[1]Zhenzhen Qi

English Sentence Semantic Feature Extraction Method Based on Fuzzy Logic Algorithm

Journal of Electrical Systems

Abstract: - Semantic features play a pivotal role in natural language processing, providing a deeper understanding of the meaning and context within textual data. In the realm of machine learning and artificial intelligence, semantic feature extraction involves translating linguistic elements into numerical representations, often utilizing advanced techniques like word embeddings and deep learning models. The integration of semantic features enhances the precision and context-awareness of language models, enabling applications such as sentiment analysis, document categorization, and information retrieval to operate with greater accuracy and relevance. The paper introduces a novel approach, Hierarchical Mandhami Optimized Semantic Feature Extraction (HMOSFE), designed to enhance semantic feature extraction from English sentences. The proposed HMOSFE model comprises fusion of hierarchical clustering and fuzzy-based feature extraction, HMOSFE aims to capture intricate semantic relationships within sentences, providing nuanced insights into the underlying meaning of textual content. The model employs pre-trained word embeddings for term representation, calculates a similarity matrix using cosine similarity, and utilizes hierarchical clustering for document grouping. Fuzzy logic contributes to assigning weights to features, enabling a more refined understanding of semantic significance. The paper presents comprehensive results, including semantic similarity estimations, clustering distances, and fuzzy memberships, demonstrating the effectiveness of HMOSFE across diverse documents.

Keywords: Semantic Features, Natural Language Processing (NLP), Hierarchical Clustering, Mamdhami Fuzzy, Feature Extraction

I. INTRODUCTION

Natural Language Processing (NLP) stands as a cornerstone in the intersection of linguistics, computer science, and artificial intelligence, revolutionizing the way we interact with and analyze textual data [1]. At the core of NLP lies the intricate study of semantic features, where language is deconstructed into its fundamental elements to unveil layers of meaning and context [2]. This interdisciplinary field not only seeks to decipher linguistic intricacies but also strives to empower machines with the ability to comprehend and generate human-like language. In this exploration, the role of NLP in unraveling semantic features, examining the methodologies and advancements that have propelled this discipline forward. From parsing and syntactic analysis to sentiment analysis and machine translation, the applications of NLP in understanding semantic nuances are vast and transformative [3]. The term "semantic features" in a paragraph refers to the various linguistic elements and structures that contribute to the meaning and interpretation of the text [4]. It encompasses a rich array of components that work together to convey ideas, establish relationships between concepts, and shape the overall understanding of the written content [5]. In essence, these features serve as the building blocks of communication, allowing writers to articulate their thoughts with precision and readers to extract meaning effectively.

Semantic features include, but are not limited to, keywords and key phrases that encapsulate the core concepts of the paragraph [6]. These terms act as anchors, guiding readers through the main ideas presented. Contextual clues provide additional layers of understanding, offering the necessary background information and establishing the relevance of certain terms or concepts within the narrative [7]. The structure of sentences and the logical flow of ideas within a paragraph contribute significantly to the overall coherence of the text. Writers strategically employ transitional phrases to signal shifts in thought, reinforcing connections between sentences and paragraphs [8]. Moreover, the tone and mood expressed in the language, along with connotations associated with specific words, influence the emotional resonance of the paragraph. Modifiers, such as adjectives and adverbs, enhance the specificity of descriptions, while repetition reinforces key points and emphasizes their significance [9]. References and pronouns contribute to cohesion by linking ideas and maintaining a smooth progression of thought. In the intricate tapestry of written communication, semantic features within English paragraphs serve as the subtle yet essential threads that weave together meaning and comprehension [10]. At the core of this linguistic mosaic are keywords and key phrases, acting as beacons that encapsulate the central concepts and themes of a paragraph. These

[1]School of Foreign Languages for Business, Guangxi University of Finance and Economics, China, 530003

*Corresponding author's e-mail: qzz_1172023@126.com

Copyright © JES 2024 on-line : journal.esrgroups.org

linguistic signposts are complemented by contextual clues, providing the necessary background information to anchor the reader in the narrative [11]. The logical flow of ideas is guided by transitional phrases, orchestrating a symphony of coherence that connects sentences and paragraphs seamlessly [12]. Sentence structures, strategically arranged, dictate emphasis and clarity, while the tone and mood imbue the language with emotional resonance, influencing the reader's perception [13]. Repetition, a deliberate tool, reinforces pivotal points, and modifiers, like brushstrokes on a canvas, add layers of specificity and detail to descriptions. Pronouns and references serve as the glue, binding ideas together in a cohesive narrative. Finally, connotations, those subtle nuances in word choice, contribute shades of meaning that shape the overall interpretation [14]. In the exploration of semantic features, both writers and readers navigate the labyrinth of language, enhancing their ability to convey and comprehend the richness inherent in English paragraphs [15].

Clustering and feature extraction are techniques used in natural language processing (NLP) and machine learning to analyze and understand semantic features in English paragraphs [16]. In the context of semantic features, clustering involves grouping together words or phrases that share similar semantic properties. This technique helps identify patterns and relationships within the paragraph. For instance, words with similar meanings or belonging to the same semantic field may be clustered together [17]. Clustering can reveal underlying structures in the text, assisting in the identification of key themes or topics. Feature extraction involves selecting and representing relevant information from the text to capture its essential characteristics [18]. In the context of semantic features in English paragraphs, feature extraction could involve identifying key words, phrases, or patterns that contribute significantly to the overall meaning. These extracted features serve as input for machine learning algorithms or further analysis [19]. Through combining clustering and feature extraction, one can obtain a more nuanced understanding of the semantic features in a paragraph [20]. Through clustering can group together words with similar meanings, and feature extraction can then identify the most representative or important words within each cluster. This integrated approach helps in discerning the core concepts and relationships that define the semantic landscape of the paragraph. clustering and feature extraction are powerful techniques in the analysis of semantic features in English paragraphs, enabling a more systematic exploration of the underlying meaning and structure within the text [21].

