This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3421595

Dialectic Feature-based Fuzzy Graph Learning for Label Propagation Assisting Text Classification

Cherukula Madhu, *Member, IEEE,* Sudhakar M S *

***Abstract*—The abundant deposits of unstructured and scarcely labeled data over social networks make text classification vital for structuring and extracting useful information. In addition, ignoring dialectal variations significantly hinders the performance of international English (especially American and British) text classification across numerous data domains. To address this multifaceted challenge, a comprehensive and adaptable framework termed Dialectic Feature-based Fuzzy Graph Learning (DFFGL) is introduced that learns feature vectors by inculcating semantics and dialect variations from the inputted text. DFFGL then proficiently extracts uniquely modified terms frequency-inverse document frequency, parts-of-speech-Tagged $\mathcal{N}-$grams, with dialect-specific dictionary features in the fuzzy feature space to realize a novel language model. Later, these fuzzified features are affined by a novel fuzzy distance measure to construct an interpretable fuzzy graph that is then optimized using a novel elastic net regularizer for characterizing nodal relations, promising efficient classification through effective label propagation. Exhaustive $F1-$score evaluations on 6 English corpora and 17 diverse datasets reveal DFFGL's superiority in consistently registering over 93% and 80% in dialect identification and text classification even with just 10 labeled samples. Furthermore, DFFGL offers remarkable $F1-$score improvements of 10.2% and 17.3% over its peers in respective tasks, highlighting its extension to real-world data classification.**

***Index Terms*—American and British English (ABE), Dialectic Feature-based Fuzzy Graph Learning (DFFGL), Dialect Identification (DI), Fuzzy Graph (FG), Label Propagation (LP), Modified Term Frequency Inverse Document Frequency (MTFIDF)**

I. INTRODUCTION

DIGITIZATION and growth of social networks have led to the exponential accumulation of unstructured text data capturing diverse and complex societal interactions. However, managing and mining this intricate information, spanning news reporting, opinion analysis, and medical communications poses significant challenges. These broadened challenges were effectively tackled upon the emergence of Text Classification (TC) by unlocking valuable insights from the input text and permitting efficient organization through clustering and annotation [1], [2], making it extremely essential in sentiment analytics [3], spam filtering [4], and classifying news [5]. Despite its remarkable achievements, TC extension to applications such as language or dialect analysis, topic detection, and authors' identity detection remains unexplored. These applications require the exploration of text context that varies demographically based on the words/phrases/sentences [6], [7]. The advent of Graph-based semi-supervised learning (GSSL) in TC has promised researchers [8]–[11] by cleverly classifying the text samples using just a few labeled examples thereby strengthening its extension to such domains. GSSL treats the Feature Vectors (FV) of text documents as nodes in a graph, interconnected by edges that reflect their semantic similarity. GSSLs' remarkable performance is hindered when handling large datasets [9], [12], [13] with limited label information [14], and the choice of distance measures [15].

Cherukula Madhu, Sudhakar M.S. (Corresponding author) are with the School of Electronics Engineering (SENSE), VIT, Vellore, 632014,Tamil Nadu, India. (e-mail: cherukulamadhu@gmail.com, sudhakar.ms@vit.ac.in)

Also, the characterization of nodes and edges in graphs built from the text is highly essential for affinity learning and representational sparsity [2], [16]. Accordingly, this work employs the textual FVs extracted from the text documents to represent nodes with their semantic affinity represented as edges. To achieve this, firstly the documents are converted to FVs by transforming the textual features using the vector space model (VSM) [17]. VSM's transformation failed whenever it encountered text containing longer sentences with ignored contextual information. Alternately, the term weighting concept [16] merged normalized term frequency (TF) with inverse document frequency (IDF) using cosine constraint [18] to ensure the term weights ranged between $[0\ 1]$ for vectorization, thereby, overcoming the text length issue. However, the method's performance deescalated when dealing with polysemous words insisting on contextual analysis [19]. Another TF variant [20] propagated relevant class labels blended with TF weights that improved the classification accuracy, while the supervised vectorization led to biased outcomes with increased complexity. The aforementioned discussions signify the weakness of the prevailing word representation in terms of contextual and demographic variations [6], [7], rather than concentrating more on the TF.

