmetry: For $Similarity_{Cos}(v_1, v_2) = \frac{1}{2}(1 + CosineDistance(v_1, v_2)) = \frac{1}{2}(1+CosineDistance(v_2, v_1)) = Similarity_{Cos}(v_2, v_1)$ and $Similarity_{Euc}(v_1, v_2) = 1/(1+EuclideanDistance(v_1, v_2)) = 1/(1+EuclideanDistance(v_1, v_2)) = Similarity_{Euc}(v_2, v_1)$, we can conclude that the two similarity calculation methods satisfy the symmetry.

2 Proofs And Details on Logical Properties

We consider the logical nature of the inference method defined above. Our consideration is mainly from the minimal set of expected properties of preferential inference relations [27] (also called system P) and the one of rational inference relations [27] (also called system R).

Although we vectorize all the axioms, for a certain maximal consistent subset, its credibility is still the sum of the credibility of all elements of the set. The score of one axiom is global and does not change for the change of the set. This is essentially the same as the approach of [26] to scoring axioms.

Proof of Theorem 1.

Necessity. For every consistent subset $\Sigma_i \subseteq \Sigma$, for every non-trivial axiom $\alpha \in \Sigma \backslash \Sigma_i$, $\Sigma_i \cup \{\alpha\} \succ_\Sigma \Sigma_i$, score $^s_{\Sigma,\oplus}(\Sigma_i \cup \{\alpha\}) > \text{score}^s_{\Sigma,\oplus}(\Sigma_i)$, $\bigoplus_{\beta \in \Sigma_i \cup \{\alpha\}} s(\Sigma, \beta) - \bigoplus_{\beta \in \Sigma_i} s(\Sigma, \beta) = s(\Sigma, \alpha) > 0$. It shows that for every non-trivial axiom $\alpha \in \Sigma$, there exists a consistent subset $\Sigma_i \subseteq \Sigma$, such that $\alpha \notin \Sigma_i . \alpha$ must be included into at least one $\Sigma_n \in \text{mcs}(\Sigma)$.

case 1 : The cardinality of $\Sigma_n, |\Sigma_n| = 1$, then there must be a subset $\Sigma_m \in mcs(\Sigma), m \neq n$, such that $\alpha \notin \Sigma_m$.

case 2: The cardinality of $\Sigma_n, |\Sigma_n| > 1$, then $\alpha \notin \Sigma_n \backslash \{\alpha\}$. Since the arbitrariness of α is proved above, we conclude that for every non-trivial axiom $\alpha \in \Sigma, s(\Sigma, \alpha) > 0$.

Sufficiency. For every non-trivial axiom $\alpha \in \Sigma, s(\Sigma, \alpha) > 0$, since $\text{score}^S_{\Sigma,\oplus}(\Sigma_i) = \bigoplus_{\beta \in \Sigma_i} s(\Sigma, \beta)$, then for each consistent set $\Sigma_i \subseteq \Sigma$, $\text{score}^s_{\Sigma,\oplus}(\Sigma_i \cup \{\alpha\}) - \text{score}^s_{\Sigma,\oplus}(\Sigma_i) = \bigoplus_{\beta \in \Sigma_i \cup \{\alpha\}} s(\Sigma, \beta) - \bigoplus_{\beta \in \Sigma_i} s(\Sigma, \beta) = s(\Sigma, \alpha) > 0$, namely $\text{score}^s_{\Sigma,\oplus}(\Sigma_i \cup \{\alpha\}) > \text{score}^s_{\Sigma,\oplus}(\Sigma_i)$, which means $Sigma_i \cup \{\alpha\} \succ_\Sigma \Sigma_i$.

Proof of Theorem 2.

Since Theorem 2 is a rewriting of Theorem 5.18 in [27], we refer the reader to the original paper for detailed proof.

Proof of Theorem 3.