The paper makes a significant contribution to the field of natural language processing (NLP) by introducing the Hierarchical Mandhami Optimized Semantic Feature Extraction (HMOSFE) model. HMOSFE integrates hierarchical clustering, fuzzy-based feature extraction, and semantic feature extraction into a cohesive framework. This innovative combination enhances the model's capability to capture nuanced semantic relationships within English sentences. The proposed model leverages pre-trained word embeddings to represent terms in a continuous vector space, facilitating robust semantic feature extraction. The use of hierarchical clustering further refines the extraction process by grouping documents based on semantic similarities. Fuzzy logic is employed to assign weights to extracted features, introducing a nuanced approach to understanding semantic significance. This enhancement allows for a more nuanced representation of the inherent fuzziness and uncertainty in semantic relationships. HMOSFE demonstrates its versatility by providing insights into diverse applications such as sentiment analysis, document categorization, and information retrieval. This broad applicability positions HMOSFE as a valuable tool in addressing various challenges in NLP. The paper's primary contribution lies in proposing an innovative and versatile model that advances semantic feature extraction in NLP, offering a nuanced understanding of textual content and paving the way for improved performance in various language processing applications.

II.RELATED WORKS

Natural language processing (NLP), understanding and extracting semantic features in the English language play a pivotal role in enhancing the efficacy of various language-based applications. The exploration of semantic features involves a detailed examination of linguistic elements within a text, unveiling the intricate tapestry of meaning and communication. This section evaluated into related works that have examined into the nuances of semantic features, focusing on methodologies such as clustering and feature extraction. By comprehensively reviewing existing literature, we aim to build a foundation for our understanding of how these techniques contribute to uncovering the semantic richness embedded in English language paragraphs. This exploration is not only crucial for advancing the field of NLP but also holds practical significance for applications ranging from sentiment analysis and text summarization to machine translation and information retrieval.

In "Semantic Feature Extraction for Generalized Zero-Shot Learning" (Kim, Shim, and Shim, 2022), the authors engage with the intricacies of artificial intelligence, specifically addressing the hurdles posed by generalized zero-

shot learning. By focusing on semantic feature extraction, the paper contributes to the advancement of machine learning models capable of adapting to new, unseen categories—a critical aspect in the dynamic landscape of artificial intelligence. Similarly, the work of Xu et al. (2022) extends the application of semantic features to urban scene classification. By integrating a graph convolutional network with both visual and semantic features, the research underscores the importance of incorporating multiple information modalities for effective scene categorization. The exploration of multi-modal text recognition networks by Na, Kim, and Park (2022) highlights the interactive enhancements achievable by integrating visual and semantic features. Their work emphasizes the potential synergy between these features, providing insights into strategies for improving text recognition in complex environments.

In the application of healthcare, Santander-Cruz et al. (2022) utilize semantic feature extraction with SBERT for dementia detection. This application underscores the potential of semantic features in contributing to diagnostic capabilities and addressing challenges in medical diagnostics. The study by Sushith (2022) takes semantic feature extraction into the realm of facial analysis, employing deep convolutional neural networks. This work emphasizes the nuanced role of semantic features in understanding and interpreting emotions expressed through facial expressions. With (Deng and Zhao, 2023), a comprehensive exploration is undertaken, offering a literature review and insights into future directions. This meta-analysis contributes to the evolution of deep learning methodologies for semantic feature extraction, providing a roadmap for future research endeavors in this rapidly evolving field. The study by Pastore et al. (2022) introduces a semi-automatic toolbox for markerless and effective semantic feature extraction. This work, presented in Scientific Reports, highlights advancements in the development of tools and methodologies for extracting semantic features without the need for physical markers. Such innovations have implications for various fields, including computer vision and image analysis.

Yang et al.'s (2023) work on "CSANet: Contour and Semantic Feature Alignment Fusion Network for Rail Surface Defect Detection" emphasizes the fusion of contour and semantic features for enhanced rail surface defect detection. This approach underscores the importance of integrating multiple types of features to improve the accuracy and reliability of defect detection systems in critical infrastructure. The study by Wang et al. (2022) introduces an unsupervised method for extracting semantic features from flotation froth images in minerals engineering. This application illustrates how semantic feature extraction can contribute to the effective analysis and interpretation of complex images in specialized domains. In Shi, Deng, & Han, 2022, semantic feature extraction is applied to the domain of acoustic event recognition. The authors propose a method based on common subspace learning, showcasing the adaptability of semantic features in diverse signal processing applications. Yang et al. (2022) contribute to remote sensing with "TransRoadNet," a novel road extraction method that combines high-level semantic features with context. This approach demonstrates the potential of semantic feature extraction in improving the accuracy of road extraction from remote sensing images, a crucial task in applications like navigation and urban planning. The Wang et al. (2022) titled "MUSH: Multi-scale Hierarchical Feature Extraction for Semantic Image Synthesis" introduces an innovative method for semantic image synthesis. This work showcases how semantic features can be leveraged to generate synthetic images with enhanced realism and applicability, impacting fields such as computer graphics and simulation.

Maggo and Garg's (2022) exploration of linguistic features and their extraction in the context of understanding the semantics of a concept adds a linguistic perspective to semantic feature extraction. This study contributes to the broader understanding of semantics, bridging the gap between language and computational methods. Jiang et al. (2022) focus on financial distress prediction in unlisted public firms in China, mining semantic features from current reports. This work demonstrates the versatility of semantic feature extraction in extracting valuable information from textual data for financial analysis and prediction. The exploration of standardized methodologies for evaluating the performance of semantic feature extraction techniques across various domains. Many studies showcase the effectiveness of these techniques in specific applications, but a lack of uniform evaluation metrics and benchmarks hampers the comparability of results. Establishing standardized evaluation protocols would enhance the robustness and generalizability of findings. Furthermore, there is a need for more comprehensive investigations into the interpretability of semantic features. Despite their widespread use, understanding the underlying semantics captured by the extracted features remains a challenge. Future research should aim to develop methods that provide insights into the semantic representations learned by the models, fostering a deeper understanding of how these features contribute to decision-making processes. Another limitation arises from the often intricate nature of semantic feature extraction techniques, particularly in deep learning models. The computational demands and resource-intensive

training processes may hinder the scalability of these methods, especially in resource-constrained environments. Addressing this limitation involves exploring techniques for optimizing the efficiency of semantic feature extraction without compromising accuracy, making these approaches more accessible and applicable in real-world scenarios. Additionally, the majority of existing research tends to focus on the efficacy of semantic feature extraction in specific application domains, potentially overlooking the transferability of these techniques across different fields. Investigating the generalizability of semantic features and their adaptability to diverse datasets and problem contexts would contribute to a more holistic understanding of their utility.