Lately, the black box deep learning models have been actively engaged in several TC tasks [21]. Rather, its fairness or transparency in decision-making is a major reason for concern when extending to real-world situations. Also, their poor judgments and lackluster explanations in predictions, make them unreliable [21]. Alternately, the inherent flexibility and interpretability of fuzzy rule-based systems with precise understandability make them a suitable choice in applications such as knowledge extraction and decision support from expert knowledge or data [22]. However, a fuzzy system modeled utilizing expert knowledge limits its performance, as they force membership functions in each rule to be in a common set whereas the data-based models address these shortcomings by compromising on their interpretability [23], [24]. Hence, underlying the demand for a model to strike a balance between accuracy and interpretability [25]. This issue is addressed by replacing fuzzy linguistics with fuzzy variables operating using minimal rule sets which are equivalent to FGs [26], [27]. Likewise, this intention replaces the crisp graph (CG) in GSSLs

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3421595

with fuzzy graph (FG) to ensure interpretability, flexibility, and adaptability [28] by enforcing dialectic variations in improving TC.

However, the extant TC models overlook dialectical variations while classifying international English text especially American English (AE) and British English (BE) [29]. Conversely, dialects play a crucial role in ensuring clarity, particularly in critical domains like news reporting, medical communication, and social media transactions. Nuances between BE and AE often lead to misinterpretations, with potentially severe consequences. For instance, in healthcare, precise understanding is vital for patient safety. Studies have shown how dialect variations on social media platforms like Twitter during the COVID-19 pandemic caused confusion in public health messages [29]. Also, the frequently used terms in a particular region reflect the characteristics of that location [6], [7], [29], [30]. This highlights the critical need for dialect awareness to achieve linguistic clarity, especially during human-human interactions (HHI) and human-computer interactions (HCI). Hence, identifying dialect-specific vocabulary within a text offers a valuable tool for both dialect determination and enhanced TC. Explicitly capturing these dialectal variations accounts for the richer contextual meaning of terms, thus improving classification efficiency. This, however, remains an unexplored avenue in current TC approaches. However, to highlight the importance of the contributions, herein a summary of the traditional and trending literature on text vectorization along with the recent GSSL schemes is ordered chronologically in Section I-A.

A. Related Work

One of the primal vectorization techniques introduced in TC is the simple binary weighting scheme that binarized word occurrences as 1s and 0s based on their presence and absence respectively in a given document. The binarization highly oscillates word connectivity by neglecting the TF which is unreasonable. Instead, TF assigns large weights to common terms that weaken the text discrimination thereby declining its performance [1], [2]. To compensate [18] combined TF with IDF coined TFIDF performed unsupervised term weighting and is widely popular due to its prioritized term weighting concept based on their frequency of occurrence. Later [31] replaced the unsupervised weighting with the supervised, labeled improved inverse gravity moment (IGM) that decreased the discerning potency of words/phrases in texts with performance and increased computational complexity. These text vectorization schemes concentrate on IDF modification [31] rather than altering TF which guarantees improved model efficiency. Likewise, the square root of TF (sqrtTF) [1] blended with IDF outperformed the IGM. Similarly, the sqrtTF in [1] coupled with the distinguishing feature selector (DFS) in [2] retained both the modified TF and IDF terms aiding text vectorization. The above discussions outline the demand for designing equally prioritized weighting schemes for the TF and IDF terms insisting on deep exploration.

