Type-2 Neutrosophic Ontology for Automated Essays Scoring in Cybersecurity Education

Huda Lafta Majeed[1], Esraa Saleh Alomari[2], Ali Nafea Yousif[3], Oday Ali Hassen[4,*], Saad M. Darwish[5], Yu Yu Gromov[6]

[1]Computer Science and Information Technology, University of Wasit, Al Kut 52001, Iraq

[2]Computer Department, College of Education for Pure Sciences, Wasit University, 52001 Al-Kut, Wasit, Iraq

[3]University of Information Technology and Communication, Baghdad, Iraq

[4]Ministry of Education, Wasit Education Directorate, Kut 52001, Iraq

[5]Department of Information Technology, Institute of Graduate Studies and Research, Alexandria University, Egypt

[6]Director of the Institute of Automation and information technology, Tambov State technical university, Russia

Emails: hulafta@uowasit.edu.iq; ealomari@uowasit.edu.iq; alinafea@uoitc.edu.iq; oday123456789.oa@gmail.com; saad.darwish@alexu.edu.eg; gromovtambov@yandex.ru

Abstract

Given the growing demand for cybersecurity education, the practice of protecting network and software systems from digital and electronic attacks, investing in educational cybersecurity helps significantly reduce the risk of data breaches and protect against security breaches, and given the urgent need and growing number of students worldwide, it is also a way to connect with and between customers by building trust-based relationships, especially regarding essays. Automated Essay Scoring (AES) is a scalable solution for grading large amounts of essays with multiple uses, making it ideal for cybersecurity certification programs, online courses, and standardized tests. In the field of educational cybersecurity, automated essay scoring poses unique challenges due to specialized terminology, persistent and evolving threats. These automated scoring systems use domain-defined ontologies to grade essays but struggle to manage uncertainties, such as ambiguous language and partially valid arguments, which can influence the accuracy of their scoring. Traditional ontologies often struggle to interpret such uncertainties, leading to inconsistent results. Type 2 neutrosophic clustering (T2NS) as a novel approach introduced in this paper is combined with an automated article scoring system based on the cybersecurity learning ontology to address these challenges. The main steps include extracting concepts relevant to this research area from the articles, formalizing the cybersecurity scoring criteria as ontological rules and extending the ontology using T2NS, as well as defining membership functions to measure uncertainty and inconsistency levels. This evaluation using benchmark datasets of cybersecurity articles shows that this approach significantly enhances the scoring reliability and robustness of the approach compared to the basic AES methods.

Keywords: Type-2 neutrosophic ontology; Cybersecurity Education (CE); Foot of Uncertainty (FOU)

1. Introduction

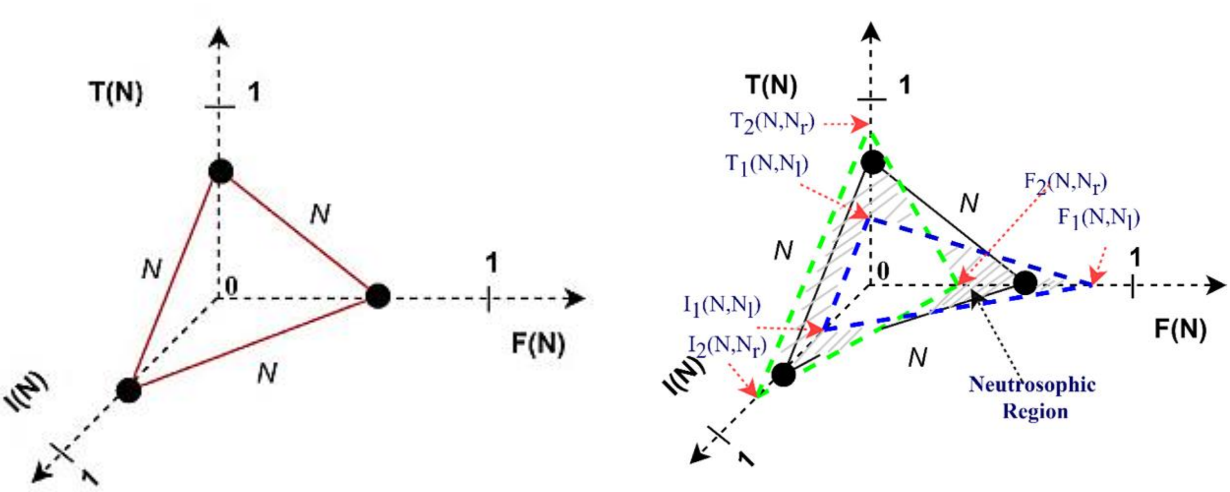
The increasing reliance on digital platforms in education has amplified the demand for scalable, efficient, and reliable assessment methods. Understanding complex technical concepts, ethical considerations, and dynamic threat landscapes is critical in cybersecurity education and protecting internet-connected devices from attacks by spammers, cybercriminals, and hackers. Traditional grading systems face many challenges in consistently and accurately grading articles. These companies use the practice to protect against phishing schemes, attacks, ransomware, identity theft, financial losses, and data breaches. Automated Essay Scoring (AES) has been shown to be a viable solution, offering the ability to process large volumes of articles and provide timely feedback. However, scoring cybersecurity articles presents unique challenges due to domain-specific terminology, sophisticated content, and the need to interpret nuanced arguments and solutions. AES must take into account the

DOI: https://doi.org/10.54216/JCIM.150222
Received: May 22, 2024 Revised: July 23, 2024 Accepted: November 11, 2024

underlying bias and variance in the data presented when modelling human-mediated scoring criteria [1–3]. AES techniques fall into several approaches [4–6]: (1) AES focus on grammar-based and syntactic systems; (2) AES uses statistical machine learning techniques to model patterns from annotated article data; and (3) Natural language processing techniques assess the coherence and cohesion of text. (4) Private ontologies use domain knowledge-based systems to evaluate the correspondence of articles to relevant concepts. (5) Hybrid approaches combine multiple techniques to improve overall performance.