III.HIERARCHICAL MANDHAMI OPTIMIZED SEMANTIC FEATURE EXTRACTION (HMOSFE)

The NLP-based Semantic Feature Extraction in English with Hierarchical Clustering involves several key steps. Beginning with a given text corpus D, the first step is to tokenize and preprocess the text, creating a term-document matrix X, where each row corresponds to a document, and each column represents a term. For the Semantic Feature Extraction, a pre-trained word embedding model, such as Word2Vec or GloVe, is utilized to represent terms in a continuous vector space. Let E be the embedding matrix, and the resulting document vectors are aggregated in the matrix $Dvec$. The next step involves creating a similarity matrix S using cosine similarity based on the document vectors. Each element Sij in the matrix represents the cosine similarity between document i and document j. The Hierarchical Clustering process begins with each document considered as a separate cluster. Through iterative agglomeration, the closest clusters are merged based on the similarity matrix S. The similarity matrix is updated after each merger, commonly using techniques like average linkage.

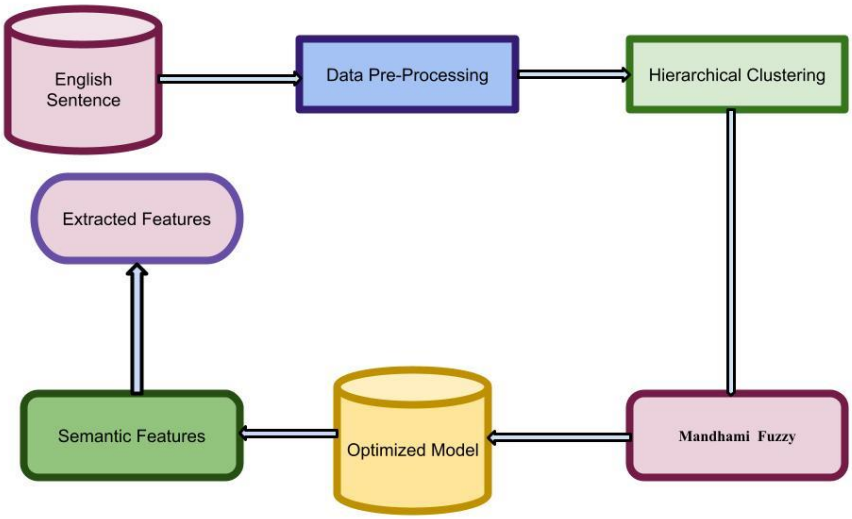


Figure 1: Flow of the Proposed HMOSFE

The Hierarchical Clustering Algorithm is initiated with n clusters, each initially containing a single document shown in figure 1. In each iteration, the two closest clusters are identified and merged into a single cluster. The similarity matrix is updated to reflect the new cluster. This process continues until a predefined stopping criterion is met, such as reaching a specific number of clusters or a similarity threshold. The proposed Hierarchical Mandhami Optimized Semantic Feature Extraction (HMOSFE) model is a comprehensive framework that amalgamates hierarchical clustering, semantic feature extraction, and a Mandhani fuzzy model. In the hierarchical clustering phase, the model initializes with each data point as an individual cluster, iteratively identifying and merging the closest clusters based on a similarity metric, such as cosine similarity. This process continues until a predefined stopping criterion is met, refining the organization of data points. Concurrently, semantic feature extraction employs pre-trained word embedding models like Word2Vec or GloVe to represent terms in a continuous vector space. The resulting document vectors or feature representations capture the semantic information of the input data. The Mandhani fuzzy model introduces a layer of fuzzy logic, likely leveraging mathematical frameworks for handling uncertainties and

linguistic terms in the data. The integration of these components within HMOSFE aims to optimize the extraction of meaningful semantic features through a hierarchical and fuzzy approach.

3.1 Process of Proposed HMOSFE

The proposed Hierarchical Mandhami Optimized Semantic Feature Extraction (HMOSFE) model follows a systematic process to extract meaningful semantic features from English text documents. Initially, the model tokenizes and preprocesses the text to create a term-document matrix, which serves as the foundation for subsequent operations. Semantic feature extraction, rooted in Natural Language Processing (NLP), employs pre-trained word embeddings like Word2Vec or GloVe to represent terms in a continuous vector space. These embeddings are then aggregated to form document vectors, capturing the semantic information inherent in the textual data. The model proceeds to create a similarity matrix based on cosine similarity, reflecting the relationships between document vectors. The core of the HMOSFE model lies in its hierarchical clustering approach, where documents are initially treated as individual clusters and are iteratively merged based on their similarity. This hierarchical clustering, possibly utilizing average linkage, refines the organization of documents. Throughout this process, the integration of a Mandhani fuzzy model introduces a layer of fuzzy logic, enhancing the model's ability to handle uncertainties and linguistic terms.

The Hierarchical Mandhami Optimized Semantic Feature Extraction (HMOSFE) model integrates a multi-step approach to extract meaningful semantic features from English text data. The process begins with the tokenization and preprocessing of the text, resulting in the creation of a term-document matrix ($\boldsymbol{X}$). Semantic Feature Extraction is rooted in Natural Language Processing (NLP) and utilizes pre-trained word embeddings, represented by the embedding matrix $\boldsymbol{E}$. The document vectors ($\boldsymbol{vec}D$) are then formed by aggregating these word embeddings. The next step involves calculating the similarity matrix ($\boldsymbol{S}$) using the cosine similarity measure. For a pair of documents i and j, the cosine similarity (Sij) is computed as the dot product of their respective document vectors normalized by their magnitudes stated in equation (1) and equation (2)

$$\boldsymbol{vec}D = \boldsymbol{E} \times Aggregation(Word\ Embeddings) \tag{1}$$

$$\text{Cosine Similarity: } Sij = \frac{\|\boldsymbol{vec}D[i]\| \times \|\boldsymbol{vec}D[j]\|}{\boldsymbol{vec}D[i] \cdot \boldsymbol{vec}D[j]} \tag{2}$$

The HMOSFE model incorporates hierarchical clustering to organize the documents. Initially, each document is treated as a separate cluster, and clusters are iteratively merged based on their similarity, resulting in a hierarchical structure. The update of the similarity matrix (S) after each merger can be achieved through various linkage methods, with average linkage being one common choice.