The other dimension of concern in TC is Label Propagation (LP), which is generally accomplished using Graph-based representations, for describing various entities and their relationships quantified by affinity [32]. The GSSL-based LP [33] performed dimensionality reduction by two-stage iterative minimization, based on the availability of FV class information, and its accuracy was dependent on the labeled samples' strength. Likewise, [34], determined the edges by unifying the pair-wise distances with the estimated density and was regularization sensitive. Alternatively, the non-negative sparse graph (NNSG) [35], formulated the margin-based discriminant embedding model using Euclidean distance registered declined performance with added complexity. Likewise, Positive and Unlabeled Learning (PUL) [36] categorically binarized the labeled and unlabeled documents while its accuracy deteriorated in comparison with one-class learning models such as k-Nearest Neighbor density (KNND), k-means, and Dense Auto Encoder (DAE).

Moreover, the recent TC frameworks [37]–[43], utilize count vectorizer, TFIDF, Word-to-Vector (Word2Vec), Global Vector (GloVe), and Bidirectional Encoder Representations from Transformers (BERT) or their combination for representing text numerically. Particularly, the count-based and TFIDF vectorizers [37]–[39] majorly miss syntactic cues and relationships, neglecting context and nuances. Likewise, Word2Vec, GloVe, and BERT feature-based classifiers [40]–[43] lack interpretability, hindering bias detection and understanding predictions. Specifically, Word2Vec and GloVe's static word embeddings fail to capture dynamic contextual information. Similarly, the traditional models (SVM, MNB, LR, KNN, and LR) [37], [43], struggle with scaling and complexity, while deep learning models (ANN/CNN) [38], [40]–[42] lack interpretability. Also, Graph-based semi-supervised learning (GSSL) [38] relies heavily on available labels and dataset size, impacting performance.

The above discussions insist on the need for 1) an efficient text vectorizer or a language model capable of distinguishing and resonating ABE dialects, 2) an interpretable and flexible graph learning model, and 3) an effective regularizer warranting a better trade-off between computational cost and performance. Accordingly, to achieve the core intention of TC of International English reposited across several social and media platforms (news, Twitter, medical, and opinions, etc.,), and to address the shortcomings of the erstwhile classification frameworks, the ensuing research objectives (RO) are formulated:

- RO1: Develop a domain-independent vectorizer offering enriched textual representations by capturing deeper semantic information and nuanced features, while mitigating data sparsity.
- RO2: Word prioritization for the distinction of International English dialects (AE & BE) supplementing dialect recognition.
- RO3: Build unique language models to induce dialectical variations into the FV
- RO4: Construction of an inherently dynamic and interpretable graph, with convex optimization and regularization.
- RO5: Creation of an interpretable (transparent) framework to classify massive volumes of text efficiently with fewer computations.

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3421595

To address the ROs listed above, the manuscript presents the following contributions aimed at efficient TC across domains.

1) RO1 and RO2 are met by implementing three distinct vectorizers:
   - Parts-of-Speech (PoS)-tagged $\mathcal{N}-$grams for enriching the semantic vector information.
   - Prioritizing the least significant terms by tweaking conventional TFIDF, to uncover hidden text aspects contributing to specific classes.
   - Construction of individual AE and BE word dictionaries, supplementing Dialect Identification (DI).
2) RO3 is addressed by developing two different language models dedicated to DI and TC, by coalescing Modified Term Frequency Inverse Document Frequency (MTFIDF) and Hybrid $\mathcal{N}-$grams with dialect-specific dictionary features for comprehensive text representation.
3) To fulfill RO4 and RO5 the FG-based LP is performed using a leveraged Fuzzy Distance Measure (FDM) with optimal similarity learning via elastic net regularizer.