Automated article scoring offers key advantages over AES ontology [7–10]. It uses semantic knowledge to understand the content of an article, and deeper analysis to capture relationships between concepts than purely statistical methods. Automated article scoring can also handle uncertainty and ambiguity through fuzzy logic or neutrosophic clustering techniques, enabling more useful evaluations even when dealing with uncertain information. Neutrosophic ontology that uses neutrosophic logic can represent uncertain and ambiguous information, and enhances automated article scoring (AES) systems by accommodating falsity, degrees of truth, and indeterminacy, allowing AES systems to better handle ambiguous and subjective content This system improves and adapts the scalability and effectiveness of AES systems across different writing tasks such as creative, technical, and argumentative essays, taking into account diverse perspectives and interpretations [11-13]. Neutrosophic ontology can also use neutrosophic logic to represent ambiguous and uncertain information.

It differs from traditional neutrosophic ontology in that it handles uncertainty at a single level through uncertainty of truth and falsity. T2NL enhances traditional neutrosophic ontology by an additional layer of uncertainty. It provides a second layer to capture situations where the degree of uncertainty itself is uncertain. This additional complexity allows for a more comprehensive and detailed representation of uncertainty. It makes the neutrosophic ontology type 2 give valuable advances for various practical applications [14][15]. By incorporating the FOU (see Figure 1), where T2NL represents not only the degree of membership but also the uncertainty associated with it. Type 2 neutrosophic logic (T2NL) with uncertainty fingerprint (FOU) offers significant benefits for accurate modelling of recursive uncertainty. This provides a more comprehensive view of uncertainty. In contrast to Type 1 Neutrosophic Logic (T1NL), which only takes into account uncertainty in membership values, T2NL extends this to the membership function itself. This allows T2NL to handle complex and ambiguous data more effectively, capturing uncertainty about uncertainty and better representing situations with unclear boundaries between categories.



(a) (b)

Figure 1. (a) Type 1 Neutrosophic truth T(N), indeterminacyI(N), and falsity membership functions F(N). (b) Type-2 neutrosophic set membership functions; the blurred region in T2NS provides two extra memberships for truth, indeterminacy, and falsify membership functions TM, IM, and FM. The two extra memberships are N_l, N_r l and r are left and right shifts.

Neutrosophic ontology-based automated essay scoring (NOAES) offers a novel approach that integrates neutrosophic logic and ontology to tackle uncertainties in AES, such as vague language and subjective interpretations. This approach allows flexible grading by defining the degrees of truth, falsity, and indeterminacy for each criterion. However, managing higher uncertainty levels remains a challenge, requiring interdisciplinary efforts in neutrosophic logic, ontology engineering, NLP, and machine learning. Type-2 neutrosophic ontology shows promise in improving uncertainty handling, offering more reliable, interpretable, and scalable AES systems for accurate scoring. This paper extends our previous work in Ref. [13], where Type-2 neutrosophic ontology was used to better represent uncertainty compared to Type-1 neutrosophic logic. Type-2 neutrosophic ontology

DOI: https://doi.org/10.54216/JCIM.150222
Received: May 22, 2024 Revised: July 23, 2024 Accepted: November 11, 2024

provides AES systems with enhanced tools to handle ambiguity and vagueness in essay content. By addressing uncertainty at various levels, it improves scoring accuracy by mitigating the effects of ambiguous language. The adaptive criteria offered by this approach ensure fair and consistent evaluation, accommodating varying levels of uncertainty across essays while promoting transparent assessment.

The structure of this paper is as follows: Section 2 reviews the literature on essay evaluation systems. Section 3 introduces the proposed AES model based on Type-2 neutrosophic ontology. Section 4 presents the model's evaluation, along with results and discussion. Finally, Section 5 concludes the work and outlines potential future research directions.

2. Related Work

Deep neural networks offer strong capabilities for AES but face challenges in interpretability, data requirements, generalization, scalability, fairness, and robustness [16]. They need large labelled datasets for good performance, which can be costly and time-consuming to create, especially for essay scoring. While DNNs excel at learning complex patterns, they may struggle to generalize across varied writing styles, topics, and linguistic differences, leading to biased or inaccurate scores. Additionally, training and deploying large-scale DNNs requires significant computational resources, making real-time processing and scaling costly. Ontology-based approaches offer key advantages for AES, especially compared to neural networks. They can utilize existing knowledge bases, dictionaries, and linguistic resources, reducing the need for large amounts of labelled training data. By explicitly encoding domain knowledge, ontologies perform robustly with fewer annotated examples. Additionally, ontology-based methods are computationally efficient, requiring less processing power than neural networks, making them suitable for resource-constrained devices [17-20]. Table 1 highlights the sate-of-the -art algorithms suggested in the literature, their benefits, and their limitations for different contexts of automated essay assessment and ontology-based evaluation systems.

Table 1: Summary of an automated essay scoring approaches.

Ref.	Algorithm/Method	Pros	Cons
[21]	Subject-focused AES system using data management, semantic matching, ontology building, UML for data design and modeling	Provides specific feedback, incorporates domain knowledge	Limited to subject-focused domains, may require extensive domain expertise for setup
[22]	Automated evaluation framework for question generation using semantic/contextual analysis	Automates question generation, draws from multiple educational sites	Performance depends on quality of external educational data
[23]	Machine learning application for automated essay grading using OntoGen and Natural Language Tool Kit (NLTK)	Effective in correlating vocabulary richness with essay scores	May over-rely on word count/vocabulary, potentially neglecting deeper linguistic quality
[24]	Ontology-based case-based question generation for medical education	Efficient in generating complex stems for medical case-based questions	Focused on medical education, challenging to adapt to other domains
[25]	Ontology-driven learning assessment using SCTonto and electronic health record data	Scalable, reduces test preparation time compared to manual methods	Restricted to automated script concordance test (SCT) generating technology and medical contexts
[26]	Ontology-based multiple-choice question grading with a pruning strategy	Does not require model answers, fully depends on domain ontologies	Limited applicability outside domains with well-established ontologies

DOI: https://doi.org/10.54216/JCIM.150222
Received: May 22, 2024 Revised: July 23, 2024 Accepted: November 11, 2024

Ref.	Algorithm/Method	Pros	Cons
[27]	Two-step topic modeling and ontology-driven keyword recognition with graph-based unsupervised method	Language and domain-independent, outperforms state-of-the-art in keyword recognition	Complexity of setup, relies heavily on quality of domain ontology
[28]	Ontology-based method for assessing test paper depth using Protégé, Jena, and Java libraries	Effective for subject-specific exams, utilizes word path analysis for depth assessment	Highly dependent on the completeness of the developed ontology
[29]	Ontology-based framework for group assessment analytics in collaborative learning	Promotes group interactions, provides feedback for improved communication skills	Limited focus on group interactions, not individual performance