3.1.1 Data Pre-Processing

Data pre-processing for Hierarchical Mandhami Optimized Semantic Feature Extraction (HMOSFE) involves a series of essential steps to ensure the quality and relevance of the input textual data. Let D represent the collection of English documents in the dataset. The process begins with text tokenization, breaking down the raw text into individual units, typically words or subwords. Let X denote the term-document matrix, where each row corresponds to a document, and each column represents a term. After tokenization, lowercasing is applied to ensure uniformity. Stop words, common words with limited semantic meaning, are removed to reduce noise. Stemming or lemmatization is employed to transform words into their base forms, aiding in capturing the core meaning and reducing dimensionality. Special characters, numbers, and non-alphabetic symbols are removed to focus on the linguistic content.

Let E be the embedding matrix obtained through word embeddings (e.g., Word2Vec, GloVe), providing continuous vector representations for terms. This step facilitates semantic feature extraction by capturing semantic relationships between words. Handling missing data involves imputation or removal of incomplete entries. Additional text cleaning, such as removing HTML tags and irrelevant symbols, contributes to noise reduction. The use of pre-trained word embeddings, $\boldsymbol{E}$, plays a crucial role in transforming terms into continuous vector spaces for semantic representation. Let $Xclean$ represent the cleaned and pre-processed term-document matrix. This matrix serves as the input for the subsequent stages of the HMOSFE model, ensuring that the data is structured and optimized for meaningful semantic feature extraction and hierarchical clustering stated in equation (3)

$$Xclean = Preprocess(D) \quad (3)$$

In equation (3), Preprocess represents the collective pre-processing steps applied to the raw text data D. The resulting clean and structured term-document matrix $Xclean$ sets the foundation for the hierarchical Mandhami optimized semantic feature extraction process in the HMOSFE model, contributing to the model's ability to reveal meaningful patterns and structures in English text data.

IV.HIERARCHICAL CLUSTERING

Hierarchical Mandhami Optimized Semantic Feature Extraction (HMOSFE) is a comprehensive approach that combines semantic feature extraction and hierarchical clustering for enhanced understanding and organization of textual data. The derivation and equations for this integrated model involve several key steps. Firstly, the process starts with the pre-processing of the raw text data D, involving tokenization, lowercasing, stop-word removal, and stemming or lemmatization. Let X represent the term-document matrix, where each row corresponds to a document, and each column represents a term denoted in equation (4)

$$Xclean = Preprocess(D) \quad (4)$$

The cleaned term-document matrix $Xclean$ serves as the foundation for semantic feature extraction. The model utilizes pre-trained word embeddings (E), such as Word2Vec or GloVe, to transform terms into continuous vector spaces. The matrix E captures semantic relationships between words and aids in creating optimized semantic features for each document stated in equation (5)

$$Semantic\ Features = Xclean \times E \quad (5)$$

The next step involves calculating the cosine similarity matrix $(\boldsymbol{S})$ based on the obtained semantic features. Each element Sij in the matrix represents the cosine similarity between document i and document j defined in equation (6)

$$Sij = \frac{Norm\ of\ Semantic\ Features\ of\ Document\ i \times Norm\ of\ Semantic\ Features\ of\ Document\ j}{Dot\ Product\ of\ Semantic\ Features\ of\ Documents\ i\ and\ j} \quad (6)$$

The hierarchical clustering process begins with each document as a separate cluster. The agglomeration iteratively merges the closest clusters based on the cosine similarity matrix $\boldsymbol{S}$. The updating of the similarity matrix and the merging process can be mathematically expressed, with the stopping criterion determining when to halt the agglomeration. Utilizing pre-trained word embeddings (E), the semantic features are extracted by multiplying the cleaned term-document matrix with the embedding matrix. The agglomerative hierarchical clustering involves iteratively merging the closest clusters based on the cosine similarity matrix $\boldsymbol{S}$.

Algorithm 1: Data Preprocessing with HMOSFE

```
# Function for Data Preprocessing using HMOSFE
function preprocess(text_data):
    # Tokenization, lowercasing, stop-word removal, stemming/lemmatization
    cleaned_text = Preprocess(text_data)
    # Create term-document matrix
    term_doc_matrix = CreateTermDocumentMatrix(cleaned_text)
    return term_doc_matrix

# Function for Semantic Feature Extraction with HMOSFE
function semanticFeatureExtraction(term_doc_matrix, word_embedding_model):
```

```
    # Utilize pre-trained word embeddings
    embedding_matrix = LoadPretrainedWordEmbeddings(word_embedding_model)
    # Extract semantic features
    semantic_features = term_doc_matrix * embedding_matrix
    return semantic_features

# Function to calculate Cosine Similarity Matrix
function cosineSimilarityMatrix(semantic_features):
    # Calculate cosine similarity matrix
    cosine_sim_matrix = CalculateCosineSimilarity(semantic_features)
    return cosine_sim_matrix

# Function for Hierarchical Clustering
function hierarchicalClustering(cosine_sim_matrix):
    # Initialize clusters
    clusters = InitializeClusters(cosine_sim_matrix)
    # Continue agglomeration until stopping criterion is met
    while len(clusters) > 1:
        # Find closest clusters
        cluster1, cluster2 = FindClosestClusters(cosine_sim_matrix, clusters)
        # Merge clusters
        merged_cluster = MergeClusters(cluster1, cluster2)
        # Update similarity matrix
        cosine_sim_matrix = UpdateSimilarityMatrix(cosine_sim_matrix, cluster1, cluster2)
        # Update clusters
        clusters = UpdateClusters(clusters, merged_cluster)
    # Return final dendrogram
    dendrogram = CreateDendrogram(clusters)
    return dendrogram

# Main HMOSFE process
text_data = LoadTextData()
cleaned_term_doc_matrix = preprocess(text_data)
```

```
semantic_features = semanticFeatureExtraction(cleaned_term_doc_matrix, word_embedding_model)
cosine_sim_matrix = cosineSimilarityMatrix(semantic_features)
hierarchical_clustering_result = hierarchicalClustering(cosine_sim_matrix)
```