The above contributions guarantee higher efficiencies with interpretability as relatively witnessed in Sections IV-C and V-C respectively. The remainder of the paper is structured as follows: Section II introduces the engaged fuzzy concepts in the context of TC along with its notations. Dialectic Feature-based Fuzzy Graph Learning (DFFGL) formulation is decomposed under the following sub-headers: class-wise feature extraction, dialect labeling, the building of a novel FDM, and graph structuring employing a unique cost function in Section III. Section IV relatively validates the DI and TC on different national corpora with the traditional and contemporary predecessors. Section V establishes the model's simplicity and understandability examined in terms of computational complexity and interpretability, and Section VI culminates with the work's efficiency and uniqueness.

II. Preliminaries

The fundamental concepts of fuzzy logic and FGs engaged for the formulation of DFFGL are introduced below.

A. Notations

The adopted representations with notations are summarized in Table I.

B. Fuzzy Theory

Fuzzy set theory mathematically fuzzifies everyday linguistics by mapping each element from the universal set U to $[0\ 1]$. Accordingly, an element of fuzzy set F, corresponding to the element u of U is represented as an ordered pair as in (1)

$$F = \{(u, \mu_F(u))\,|u \in U\} \tag{1}$$

where $\mu_F(u) \in F$ is the degree of membership for the crisp element $u \in U$ and according to fuzzy set theory, for all $u \in \mathbb{R},\ \ \mu_F(u) \in [0\ \ 1]$ holds. Also, F is said to be a convex fuzzy set, if it satisfies the condition in (2) for all $\alpha \in [0\ \ 1]$.

$$F(\alpha u + (1-\alpha)v) \geq \min\{\mu_F(u), \mu_F(v)\} \tag{2}$$

TABLE I
Symbols and their Descriptions

Symbol	Description
ω	Word
$\alpha_0,\ \alpha_1$	Weights involved in Language Model
$\beta_0,\ \beta_1,\ ...$	Weights in defining the term frequency
η	Entropy of text document
d	Text document
T	Threshold to segregate the terms
$\mathcal{X}$	Dataset
n	Total number of FVs (Documents) in a Dataset (Corpus)
$d_{ij}^{\mathcal{X}}$	Fuzzy Distance measure obtained from $\mathcal{X}$
$x_i,\ x_j$	i^{th} and j^{th} FVs in the in the dataset $\mathcal{X}$
$\\|x_i - x_j\\|^2$	Euclidean distance between x_i and x_j
$\mu_{\mathcal{X}}(x_i)$	Fuzzified version of i^{th} FV in $\mathcal{X}$
σ_x	Standard deviation of an FV
$x_l,\ x_{ul}$	Number of labeled and unlabeled FVs in the dataset
$\mathcal{G}$	Fuzzy Graph
$\mathcal{V}$	Vertex Set
$\upsilon,\ \varepsilon$	Membership values of vertices and edges respectively
$S_{ij}^{\mathcal{X}}$	Similarity Matrix
$\lambda_1,\ \lambda_2$	elastic net regularization coefficients
$\gamma,\ \psi$	Lagrangian Multipliers
k	Nearest neighbors

Similarly, the relation R between the two fuzzy sets F and G is the cartesian product $R \rightarrow F \times G$ defined in (3)

$$\mu_R(u,v) = \mu_{F\times G}(u,v) = \min(\mu_F(u), \mu_G(v)) \tag{3}$$

The affinity between two fuzzy sets is given by their fuzzy relation R, which is a crucial quality for LP and FG building.

FG of Zadeh [44], is generalized here to construct an interpretable FG (IFG) with fuzzy FVs having membership values $\mu_{\mathcal{X}}(x_i)$ representing nodes bounded in $[0\ 1]$ [28]. Also, if the probability of occurrence of two nodes is similar or their degrees of memberships $(\mu_{\mathcal{X}}(x_i))$ are closer, then an edge is formed. Further, the non-empty vertex set $\mathcal{V}$ with membership values $\{\upsilon : \mathcal{V}|0 \leq \upsilon \leq 1\}$ and their relation $\{\mu \rightarrow \mathcal{V} \times \mathcal{V}|0 \leq \mathcal{V} \leq 1\}$, representing edges form an FG triplet $\mathcal{G} = (\mathcal{V}, \upsilon, \varepsilon)$ and should satisfy the condition given in (4).