Designing an ontology for essay scoring poses challenges, requiring domain expertise and careful consideration of relevant concepts and relationships. Article interpretation requires the use of ontology-based approaches as well as advanced natural language processing techniques to extract semantic information accurately, given the diversity of language and writing styles. Fuzzy ontology helps to deal with ambiguity by allowing degrees of membership, which reflects the uncertainty in human language. However, the creation of fuzzy ontology is based on subjective decisions about fuzzy grammar, linguistic variables, and membership functions, which leads to varying effectiveness depending on the choices made by different ontologists. This paper aims to develop AES, a development system that provides the highest level of security for the most sensitive data and security standard due to its comprehensive and impenetrable protection. AES is based on the Type 2 neutrosophic ontology to handle higher levels of uncertainty in article scoring, ensuring more reliable and accurate evaluations. The fingerprint of uncertainty (FOU) in this system is an important key to truth and falsehood to manage uncertainty, which provides a more accurate representation than Type 1 neutrosophic sets. By incorporating higher-order uncertainty, the T2NS ontology provides a flexible framework for representing complex and uncertain information. The next section shows the proposed AEE model based on this ontology, which improves semantic evaluation.

3. Proposed Method

3.1 Preliminaries

The ontology-based automated essay scoring system (OAES) combines ontological knowledge representation with automated essay scoring techniques. Let O be the ontology that represents domain-specific and organizational knowledge as a set of concepts, axioms, and relationships. Let S be the set of possible scores, represented as $S = \{s_1, s_2, \ldots, s_m\}$ where m is the total number of possible scores. Let F be the set of features extracted from an article, represented as $F = \{f_1, f_2, \ldots, f_n\}$, where n is the total number of features. OAES can then be mathematically defined as a function $\mathrm{OAES}: F \times O \to S$, where $\mathrm{OAES}(f_1, f_2, \ldots, f_n, O) = s_i$. This function associates the ontological knowledge O with the set of extracted features in an article F with a given score s_i to the set of possible scores S. The ontology O in OAES serves as a knowledge base that encodes the domain-specific relationships, constraints, and concepts relevant to article evaluation. OAES uses the ontological knowledge function to reason about the content, relevance, and coherence of the article, as well as to resolve the uncertainty and ambiguity in the text [7][9][10][21][22].

Incorporating the concept of "uncertainty fingerprint" into the Type 2 Neutrosophic Ontology (T2NO) provides a new way to measure the uncertainty associated with ontology elements. Let C be the set of concepts in the ontology, represented as $R = \{r_1, r_2, \ldots, r_l\}$, where l is the total number of relationships. represented as $C = \{c_1, c_2, \ldots, c_k\}$, where k is the total number of concepts. Let R be the set of relationships between concepts, Let A be the set of axioms defining constraints and logical relationships between concepts and relationships. Let T_1 be the set of layer 1 neutrosophic elements associated with concepts and relationships, represented as $T_1 = \{t_{1c}, i_{1c}, f_{1c}\}$ for concepts and $T_1 = \{t_{1r}, i_{1r}, f_{1r}\}$ for relationships, where t_{1c}, i_{1c}, f_{1c} and t_{1c}, i_{1c}, f_{1c} are the degrees of truth, indeterminacy, and falsity, respectively, at the first layer. Let T_2 be the set of layer 2 neutrosophic elements associated with concepts and relationships, represented as $T_2 = \{t_{2c}, i_{2c}, f_{2c}\}$ for concepts and $T_2 = \{t_{2r}, i_{2r}, f_{2r}\}$ for relationships, where t_{2c}, i_{2c}, f_{2c} and t_{2c}, i_{2c}, f_{2c} are the degrees of truth, indeterminacy, and falsity, respectively, at the second layer.

Let FOU be the set of footprints of uncertainty associated with concepts and relationships, represented as $\mathrm{FOU} = \{u, v\}$, where u and v are the lower (layer 1) and upper bounds (layer 2) of the footprint of uncertainty, respectively.

DOI: https://doi.org/10.54216/JCIM.150222
Received: May 22, 2024 Revised: July 23, 2024 Accepted: November 11, 2024

A type-2 neutrosophic ontology can then be formally defined as a tuple $T2NO = \{C, R, A, T_1, T_2, FOU\}$, where C, R, and A represent the set of concepts, relationships, and axioms, respectively, defining the ontology structure. T_1 and T_2 represent the set of layer 1 and layer 2 neutrosophic elements associated with concepts and relationships, capturing the degrees of truth, indeterminacy, and falsity at the first and second layers, respectively. Figure 2 illustrates the graphical representation of the type-2 neutrosophic membership function.

In formal, a single-valued neutrosophic set (SVNS) $\breve{S}$ on universal set U is characterized by truth membership function$TMF(\emptyset_{\breve{S}})$, indeterminacy membership function $IMF(\emptyset_{\breve{S}})$ and falsity membership function $FMF(\emptyset_{\breve{S}})$ respectively, in the following way [30]

$$\breve{S} = \{\langle \xi, (\emptyset_{\breve{S}}(\xi), \psi_{\breve{S}}(\xi), \varphi_{\breve{S}}(\xi))\rangle : \xi \in U, \emptyset_{\breve{S}}(\xi), \psi_{\breve{S}}(\xi), \varphi_{\breve{S}}(\xi) \in [0,1]\} \quad (1)$$

such that $0 \le \emptyset_{\breve{S}}(\xi), \psi_{\breve{S}}(\xi), \varphi_{\breve{S}}(\xi) \le 3$. Let $\breve{S}(\xi) = [\breve{S}^U(\xi), \breve{S}^L(\xi)]$ be an interval type-2 neutrosophic set (IT2NS) on universal set U where $\xi \in U$ and $\breve{S}^U : U \to [0,1]$ and $\breve{S}^L : U \to [0,1]$ are two type-1 neutrosophic sets (T1NSs) known as upper and lower neutrosophic sets respectively having the condition $0 \le \breve{S}^L(\xi) \le \breve{S}^U(\xi) \le 1$ defined as follows:

$$\breve{S} = \{\langle \xi, ([\emptyset_{\breve{S}}{}^U(\xi), \emptyset_{\breve{S}}{}^L(\xi)], [\psi_{\breve{S}}{}^U(\xi), \psi_{\breve{S}}{}^L(\xi)], [\varphi_{\breve{S}}{}^U(\xi), \varphi_{\breve{S}}{}^L(\xi)])\rangle : \xi \in U\},$$

$$[\emptyset_{\breve{S}}{}^U(\xi), \emptyset_{\breve{S}}{}^L(\xi)], [\psi_{\breve{S}}{}^U(\xi), \psi_{\breve{S}}{}^L(\xi)], [\varphi_{\breve{S}}{}^U(\xi), \varphi_{\breve{S}}{}^L(\xi)] \in [0,1] \quad (2)$$