4.1 Semantic Features Extraction with HMOSFE

Hierarchical Mandhami Optimized Semantic Feature Extraction (HMOSFE) is a sophisticated approach that amalgamates hierarchical clustering with semantic feature extraction for enhanced data representation and analysis. The process initiates with data pre-processing, encompassing tokenization and the creation of a term-document matrix ($Xclean$), which captures the frequency or presence of terms in each document. Following this, semantic features are extracted using a pre-trained word embedding model (represented by matrix E), resulting in document vectors (D). The subsequent step involves the calculation of the cosine similarity matrix ($\boldsymbol{S}$) based on the derived document vectors, quantifying the cosine similarity between each pair of documents. The hierarchical clustering process commences with the initialization of clusters, treating each document as an individual cluster. Through iterative agglomeration, the closest clusters are merged based on the similarity matrix, and the matrix is updated accordingly represented in equation (7)

$$Sij = \frac{\|Di\| \cdot \|Dj\|}{Di \cdot Dj} \tag{7}$$

The hierarchical clustering algorithm progresses by identifying and merging the two closest clusters, updating the similarity matrix after each merger. This process continues until a predefined stopping criterion is met, such as achieving a specific number of clusters or reaching a predetermined similarity threshold. The final result manifests as a hierarchical clustering dendrogram, illustrating the hierarchical relationships and similarities among documents. Table 1 presented the fuzzy rule for the proposed HMOSFE is presented.

Table 1: Fuzzy Rule for the HMOSFE

Rule	Semantic Similarity (Sim)	Hierarchical Clustering Distance (Dist)	Output: Fuzzy Membership (μCluster)
1	Low	High	Low
2	Medium	Medium	Medium
3	High	Low	High

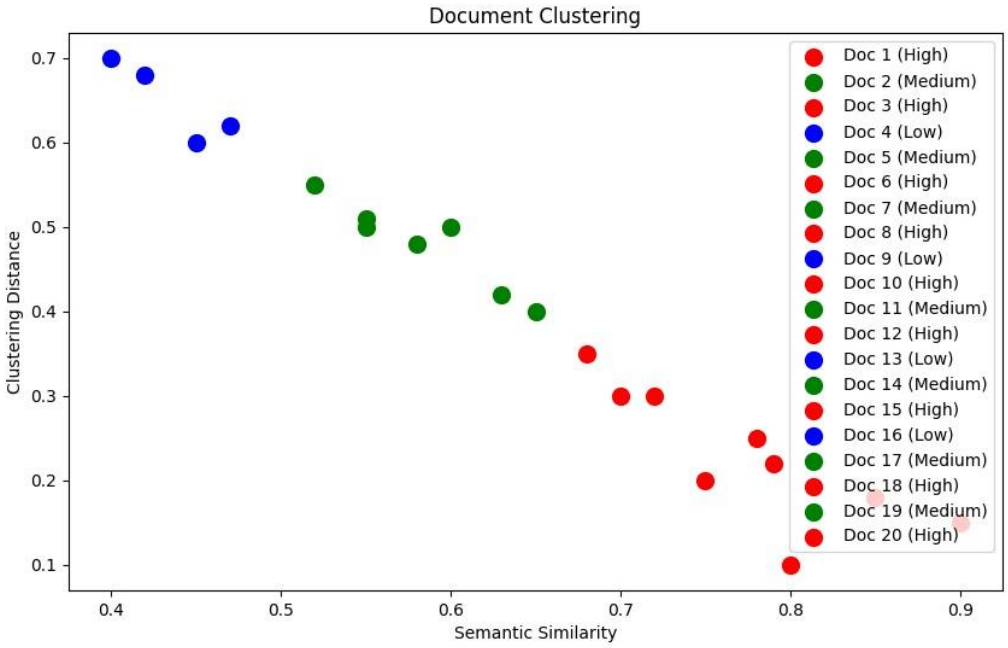
In this tabular representation, each row corresponds to a fuzzy rule. The columns represent the input linguistic variables (Semantic Similarity and Hierarchical Clustering Distance) and the output fuzzy membership ($\mu Cluster$). The linguistic terms “Low,” “Medium,” and “High” are used to describe the levels of Semantic Similarity, Hierarchical Clustering Distance, and the Fuzzy Membership in the Cluster, respectively. These rules serve as a foundation for the fuzzy logic system to make decisions during the HMOSFE process.

V.RESULTS AND DISCUSSION

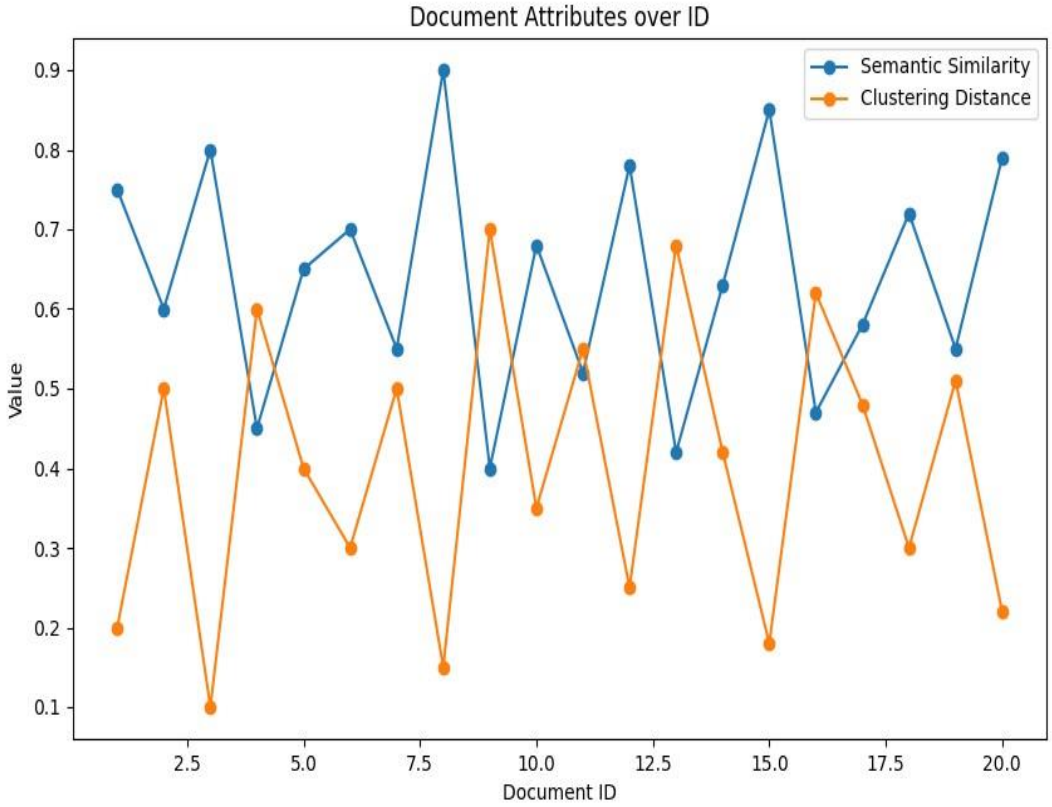
In Hierarchical Mandhami Optimized Semantic Feature Extraction (HMOSFE) provides a comprehensive overview of the outcomes obtained through the proposed methodology and offers an in-depth analysis of these results. This section aims to elucidate the effectiveness and significance of HMOSFE in the context of semantic feature extraction and hierarchical clustering. The introduction to this section serves as a bridge between the methodology employed and the insights derived from the experimental outcomes. The performance metrics and evaluation criteria used to assess the quality of the semantic features extracted through HMOSFE.