Suppose an arbitrary non-trivial axiom $\alpha \in \Sigma$. Considering in the Grammatical Reflexivity of *Similarity*:

$$\begin{aligned} s(K, \alpha) &= mc(\Sigma, \alpha) \\ &= \sum_{\{\Sigma_i \in mcs(\Sigma) | \alpha \in \Sigma_i\}} agg(\Sigma_i, \alpha) \\ &= \sum_{\{\Sigma_i \in mcs(\Sigma) | \alpha \in \Sigma_i\}} (\frac{1}{|\Sigma_i|} \sum_{\beta \in \Sigma_i} Sim(\alpha, \beta)) > 0 \end{aligned}$$

So the selection relation based on our proposed method is monotonic relation according to Theorem 1. And Theorem 2 shows the rationality of the corresponding inference relation. Due to the rationality of our proposed method, the reasoning satisfies:

$$\begin{array}{llll} \textbf{Ref} & \alpha \mid\sim \alpha & \textbf{Cut} & \dfrac{\alpha \wedge \beta | \sim \gamma, \alpha | \sim \beta}{\alpha \mid\sim \gamma} \\ \textbf{LLE} & \dfrac{\alpha \leftrightarrow \beta, \alpha \mid\sim \gamma}{\beta \mid\sim \gamma} & \textbf{Or} & \dfrac{\alpha | \sim \gamma, \beta | \sim \gamma}{\alpha \vee \beta \mid\sim \gamma} \\ \textbf{RW} & \dfrac{\alpha \to \beta, \gamma \mid\sim \alpha}{\gamma \mid\sim \beta} & \textbf{CM} & \dfrac{\alpha | \sim \beta, \alpha | \sim \gamma}{\alpha \wedge \beta \mid\sim \gamma} \\ \textbf{RM} & \dfrac{\alpha | \not\sim \neg\beta, \alpha | \sim \gamma}{\alpha \wedge \beta \mid\sim \gamma} & & \end{array}$$

Ref is Reflexivity.

Cut expresses the fact that one may, in his way towards a plausible conclusion, first add a hypothesis to the facts he knows to be true and prove the plausibility of his conclusion from this enlarged set of facts, and then deduce (plausibly) this added hypothesis from the facts.

LLE is Left Logical Equivalence. Left Logical Equivalence expresses the requirement that logically equivalent formulas have exactly the same consequences.
Or says that any formula that is, separately, a plausible consequence of two different formulas, should also be a plausible consequence of their disjunction.
RW is Right Weakening. Right Weakening obviously implies that one may replace logically equivalent formulas by one another on the right of the $|\sim$ symbol.
CM is Cautious Monotonicity. Cautious Monotonicity expresses the fact that learning a new fact, the truth of which could have been plausibly concluded should not invalidate previous conclusions.
RM is Rational Monotonicity. It expresses the fact that only additional information, the negation of which was expected, should force us to withdraw plausible conclusions previously drawn. It is an important tool in minimizing the updating we have to do when learning new information.

3 Detailed Results on Efficiency

Model	**Method**	**AUTO.**	**biop.**	**UOBM-35**	**UOBM-36**
#mc		0.1	0.6	0.1	0.6
Sentence-BERT	Cosine	129.2	112.5	28.1	42.5
	Euclid	121.3	109.5	25.6	41.6
ConSERT	Cosine	134.9	126.9	32.1	49.3
	Euclid	129.2	132.0	31.1	47.6
TransE	Cosine	593.6	574.7	452.7	461.1
	Euclid	577.8	409.2	449.2	461.5
RDF2Vec	Cosine	368.1	389.2	359.6	361.5
	Euclid	350.0	387.2	359.4	361.8

Table 1: Evaluation results on efficiency (The unit of all data is seconds).

All the evaluation results of efficiency are shown in Table 1. We do not present the results of skeptical inference and CMCS because no selection needs to be executed by Skeptical inference and it is very efficient for CMCS to select those cardinality-maximal subsets (i.e., no more than 1 second for each ontology). For the four ontologies in our experiment, the consumed time of our proposed method is within 10 minutes. Although our proposed methods spend more time than the baselines, they are efficient enough in practice as the selection only needs to be performed once for each ontology and this process can be done offline. For all methods proposed in this paper, they are very efficient and a query can be answered within about half a second.

Overall, for efficiency, Sentence Embedding is better than KG Embedding and Euclid Distance is better than Cosine Distance. Among the models of Sentence Embedding used in our experiment, generally Sentence-BERT is the most efficient .