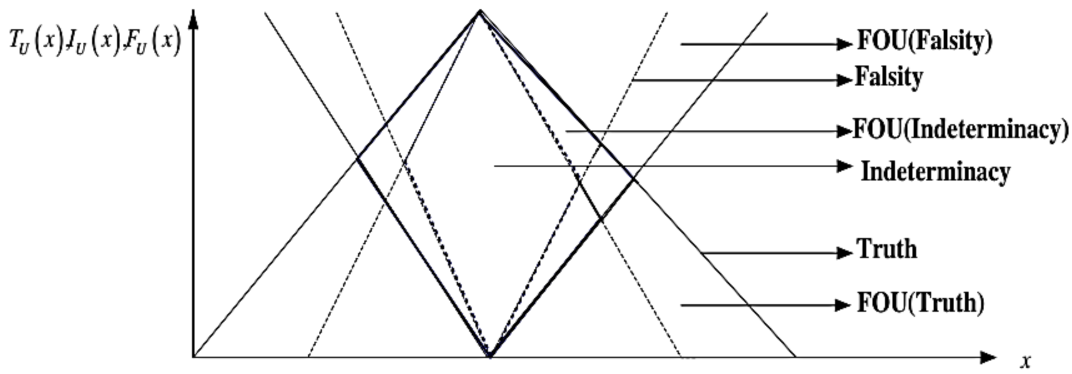


Figure 2. Graphical representation of type-2 neutrosophic membership function [30].

$$\emptyset_{\breve{S}}(\xi) = \begin{cases} \frac{(\xi - \breve{S}_1)\emptyset_{\breve{S}}}{\breve{S}_2 - \breve{S}_1} & \breve{S}_1 \le \xi \le \breve{S}_2 \\ \emptyset_{\breve{S}} & \breve{S}_2 \le \xi \le \breve{S}_3 \\ \frac{(\breve{S}_4 - \xi)\emptyset_{\breve{S}}}{\breve{S}_4 - \breve{S}_3} & \breve{S}_3 \le \xi \le \breve{S}_4 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$\psi_{\breve{S}}(\xi) = \begin{cases} \frac{\breve{S}_2 - \xi + (\xi - \breve{S}_1)\psi_{\breve{S}}}{\breve{S}_2 - \breve{S}_1} & \breve{S}_1 \le \xi \le \breve{S}_2 \\ \psi_{\breve{S}} & \breve{S}_2 \le \xi \le \breve{S}_3 \\ \frac{\xi - \breve{S}_3 + (\breve{S}_4 - \xi)\psi_{\breve{S}}}{\breve{S}_4 - \breve{S}_3} & \breve{S}_3 \le \xi \le \breve{S}_4 \\ 1 & \text{otherwise} \end{cases} \quad (4)$$

$$\varphi_{\breve{S}}(\xi) = \begin{cases} \frac{\breve{S}_2 - \xi + (\xi - \breve{S}_1)\varphi_{\breve{S}}}{\breve{S}_2 - \breve{S}_1} & \breve{S}_1 \le \xi \le \breve{S}_2 \\ \varphi_{\breve{S}} & \breve{S}_2 \le \xi \le \breve{S}_3 \\ \frac{\xi - \breve{S}_3 + (\breve{S}_4 - \xi)\varphi_{\breve{S}}}{\breve{S}_4 - \breve{S}_3} & \breve{S}_3 \le \xi \le \breve{S}_4 \\ 1 & \text{otherwise} \end{cases} \quad (5)$$

where $\emptyset_{\breve{S}} = [\emptyset_{\breve{S}}{}^U, \emptyset_{\breve{S}}{}^L], \psi_{\breve{S}} = [\psi_{\breve{S}}{}^U, \psi_{\breve{S}}{}^L]$ and $\varphi_{\breve{S}} = [\varphi_{\breve{S}}{}^U, \varphi_{\breve{S}}{}^L]$ are interval type-2 neutrosophic numbers (IT2NNs). The number $\breve{S}$ can be represented as (see Figure 3):

$$\breve{S} = [\breve{S}^U, \breve{S}^L] = [(\breve{S}_1^U, \breve{S}_2^U, \breve{S}_3^U, \breve{S}_4^U; \emptyset_{\breve{S}}^U, \psi_{\breve{S}}^U, \varphi_{\breve{S}}^U), (\breve{S}_1^L, \breve{S}_2^L, \breve{S}_3^L, \breve{S}_4^L; \emptyset_{\breve{S}}^L, \psi_{\breve{S}}^L, \varphi_{\breve{S}}^L)] \quad (6)$$

DOI: https://doi.org/10.54216/JCIM.150222
Received: May 22, 2024 Revised: July 23, 2024 Accepted: November 11, 2024

and is called interval type-2 trapezoidal neutrosophic logic number (IT2TrNN) where

$$0 \leq \breve{S}_1^U \leq \breve{S}_2^U \leq \breve{S}_3^U \leq \breve{S}_4^U \leq 1,\ 0 \leq \breve{S}_1^L \leq \breve{S}_2^L \leq \breve{S}_3^L \leq \breve{S}_4^L \leq 1, \quad (7)$$

$$0 \leq \emptyset_{\breve{S}}^L \leq \emptyset_{\breve{S}}^U \leq 1, 0 \leq \psi_{\breve{S}}^L \leq \psi_{\breve{S}}^U \leq 1, 0 \leq \varphi_{\breve{S}}^L \leq \varphi_{\breve{S}}^U \leq 1 \quad (8)$$

In our case, the FOU represents the uncertainty associated with concepts and relationships, quantifying uncertainty for each element in the ontology. Type-2 neutrosophic ontologies offer structured knowledge representations with high uncertainty, enhancing the evaluation of essays. By integrating this ontological knowledge, OAES improves the accuracy of essay scoring, capturing both surface features and deeper semantic content.