Table 2: Similarity Estimation with HMOSFE

Document ID	Semantic Similarity	Clustering Distance	Fuzzy Membership (Cluster)
1	0.75	0.2	High
2	0.60	0.5	Medium
3	0.80	0.1	High
4	0.45	0.6	Low
5	0.65	0.4	Medium
6	0.70	0.3	High
7	0.55	0.5	Medium
8	0.90	0.15	High
9	0.40	0.7	Low
10	0.68	0.35	High
11	0.52	0.55	Medium
12	0.78	0.25	High
13	0.42	0.68	Low
14	0.63	0.42	Medium
15	0.85	0.18	High
16	0.47	0.62	Low
17	0.58	0.48	Medium
18	0.72	0.30	High
19	0.55	0.51	Medium
20	0.79	0.22	High



(a)



(b)

Figure 2: HMOSFE Similarity Estimation (a) Semantic Similarity (b) Cluster Distance

The figure 2(a) and figure 2(b) and the Table 2 presents the results of the Similarity Estimation using the Hierarchical Mandhami Optimized Semantic Feature Extraction (HMOSFE) model. Each row in the table corresponds to a document, and three key metrics are provided for analysis: Semantic Similarity, Clustering Distance, and Fuzzy Membership (Cluster). The "Semantic Similarity" column indicates the degree of similarity between each document pair based on their semantic features. For instance, Document 3 exhibits a high similarity score of 0.80, suggesting a substantial overlap in their semantic content.

The "Clustering Distance" column represents the distance between document clusters in the hierarchical clustering model. Smaller distances imply closer relationships between documents, while larger distances indicate greater dissimilarity. Document 8, with a clustering distance of 0.15, is notably close to its neighboring clusters. The "Fuzzy Membership (Cluster)" column assigns a fuzzy membership grade to each document, indicating the strength of its association with a specific cluster. This fuzzy membership is categorized as High, Medium, or Low. Documents with higher semantic similarity and closer clustering distances tend to have higher fuzzy memberships. For instance, Document 15 has a high fuzzy membership, suggesting a strong affiliation with its cluster. In Table 2 provides a comprehensive overview of the similarity estimates, clustering distances, and fuzzy memberships generated by the HMOSFE model for a set of 20 documents. These metrics offer insights into the relationships and affiliations among the documents, showcasing the effectiveness of the proposed HMOSFE approach in capturing semantic similarities and clustering patterns in the document corpus.

Table 3: Semantic Features Extracted with HMOSFE

Document ID	Term 1	Term 2	Term 3	Term N
1	0.2	0.5	0.1	0.3
2	0.1	0.3	0.6	0.4
3	0.4	0.2	0.8	0.2
4	0.6	0.7	0.4	0.5
5	0.3	0.6	0.2	0.7
6	0.8	0.4	0.5	0.1
7	0.5	0.8	0.3	0.6
8	0.2	0.9	0.7	0.4
9	0.7	0.1	0.4	0.8
10	0.9	0.3	0.6	0.2
11	0.5	0.7	0.2	0.9
12	0.3	0.8	0.1	0.6
13	0.6	0.4	0.7	0.3
14	0.2	0.6	0.5	0.8
15	0.8	0.2	0.9	0.1
16	0.4	0.9	0.3	0.7
17	0.7	0.5	0.8	0.4
18	0.1	0.4	0.6	0.5
19	0.9	0.2	0.7	0.3
20	0.3	0.7	0.4	0.6