$$\varepsilon(a,b) \leq \min\{\upsilon(a), \upsilon(b)\}\,|\,\forall(a,b) \in \mathcal{V};\ 0 \leq \upsilon, \varepsilon \leq 1 \tag{4}$$

Moreover, DFFGL learns similarity and labels k nearest neighbors which is equivalent to the number of FV in a class, constrained as $k \ll n$ offers improved interpretability. Upon constructing FG from the developed FDM ($d_{ij}^{\mathcal{X}}$), the optimal similarity matrix ($S_{ij}^{\mathcal{X}}$) is built using the modeled objective function. **(An instance of FG Construction for a snippet of dataset is presented in Supplementary Material)**

III. Methodology

Prevailing GSSL schemes neglect interpretability which is an extremely essential quality in LP determining the model's efficiency. Instead, these models propagate labels by relating edges based on non-probabilistic distances. To overcome these shortcomings, a unique FG model illustrated in Fig. 1 is introduced.

The DFFGL processing modules in Fig. 1 mainly focus on two things 1) Determining the dialect of text documents and

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3421595

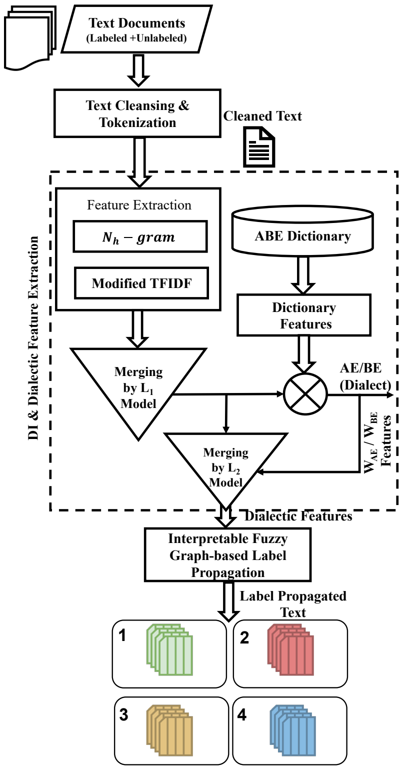


Fig. 1. Flow diagram of DFFGL Framework Assisting TC

2) LP to the text corpus with the limited labels. To achieve this, the text corpus undergoes the following preprocessing steps for handling noise and redundancy with simultaneous preservation of relevant information for DI and TC.

Text Cleansing & Tokenization:

- Case folding (e.g. Ab $\rightarrow$ ab) to eliminate case sensitivity, and simplify dictionary matching.
- Elimination of punctuations, common stop or short words (less than 3 characters), and long words (exceeding 15 characters) as they do not contribute significantly to the semantic content and often represent noise, typos, or irrelevant terms.
- Sentence-level tokenization based on "." delimiter to decompose text into smaller units for further processing followed by PoS-tagging.

DI & Dialectic Feature Extraction: To capture semantically rich contextual information, DFFGL adopts two distinct vectorizers generating unique and distinct FVs that are subsequently fused using the introduced language model (L_1). The resultant along with the pre-constructed AE and BE word dictionaries are inputted for DI followed by the fusion of L_1 and respective dictionary features using L_2 to produce dialectic FVs that are dialect-specific.

FG Construction & LP: To induce interpretability in the dialectic FVs, the Adaptive Bezier Curve-based Membership Function (ABCMF) [22] is adopted for fuzzification. The resulting fuzzy FVs, serve as nodes of the constructed FG with the edges representing the semantic similarity between them, quantified by a customized similarity function. The formulated FG is then optimized using an elastic-net regularizer followed by LP for TC.