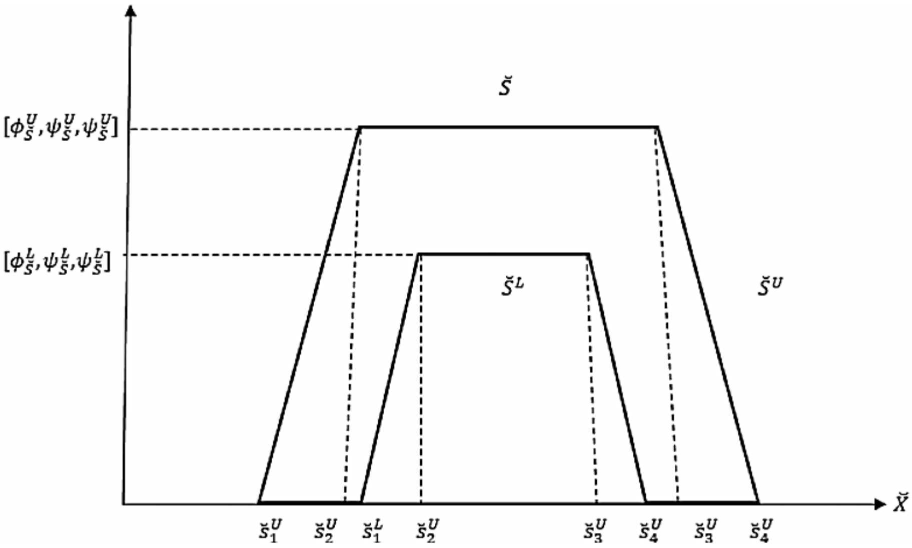


Figure 3. An interval type-2 trapezoidal neutrosophic number [30].

3.2 Type-2 Neutrosophic Ontology-based AES

An ontology-based automated article scoring system (OAES) follows several main steps, starting with pre-processing articles, where texts are cleaned, formatted and irrelevant information is removed, and sometimes-semantic tags are added to support ontology-based analysis or metadata. Features are then extracted, integrating traditional structural and linguistic aspects along with ontology-based features such as semantic similarity and concept coverage, which are represented as numerical vectors for scoring. Semantic analysis (NLP) uses natural language processing to interpret texts in terms of presence, identify key concepts, assess article coherence, develop the presence of the article domain, and integrate it into its scoring framework for content analysis. Machine learning models are trained on labelled datasets, using both traditional and presence-based features to predict outcomes. These models are evaluated for accuracy using different metrics before publishing new article scoring, providing a comprehensive assessment through the combined use of semantic analysis and domain-specific knowledge [7-9].

Article pre-processing is a crucial step in preparing texts for automated scoring systems. The main tasks include segmentation, during which text is broken down into individual words or phrases, which are then converted to lowercase to ensure that the words are represented in a single form. This segmentation process reduces words to their root or lexical forms, ensuring consistency. Punctuation is then removed, as it does not contribute to semantic meaning, and stop words—common words like “the” and “is”—are filtered out. The text is broken down into sentences for structural analysis, with parts of speech tagged for grammatical information. Named entity recognition (NER) is then applied to identify key entities such as people or locations. HTML tags, special characters, and irrelevant information, such as titles or citations, are removed. This comprehensive pre-processing ensures a clean and structured dataset for accurate essay scoring, and is facilitated by tools such as NLTK, Stanford NER, and the standard text-processing library Python.

Various approaches to essay evaluation systems for the purpose of analyzing and interpreting the content of articles include feature extraction methods and dividing sentences into sub-phrases, such as noun and verb phrases, which provide a hierarchical picture of the sentence structure. In order to gain knowledge of the grammatical structure, it is necessary to conduct analysis, specifically constituency analysis. The analysis process can be accomplished with the help of Stanford Analyzer. It focuses on analyzing vocabulary and identifying spelling or grammatical problems, which provides vital insights into the quality of the text. Evaluating the overall lexical richness of the articles and detecting errors with the help of these studies.

DOI: https://doi.org/10.54216/JCIM.150222
Received: May 22, 2024 Revised: July 23, 2024 Accepted: November 11, 2024

$$\text{Lexical Variation (With Error): } LV_1 = \sum Lds / \sum Ls \quad (9)$$

$$\text{Lexical Variation (Without Error): } LV_2 = \sum (Lds - \sum LEs) / \sum Ls \quad (10)$$

$$\text{Percentage of Lexical Error: } LE = \sum LEs / \sum Ls \quad (11)$$

$$\text{Lexical Density: } Ld = \sum L / \sum Tw \quad (12)$$

Where:

Lds represents the total number of items in each part.

LEs represents the total number of lexical errors in each part.

Ls represents the total number of lexical items (excluding stop words) in each part.

Tw represents the total number of words used in the article (including stop words).

L represents the total number of items.

Semantic analysis of essay evaluation systems aims to interpret and evaluate the meaning of the text, the quality of the essay's processing, and the coherence and depth of its arguments and content, respectively. Its main components are identifying main and sub-topics, measuring the similarity between the essay and the expected responses, and evaluating the coherence and flow of ideas through lexical coherence (key repeated terms) and logical coherence (presented idea). Additionally, the system checks if the essay remains focused on the topic, as irrelevant content can negatively affect the score. A word's significance in a corpus can be quantified using its term frequency-inverse document frequency (TF-IDF). The term frequency (TF) measures how often a word (t) appears in a document (d), while the inverse document frequency (IDF) reflects how rare the word is across the document collection (Ds), in which w represents the relative weight of the words in each segment of the essay in relation to the word frequency of the essay as a whole [8].

$$TF - IDF(t, d, Ds) = TF(t, d) \cdot IDF(t, Ds) \quad (13)$$

$$TF - IDF(t, d, Ds) = \frac{|\{t \in d\}|}{|\{w \in d\}|} \cdot \log \frac{|\{d \in Ds\}|}{|\{d \in Ds : t \in d\}|} \quad (14)$$