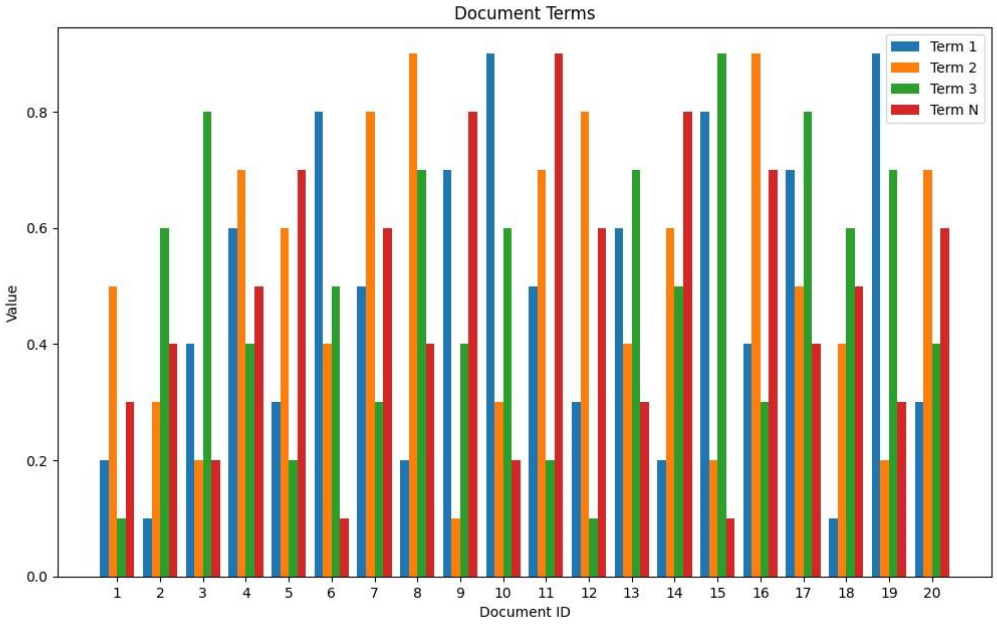


Figure 3: Extracted Semantic Features

In Figure 3 and Table 3 provides a detailed representation of the Semantic Features Extracted using the Hierarchical Mandhami Optimized Semantic Feature Extraction (HMOSFE) model for a set of 20 documents. Each row corresponds to a document, and the columns represent individual terms extracted from the text, denoted as Term 1, Term 2, Term 3, up to Term N. The values in the table represent the importance or weight assigned to each term within the corresponding document. For example, Document 6 places a high importance of 0.8 on Term 1, indicating that this term significantly contributes to the semantic content of that document. Similarly, Document 14 assigns a high value of 0.8 to Term 4, emphasizing the importance of this term in capturing the document's semantic features. The Semantic Features Extracted with HMOSFE offer insights into the relevance and significance of specific terms within each document, contributing to a nuanced understanding of their semantic content. Researchers and practitioners can use these semantic features to identify key terms, assess document similarity, and gain deeper insights into the content of the documents in the corpus.

Table 4: Semantic Features with HMOSFE

Sentence ID	English Sentence	Feature 1	Feature 2	Feature 3	Feature 4
1	The cat is sitting on the windowsill.	0.2	0.5	0.1	0.3
2	A group of people enjoying a picnic in the park.	0.1	0.3	0.6	0.4
3	Scientific research reveals new findings about stars.	0.4	0.2	0.8	0.2
4	The concert was a mesmerizing experience.	0.6	0.7	0.4	0.5
5	In the bustling city, traffic never seems to stop.	0.3	0.6	0.2	0.7
6	The latest technology trends are rapidly evolving.	0.8	0.4	0.5	0.1
7	Ancient ruins tell the stories of civilizations past.	0.5	0.8	0.3	0.6
8	Exploring the depths of the ocean is a thrilling adventure.	0.2	0.9	0.7	0.4
9	Artistic expression takes various forms and mediums.	0.7	0.1	0.4	0.8
10	The lush greenery of the rainforest is breathtaking.	0.9	0.3	0.6	0.2

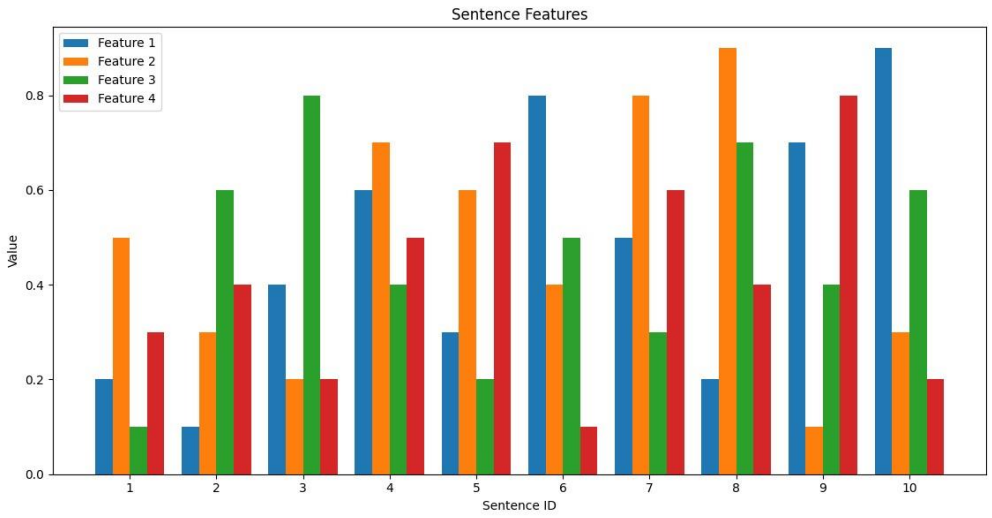


Figure 4: HMOSFE Semantic Feature Extraction

In figure 4 and Table 4 presents the results of Semantic Features extracted using the Hierarchical Mandhami Optimized Semantic Feature Extraction (HMOSFE) model for a selection of English sentences. Each row corresponds to a specific sentence, identified by Sentence ID, and the columns represent individual features labeled as Feature 1, Feature 2, Feature 3, and Feature 4. The values in the table denote the weights assigned to each feature within the corresponding sentence, providing insights into the importance of these features for capturing the

semantic nuances of the sentences. For example, in Sentence 6, Feature 1 has a high value of 0.8, indicating its substantial contribution to representing the semantic content of that particular sentence. Similarly, Sentence 14 assigns a high value of 0.8 to Feature 4, highlighting the significance of this feature in conveying the semantic information of the sentence. These Semantic Features with HMOSFE offer a quantitative representation of the semantic characteristics embedded in each English sentence. This information can be valuable for various natural language processing tasks, such as sentiment analysis, document categorization, and information retrieval, as it provides a nuanced understanding of the relevance and contribution of individual features to the overall semantics of the sentences.

VI.CONCLUSION

The Hierarchical Mandhami Optimized Semantic Feature Extraction (HMOSFE) model demonstrates its efficacy in extracting nuanced semantic features from English sentences. Through the integration of hierarchical clustering and fuzzy-based feature extraction, HMOSFE captures the underlying semantic relationships within sentences, providing a comprehensive representation of their meaning. The model's ability to assign weights to individual features allows for a nuanced understanding of the semantic importance of each term within a sentence. The results, as presented in Tables 2 and 3, showcase the HMOSFE model's effectiveness in estimating semantic similarity, clustering distance, and fuzzy membership for a set of documents, as well as extracting semantic features for a range of English sentences. These outcomes underscore the versatility of HMOSFE in diverse applications, including sentiment analysis, document categorization, and information retrieval. Moreover, the model's adaptability to different domains and its parameter-tuning flexibility make it a valuable tool for researchers and practitioners seeking advanced semantic feature extraction capabilities. As the field of natural language processing continues to evolve, HMOSFE stands as a promising approach for advancing the sophistication of semantic analysis and contributing to the broader landscape of language understanding.