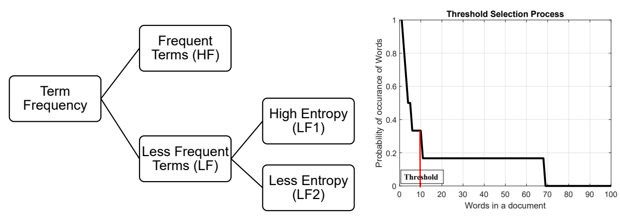
A. Feature Extraction

The introduced DFFGL initiates feature extraction for determining the text dialect and assists in LP. Especially for DI, the FVs need to capture various linguistic characteristics for contextual vector representations benefiting downstream tasks. Therefore, an ABE word dictionary is realized from the text phrases, and vocabulary deciding the language variant [45] in the extracted features [29]. Accordingly, DFFGL adopts a simple feature extraction scheme distinguishing ABE dialects by blending the features from three different modules namely MTFIDF for maintaining the text structure, PoS-tagged $\mathcal{N}-$ grams (word sequences) for syntactic analysis, and finally dictionary-based features for capturing semantic variations. Detailed discussions of these processing modules are dealt with under the relevant headers.

1) MTFIDF: Computer-aided TC requires an algebraic model to represent the documents as a vector as inferred from the recent vectorization versions reviewed in Section I-A. Particularly, the TFIDF gained wider research interest in a shorter time due to its simplicity and effectiveness in prioritizing common words. However, the omission of syntactic cues and relationships insisted on its modification thereby, seeking suitable alternatives. Accordingly, the DFFGL adopts a simple yet efficient weighting scheme presented in (5).

$$MTF = \beta_0 + \beta_1 LF_1 + \beta_2 LF_2 + \beta_3 HF \tag{5}$$

The chosen model assigns lower, and higher weights to the frequently and scarcely occurring words. The process for bifurcation of terms for term weighting and threshold selection are presented in Fig. 2a and Fig. 2b respectively.



(a) *Proposed term weighting scheme* (b) *Threshold Selection*

Fig. 2. Term Weighting Procedure

The introduced model bifurcates all the words (ω) in the text into 3 groups across two stages as presented in Fig. 2a. In the first stage, the terms or words are split into High Frequency (HF) and Low Frequency (LF) terms based on their probability distribution $(p(\omega))$ using a threshold T. Threshold selection for segregation in first stage is based on the term probability transformed into histograms that is normalized by the total number of words present in the document producing its envelope as shown in Fig. 2b. This distribution curve resembles a decaying step function, from which the threshold T intercepting the x-axis is selected after two steps of decay as highlighted in red color. The choice of thresholding is owed to the fact that the scarce terms exhibit lesser probabilities when compared with frequent terms that occur before the second

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3421595

decaying step. Hence, this position is chosen as T in all the text corpora utilized for framework validation to separate the LF and HF terms calculated using (6) and (7).

$$HF = p(\omega) \leq T \tag{6}$$

$$LF = p(\omega) > T \tag{7}$$

Subsequently, in the second stage, the LF terms are further split into LF1 and LF2 based on their entropy (η) represented as $\eta(LF)$, and are determined using (8) and (9).

$$LF_1 = \eta(LF) \geq \max(\eta) - mean(\eta) \tag{8}$$

$$LF_2 = \eta(LF) < \max(\eta) - mean(\eta) \tag{9}$$

Finally, LF1, LF2, and HF terms are weighted with β_1, β_2, and β_3 respectively constrained as $\beta_1 > \beta_2 > \beta_3$ prioritizing less frequent high entropy LF1 terms. The less frequent terms with the highest entropy attained from the above process help in discerning the features, thereby elevating them that are more specific to a particular class or dialect guaranteeing improved classification performance. To achieve this the regression coefficients β_0, β_1, β_2, β_3 are adopted from the Blackman-Harris window [46], as shown in (10).