The frequency of a term in a text determines its weight, and coherent essays are expected to have closely related points in a high-dimensional semantic space, represented by TF-IDF vectors. Coherence is measured by calculating the distance between essay parts in this semantic space. In the proposed model, cosine similarity is used to evaluate cohesiveness; measuring how conceptually, close words are within the essay. (Tokens) A and B are two points, the proximity between them in the semantic space is reflected in higher cosine similarity values, which leads to stronger coherence in meaning. In our model, we use a transformer-based model known as Bidirectional Encoder Representation from Transformers (BERT). It is able to capture deep semantic information by performing a bidirectional analysis of the context of words. Word2Vec is also able to produce word embeddings that are able to construct the context of words in a text from identifying semantic similarities.

$$\cos \theta = A \cdot B / ||A||.||B|| \quad (15)$$

Spatial autocorrelation of articles measures the similarity or dissimilarity in their characteristics based on their geographic distribution. Semantic analysis focuses on the meaning and context of the text, especially when students write articles from different regions. Major spatial autocorrelation measures, such as Moran's (I), Geary's (C), and Getis's (G), have been adapted to high-dimensional semantic space using the Python spatial analysis library. Combining these analyses provides insights into how article content varies across geographic regions.

$$I = \frac{N}{S} \cdot \frac{1}{n} \sum_{k=1}^{n} \left[\frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \left(D_i^k - \overline{D_c^k} \right) \left(D_j^k - \overline{D_c^k} \right)}{\sum_{i=1}^{cc} (D_i^k - \overline{D_c^k})^2} \right] \quad (16)$$

$$C = \frac{(N-1)}{2} \cdot \frac{1}{n} \sum_{k=1}^{n} \left[\frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} (D_i^k - D_j^k)^2}{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} (D_i^k - \overline{D_c^k})^2} \right] \quad (17)$$

$$G(AD) = \frac{1}{n} \sum_{k=1}^{n} \left[\frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(d) D_i^k D_j^k}{\sum_{i=1}^{cc} \sum_{j=1}^{cc} D_i^k D_j^k} \right] \quad (18)$$

S is the sum of all weights w_{ij}, where each pair of points i and j is given a weight w_{ij} of 1 according to whether or not they are neighbors, and 0 according to whether or not they are not neighbors. I can take on values ranging from

DOI: https://doi.org/10.54216/JCIM.150222
Received: May 22, 2024 Revised: July 23, 2024 Accepted: November 11, 2024

-1 to +1; a positive I corresponds to spatial autocorrelation, which means that points in close proximity cluster together. On the other hand, values that are close to zero indicate that there is absolute spatial randomness. AD represents the average distance between two points in semantic space. A weighting function $w_{ij}(d)$ assigns binary weights to pairs of points, where $w_{ij}(d) = 1$ if they are within distance AD, and$w_{ij}(d) = 0$ otherwise. Each part of the essay generates a vector with three elements, resulting in a final vector that averages the corresponding elements from each part.

4. Results and Discussions

The benchmark datasets used for the experiments were adapted to evaluate cybersecurity essays, leveraging data structures and methodologies from the Automated Essay Scoring competition hosted on the Kaggle platform (www.kaggle.com/datasets) [30]. These datasets consist of essays written by students on six diverse topics, with potential for customization to focus on cybersecurity themes. Prior to any training session, essays are pre-scored by two human experts to establish a reference standard. The dataset providers divided the essays into training and testing sets, enabling consistent evaluation across experiments. The proposed model's precision was assessed by comparing its scoring outputs with human evaluations using identical training and testing data.

The model is constructed using a HP Pavilion g6 Series laptop with an Intel Core i5 CPU, 16GB RAM, 64-bit operating system, and Microsoft Windows 10 professional as the running operating system. The following metrics were used to assess the accuracy of the prediction models [2] : The Z-score measures the percentage of essays with the same score from both human evaluators and the AEE model, comparing the number of scores that agree with human raters. The Kappa statistic, also known as the quadratic weighted Kappa (Kp), measures inter-rater reliability by comparing the agreement between human and system evaluation scores. It ranges from 0 to 1, typically falling between arbitrary agreement between evaluators and complete agreement between graders. All experiments were done after transforming the essay into a semantic space by moving a window over it by 10 words.

In Experiment 1, we compare the suggested model to two related statistical-based AES systems: AGE (Automated Grader for Essays) with only linguistic and content attributes and AGE+ (the AGE system with extra coherence attributes) [2][4][5] The purpose of this comparison is to find out if the inclusion of semantic attributes improves model performance. Table 2 displays the quadratic weighted kappa metric, whereas Table 3 displays the exact agreement metric for the comparative models. The model's accuracy increases by 10% to 12% for different datasets compared to AGE and AGE+, primarily due to its simultaneous consideration of syntax, coherence, and consistency. The AGE system only validates syntactic characteristics, while the proposed approach considers all three simultaneously, resulting in better results. This approach addresses grammar and spelling mistakes, smooth paragraph flow, and consistent essay structure.

Table 2: Comparison of AEG, AGE+ and the Suggested Model in terms of Kp.

Models	Dataset 1	Dataset 2	Dataset 3
AGE System	0.91	0.75	0.82
AGE+ system	0.93	0.80	0.84
Proposed model	0.97	0.88	0.93

Table 3: Comparison of AEG, AGE+ and the Suggested Model in terms of Z.

Models	Dataset 1	Dataset 2	Dataset 3
AGE System	0.88	0.72	0.78
AGE+ system	0.90	0.79	0.82
Proposed model	0.96	0.89	0.90

In experiment 2, we compare the proposed model with state-of-the-art systems like Project Essay Grade (PEG), E-Rator, and Lexile Writing Analyzer. PEG uses a database of human-graded training essays to evaluate the quality

DOI: https://doi.org/10.54216/JCIM.150222
Received: May 22, 2024 Revised: July 23, 2024 Accepted: November 11, 2024

of unscored essays using statistical methods. E-Rator assigns grades to essays by identifying key components of writing quality and assigning weights based on statistical procedures. Lexile Writing Analyser calculates the Lexile level of polished, edited, and complete writing using a web-based application. Commercial systems dominate the automated scoring market in the United States, providing grades without feedback. The results shown in Table 4 confirm that the proposed model outperforms other systems in terms of quadratic weighted kappa. These systems extract lemmas into vectors, treating them as numbers without considering word order or coherence. The proposed model, which incorporates semantic features, improves accuracy by incorporating extra features like spatial analysis and spatial autocorrelation.