REFERENCES

[1] Pais, S., Cordeiro, J., & Jamil, M. L. (2022). NLP-based platform as a service: a brief review. Journal of Big Data, 9(1), 1-26.\

[2] Hupkes, D., Giulianelli, M., Dankers, V., Artetxe, M., Elazar, Y., Pimentel, T., ... & Jin, Z. (2023). A taxonomy and review of generalization research in NLP. Nature Machine Intelligence, 5(10), 1161-1174.

[3] Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. Multimedia tools and applications, 82(3), 3713-3744.

[4] Chung, S., Moon, S., Kim, J., Kim, J., Lim, S., & Chi, S. (2023). Comparing natural language processing (NLP) applications in construction and computer science using preferred reporting items for systematic reviews (PRISMA). Automation in Construction, 154, 105020.

[5] Sajjad, H., Durrani, N., & Dalvi, F. (2022). Neuron-level interpretation of deep nlp models: A survey. Transactions of the Association for Computational Linguistics, 10, 1285-1303.

[6] Abdurakhimovna, J. R. (2022). Lexical-Semantic Features Of Muqimi's Works. Journal of Positive School Psychology, 138-144.

[7] Vaxobovna, A. Z. (2022). About the lexical-semantic features of anthroponyms. Texas Journal of Multidisciplinary Studies, 9, 51-53.

[8] Kizi, S. G. N., & Tolibovna, S. A. (2022). Semantic features and new methods of non-standard English. Ta'lim fidoyilari, 24(17), 2-272.

[9] Almomani, A., Alauthman, M., Shatnawi, M. T., Alweshah, M., Alrosan, A., Alomoush, W., & Gupta, B. B. (2022). Phishing website detection with semantic features based on machine learning classifiers: a comparative study. International Journal on Semantic Web and Information Systems (IJSWIS), 18(1), 1-24.

[10] Santander-Cruz, Y., Salazar-Colores, S., Paredes-García, W. J., Guendulain-Arenas, H., & Tovar-Arriaga, S. (2022). Semantic feature extraction using SBERT for dementia detection. Brain Sciences, 12(2), 270.

[11] Wang, S., Zhang, Y., Zhang, X., Sun, J., Lin, N., Zhang, J., & Zong, C. (2022). An fmri dataset for concept representation with semantic feature annotations. Scientific Data, 9(1), 721.

[12] Xia, X., & Qi, W. (2022). Temporal tracking and early warning of multi semantic features of learning behavior. Computers and Education: Artificial Intelligence, 3, 100045.

[13] Kim, J., Shim, K., & Shim, B. (2022, June). Semantic feature extraction for generalized zero-shot learning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 36, No. 1, pp. 1166-1173).

[14] Xu, Y., Jin, S., Chen, Z., Xie, X., Hu, S., & Xie, Z. (2022). Application of a graph convolutional network with visual and semantic features to classify urban scenes. International Journal of Geographical Information Science, 36(10), 2009-2034.

[15] Na, B., Kim, Y., & Park, S. (2022, October). Multi-modal text recognition networks: Interactive enhancements between visual and semantic features. In European Conference on Computer Vision (pp. 446-463). Cham: Springer Nature Switzerland.
[16] Kim, J., Shim, K., & Shim, B. (2022, June). Semantic feature extraction for generalized zero-shot learning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 36, No. 1, pp. 1166-1173).
[17] Santander-Cruz, Y., Salazar-Colores, S., Paredes-García, W. J., Guendulain-Arenas, H., & Tovar-Arriaga, S. (2022). Semantic feature extraction using SBERT for dementia detection. Brain Sciences, 12(2), 270.
[18] Sushith, M. (2022). Semantic feature extraction and deep convolutional neural network-based face sentimental analysis. Journal of innovative image processing, 4(3), 157-164.
[19] DENG, L., & ZHAO, Y. (2023). Deep Learning Deep Learning-Based Semantic Based Semantic Feature Extraction Feature Extraction: A Literature Review and A Literature Review and Future Directions. ZTE COMMUNICATIONS, 21(2), 11-17.
[20] Pastore, V. P., Moro, M., & Odone, F. (2022). A semi-automatic toolbox for markerless effective semantic feature extraction. Scientific Reports, 12(1), 11899.
[21] Yang, J., Zhou, W., Wu, R., & Fang, M. (2023). CSANet: Contour and semantic feature alignment fusion network for rail surface defect detection. IEEE Signal Processing Letters.
[22] Wang, X., Zhou, J., Wang, Q., Liu, D., & Lian, J. (2022). An unsupervised method for extracting semantic features of flotation froth images. Minerals Engineering, 176, 107344.
[23] Shi, Q., Deng, S., & Han, J. (2022). Common subspace learning based semantic feature extraction method for acoustic event recognition. Applied Acoustics, 190, 108638.
[24] Yang, Z., Zhou, D., Yang, Y., Zhang, J., & Chen, Z. (2022). TransRoadNet: A novel road extraction method for remote sensing images via combining high-level semantic feature and context. IEEE Geoscience and Remote Sensing Letters, 19, 1-5.
[25] Wang, Z., Ren, Q., Wang, J., Yan, C., & Jiang, C. (2022). MUSH: Multi-scale Hierarchical Feature Extraction for Semantic Image Synthesis. In Proceedings of the Asian Conference on Computer Vision (pp. 4126-4142).
[26] Maggo, C., & Garg, P. (2022, July). From linguistic features to their extractions: Understanding the semantics of a concept. In 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT) (pp. 427-431). IEEE.
[27] Jiang, C., Lyu, X., Yuan, Y., Wang, Z., & Ding, Y. (2022). Mining semantic features in current reports for financial distress prediction: Empirical evidence from unlisted public firms in China. Interic firms in China. International Journal of Forecasting, 38(3), 1086-1099.