$$\begin{aligned} W(n) = 0.35875 - 0.48829 \cos\left(\tfrac{2\pi n}{N}\right) + \\ 0.14128 \cos\left(\tfrac{4\pi n}{N}\right) - 0.01168 \cos\left(\tfrac{6\pi n}{N}\right) \end{aligned} \tag{10}$$

where the fundamental frequency is packed with the highest magnitude in comparison with the harmonics followed by their assignment to the high entropy less frequent terms. The choice of the Blackman-Harris window for term weighting is attributed mainly to the suppression of the sidelobe level to a large extent with simultaneous maximization of the roll-off factor [46]. Fitting this arrangement enhances the TF weighting scheme by adapting to their frequency distribution. The capturing of significant terms using the adopted window sensitizes term distribution, adaptability, and non-linearity. This process mitigates information loss and enhances the representation of term importance in documents. Accordingly, the window coefficients in (10) replace the regression constants in (5) with the TF modified as in (11).

$$\begin{aligned} MTF = 0.35875 + 0.48829 LF_1 + 0.14128 LF_2 + \\ 0.01168 HF \end{aligned} \tag{11}$$

Also, to address the shortcomings of the TF [47], the inverse document frequency (IDF) is evaluated using document frequency $DF(\omega)$, that furnishes the information about the occurrence of a word ω in all the documents as in (12).

$$IDF(\omega) = \log\left(\frac{n}{DF(\omega)}\right) \tag{12}$$

n - total number of documents in the corpus. Finally, the FV corresponding to ω in document d is the product of TF and IDF as given in (13).

$$FV(d) = MTF(\omega_i, d) \times IDF(\omega_i) \tag{13}$$

The formalized MTFIDF represents the structural and syntactic text information of the given document. However, to account for the semantic information the relevant feature extraction modules are dealt with below.

2) Hybrid $\mathcal{N}-grams$: The semantic information provides a better understanding of the word's context in a sentence, hence, it is essential to infuse the same in FV in addition to the structural and syntactic text information. Accordingly, the conventional $\mathcal{N}$ − grams are tagged with PoS to embed the semantic information in FVs by determining the word relations grammatically. Herein, the value of $\mathcal{N}$ remains vital in determining the classification accuracy, as the two-word phrases (bigrams) and single words (unigrams) do not contribute much to identifying the idiosyncratic differences, and PoS in a sentence. Hence, the value of $\mathcal{N}$ is chosen greater than 2 (3 to 6). These $\mathcal{N}$ − grams are further strengthened by hybridizing the conventional word sequences with respective PoS tags p_i in the word context for an hybrid $\mathcal{N}$ − grams ($\mathcal{N}_h$ − grams) in (14).

$$\mathcal{N}_h - gram = w_1 w_2 w_3 p_1 p_2 p_3 \tag{14}$$

Finally, $\mathcal{N}_h$ − grams are vectorized using MTFIDF to obtain the FV. For a text instance of the BNC corpus are shown in Fig. 3.

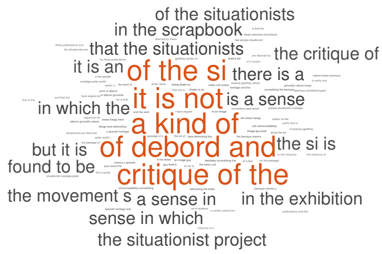


Fig. 3. Trigram Features of a text from BNC corpus

From the word cloud shown in Fig. 3, it is clear that the most frequent trigram combination is given the highest priority which is distinguished based on their font size based on the probability of occurrences. The $\mathcal{N}$ − grams solely is insufficient to characterize text, therefore, the phrases created using $\mathcal{N}_h$ − grams are furthered for determining the dictionary-based features by finding their semantic relations and remain highly essential in TC for identifying the text origin [48].