Table 4: Comparison of the proposed model with state-of-the-art commercial systems in terms of Kp

AEE Systems	Dataset 1	Dataset 2	Dataset 3	Average	Rank
PEG	0.84	0.72	0.84	0.81	2
e-rater	0.83	0.70	0.83	0.78	3
Lexile	0.69	0.63	0.68	0.64	4
Proposed Model	0.97	0.90	0.91	0.92	1

In experiment 3, the proposed model was compared to various machine learning models in essay scoring, including recurrent neural network, similarity measure (cosine coefficient), and word order graph. The results in Table 5 show that the proposed model outperformed the neural network-based model by 18%, the similarity based model by 15%, and the word graph algorithm by 7%. Reasoning that recurrent neural networks are not ideal for all NLP tasks—specifically, they struggle with lengthy sentences—is the basis for the suggested model's superiority. Synonyms, in which different words have the same meaning and polysemy, in which the same word has more than one meaning, are outside the scope of similarity metrics-based AES model. In the suggested model, the essay was evaluated well from a semantic perspective after latent semantic analysis converted it into semantic space, where each word was viewed as an individual point. When extracting features, the word order graph-based scoring system turned to word2vector and doc2vector. One major issue with these two methods is that they fail to provide a precise definition of the sub-linear relationships making them less effective for complex sentences with multiple words.

Table 5: Comparison between the proposed model and machine learning-based AEE models model on selected dataset in terms of Kp.

AEE Systems	Dataset 1	Dataset 2	Dataset 3
Neural Network based model	0.70	0.81	0.80
Measure-based AEE model	0.79	0.73	0.79
Word order graph-based Model	0.83	0.82	0.85
Proposed Model	0.96	0.91	0.89

To further demonstrate the efficacy of type 2 neutrosophic logic and an ontology builder in computer-assisted essay grading, an additional set of experiments was conducted. Here, we compared our prior work [8] [13], which uses fuzzy ontology and type-1 neutrosophic ontology, with our suggested work to express semantic features for automated essay grading. Table 6 results back up the study hypothesis that the system's kp correctness will be enhanced by employing the type-2 neutrosophic ontology to represent essay's semantic properties. When compared to the closest combination that shields type 1 neutrosophic logic and ontology, the proposed combination increased kp by at least 5%. By encompassing a wider range of possible values, type-2 neutrosophic ontology offers robustness against variations and uncertainties in data. This can be particularly beneficial in environments where data is incomplete, inconsistent, or subject to rapid change. Table 7 compares between the three types of ontology.

Table 6: Comparison of the proposed model with other state-of-the-art fuzzy and type-1 ontology-based AEE systems in terms of kp

DOI: https://doi.org/10.54216/JCIM.150222
Received: May 22, 2024 Revised: July 23, 2024 Accepted: November 11, 2024

Models	Dataset 1	Dataset 2	Dataset 3
Fuzzy Ontology model [8]	0.90	0.78	0.86
Type-1 Neutrosophic Ontology model [13]	0.95	0.83	0.88
Type-2 Neutrosophic Ontology model	0.98	0.89	0.92

Table 7: Comparison between fuzzy, type-1 and type-2 neutrosophic ontology

Feature	Fuzzy Ontology	Type-1 Neutrosophic Ontology	Type-2 Neutrosophic Ontology
Representation	Single value for truth	Single values for T, I, and F	Ranges or distributions for T, I, and F
Complexity	Lowest	Moderate	Highest
Computation	Least intensive	Moderately intensive	Most intensive
Expressiveness	Least expressive	More expressive than fuzzy	Most expressive
Handling of Uncertainty	Only truth	Truth, indeterminacy, and falsity	Truth, indeterminacy, and falsity with ranges(FOU)
Interpretability	Easiest	Easier than type-2, more complex than fuzzy	Most complex
Example	0.7	(T=0.6, I=0.2, F=0.2)	(T=[0.5, 0.7], I=[0.1, 0.3], F=[0.1, 0.3])
Use Case	Simple scenarios	Scenarios with moderate complexity and uncertainty	Highly uncertain and dynamic scenarios

5. Conclusion

The integration of Type-2 Neutrosophic Sets into ontology-based Automated Essay Scoring systems for cybersecurity education has proven to be a transformative approach, significantly enhancing scoring reliability and robustness. By effectively handling uncertainties such as ambiguous language, partially correct arguments, and contradictory information, the proposed model overcomes limitations of traditional ontology-based methods. With T2NS, the system models uncertainty, indeterminacy, and contradiction with greater precision, enabling more precise assessments of essays that address the specialized and evolving nature of cybersecurity topics. Evaluation on benchmark datasets demonstrates that this model outperforms baseline AES methods, offering scalable and consistent scoring for cybersecurity certification programs, standardized tests, and online courses. This advancement represents a meaningful step towards automating assessments in highly specialized educational domains. Despite its strengths, the proposed model has limitations that provide avenues for future work. The computational complexity of T2NS and the need for domain-specific ontologies tailored to the dynamic cybersecurity landscape pose challenges for large-scale applications. Improving computational efficiency, expanding the system's applicability to different domains, and incorporating hybrid models that combine human supervision and automated scoring should be considered. The effectiveness of the system in addressing highly novel or unusual arguments deserves further examination, although it has performed well when it comes to scoring well-structured articles. These are the focuses of future studies. Implementing these innovations will ensure the scalability, flexibility, and efficiency of AES systems in a variety of educational settings.

References

[1] Link, S., Koltovskaia, S. Automated Scoring of Writing. In Digital Writing Technologies in Higher Education: Theory, Research, and Practice, pp. 333-345, Cham: Springer International Publishing, 2023.

DOI: https://doi.org/10.54216/JCIM.150222
Received: May 22, 2024 Revised: July 23, 2024 Accepted: November 11, 2024

[2] Ramesh, D., Sanampudi, S. An automated essay scoring systems: a systematic literature review, Artificial Intelligence Review. 55(3):2495-527, 2022.
[3] Honko, M., Neittaanmäki, R., Jarvis, S., Huhta, A. Beyond literacy and competency–The effects of raters' perceived uncertainty on assessment of writing. Assessing Writing, 57(6):1-14, 100768, 2023.
[4] Huawei, S., Aryadoust, V. A systematic review of automated writing evaluation systems. Education and Information Technologies, 28(1):771-795, 2023.
[5] Susanti, M., Ramadhan, A., Warnarsa, H. Automatic essay exam scoring system: a systematic literature, Procedia Computer Science, 216(1):531-538, 2023.
[6] Shang, W., Men, H., Du, X. Machine Learning Models: A Study of English Essay Text Content Feature Extraction and Automatic Scoring. Journal of ICT Standardization, 11(4):379-390, 2023.
[7] Ajetunmobi, S., Daramola, O. Ontology-based information extraction for subject-focused automatic essay evaluation. In Proceedings of the International Conference on Computing Networking and Informatics, pp. 1-6, IEEE, 2017.
[8] Contreras, J., Hilles, S., Abubakar, Z. Automated essay scoring using ontology generator and natural language processing with question generator based on blooms taxonomy's cognitive level. International Journal of Engineering and Advanced Technology, 9(1):2448-2457, 2019.
[9] Almuayqil, S., Abd El-Ghany, S., Shehab, A. Towards an Ontology-Based Fully Integrated System for Student E-Assessment. Journal of Theoretical and Applied Information Technology, 98(21): 3514-3527, 2020.
[10] Mohammed Abed, M., Jarwan, D. A., & Salih, M. A. (2023). On Neutrosophic Relations in Group Theory. International Journal of Mathematics, Statistics, and Computer Science, 1, 48–52. https://doi.org/10.59543/ijmscs.v1i.7739
[11] Bhutani, K., Aggarwal, S. Experimenting with neutrosophic ontologies for medical data classification. In Proceedings of the IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions, pp. 1-6, 2015.
[12] Zolfani, S., Görçün, Ö., Çanakçıoğlu, M., Tirkolaee, E. Efficiency analysis technique with input and output satisficing approach based on Type-2 Neutrosophic Fuzzy Sets: A case study of container shipping companies. Expert Systems with Applications. 218(1):1-21, 119596, 2023.
[13] Karmakar, S., Seikh, M. Type-2 Neutrosophic Fuzzy Bimatrix Games Based on a New Distance Measure. In Recent Trends on Type-2 Fuzzy Logic Systems: Theory, Methodology and Applications, pp. 217-233, Cham: Springer International Publishing, 2023.
[14] Uto, M. A review of deep-neural automated essay scoring models. Behaviormetrika, 48(2):459-484, 2021.
[15] Lim, C., Bong, C., Wong, W., Lee N. A comprehensive review of automated essay scoring (AES) research and development. Pertanika Journal of Science & Technology, 29(3):1875-1899, 2021.
[16] Arbaaeen, A., Shah, A. Ontology-based approach to semantically enhanced question answering for closed domain: A review. Information, 12(5):1-25, 200, 2021.
[17] Srivastava, K., Dhanda, N., Shrivastava, A. An analysis of automated essay grading systems. International Journal of Recent Technology and Engineering, 8(6):5438- 5441, 2020.
[18] Ramnarain-Seetohul, V., Bassoo, V., Rosunally, Y. Similarity measures in automated essay scoring systems: A ten-year review. Education and Information Technologies, 7(4):5573-5604, 2022.
[19] Goura, V., Moulesh, M., Madhusudanarao, N., Gao, X.. An efficient and enhancement of recent approaches to build an automated essay scoring system. Procedia Computer Science, 215(2):442-451, 2022.
[20] Hadyaoui, A., Cheniti-Belcadhi, L. Towards an Ontology-based Recommender System for Assessment in a Collaborative ELearning Environment. In Proceeding of the 18th International Conference on Web Information Systems and Technologies, pp. 294-301, 2022.
[21] Contreras, J., Hilles, S., Abubakar, Z. Automated essay scoring with ontology based on text mining and NLTK tools. In Proceedings of the International Conference on Smart Computing and Electronic Enterprise, pp. 1-6, IEEE, 2018.
[22] Leo, J., Kurdi, G., Matentzoglu, N., Parsia, B., Sattler, U., Forge, S., Donato, G., Dowling, W. Ontology-based generation of medical, multi-term MCQs. International Journal of Artificial Intelligence in Education, 29(4):145-188, 2019.
[23] Radovic, M., Petrovic, N., Tosic, M. An ontology-driven learning assessment using the script concordance test. Applied Sciences. 12(3):1-24, 1472, 2022.
[24] Ramnarain-Seetohul, V., Bassoo, V., Rosunally, Y. OBMAAQ: Ontology-Based Model for Automated Assessment of Short-Answer Questions. In Proceedings of the First International Conference on Advances in Electrical, Electronics and Computational Intelligence, pp. 1-8, 2023.

DOI: https://doi.org/10.54216/JCIM.150222
Received: May 22, 2024 Revised: July 23, 2024 Accepted: November 11, 2024

[25] Beck, M., Rizvi, S., Dengel, A., Ahmed, S. From automatic keyword detection to ontology-based topic modeling. In Proceedings of the International workshop on document analysis systems, pp. 451-465, Cham: Springer International Publishing, 2020.

[26] Premathilaka, Y., Banujan, K., Kumara, B. Ontology-Based Approach to Determine the Coverage of Examination Papers. In Proceedings of the International Conference on Decision Aid Sciences and Application, pp. 613-617, IEEE, 2020.

[27] Hadyaoui, A., Cheniti-Belcadhi, L. Ontology-based group assessment analytics framework for performances prediction in project-based collaborative learning. Smart Learning Environments, 10(1):1-27, 43, 2023.

[28] Bakali, A., Broumi, S., Nagarajan, D., Talea, M., Lathamaheswari, M., Kavikumar, J. Graphical representation of type-2 neutrosophic sets. Neutrosophic Sets and Systems, 42(3):28-38, 2021.

[29] Singh, P. A type-2 neutrosophic-entropy-fusion based multiple thresholding method for the brain tumor tissue structures segmentation. Applied Soft Computing, 103(1):1-23, 107119, 2021.

[30] Nandini, V.; Maheswari, P.U. Automatic assessment of descriptive answers in online examination system using semantic relational features. Journal of Supercomputing, 76(3), 4430–4448, 2020.

DOI: https://doi.org/10.54216/JCIM.150222
Received: May 22, 2024 Revised: July 23, 2024 Accepted: November 11, 2024