3) Word Dictionary Features (WD): Languages or dialects are well distinguished with vocabulary and phrases present in the text, so, herein a third set of features organized in a dictionary are extracted to get semantic relations between the words. The fabricated word dictionary comprises vocabulary and idioms collected from the English Oxford Living Dictionaries [49] and Wikipedia [6] separately for AE & BE. Later, binarized $\mathcal{N}$ − grams are obtained based on the words/phrases identified in a text matching with the dictionary as in (15). **(The word dictionary Feature Extraction process is detailed with an example in Supplementary Materials)**

$$\begin{aligned} W_{AE} &= \begin{cases} 1, & \text{if} \quad \mathcal{N} - \text{gram} \in AE \\ 0. & \text{elsewhere} \end{cases} \\ W_{BE} &= \begin{cases} 1, & \text{if} \quad \mathcal{N} - \text{gram} \in BE \\ 0. & \text{elsewhere} \end{cases} \end{aligned} \tag{15}$$

The resultant is then multiplied with the MTFIDF of the text, to determine the final dictionary features. Upon assem-

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3421595

bling the extracted diverse features in the form of FVs, the same is subjected to DI discussed in the below section.

B. Dialect Identification

After the formulation of three unique and distinct individual features (detailed in Sections III-A1, III-A2, and III-A3 respectively), the dialect of the inputted text is determined as showcased in Fig. 1.

Herein, the individual MTFIDF and $\mathcal{N}_h$ − grams features are merged using the language model defined in (16).

$$L_1 = \alpha_0 \left(MTFIDF\right) + \alpha_1 \left(\mathcal{N}_h - \text{grams}\right) \tag{16}$$

$\alpha_0,\ \alpha_1-$ equally weighted linear regression coefficients accommodating both contextual and structural features. Finally, the features attained from L_1 model are compared with the American and British dictionary-based features (WD_{AE} and WD_{BE}) utilizing a similarity function modeled in (17) and (18)

$$D_x = \cos\left(\frac{\pi}{2}\left\{\frac{1}{m}\sum_{k=1}^{m}\left\|L_1\left(i\right) - WD_{AE}\left(i\right)\right\|^2\right\}\right) \tag{17}$$

$$D_y = \cos\left(\frac{\pi}{2}\left\{\frac{1}{m}\sum_{k=1}^{m}\left\|L_1\left(i\right) - WD_{BE}\left(i\right)\right\|^2\right\}\right) \tag{18}$$

D_x and D_y - Dialect Correlation between L_1 with WD_{AE} and WD_{BE} respectively. Equations (17) and (18) represent modified version of cosine similarity derived utilizing the Euclidean distance between the features from (16) to WD_{AE} and WD_{BE} $\left(\left\|L_1\left(i\right) - WD_{AE/BE}\left(i\right)\right\|^2\right)$. The magnitude of W_{AE} and W_{BE} determines the closeness of the input text to a particular dialect as given in (19).

$$Dialect = \begin{cases} AE; & if \quad D_x > D_y \\ BE; & else \end{cases} \tag{19}$$

C. Text Classification

The assembled array of features representing the text extracted from the diverse modules needs to be efficiently classified. Accordingly, this work models an IFG framework for effective TC and LP to assign fewer known labels to the entire text corpus as shown in Fig. 4.

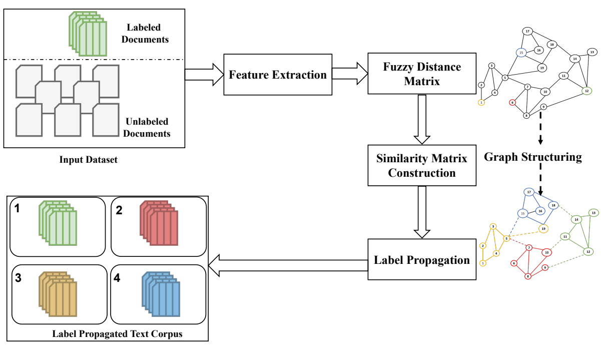


Fig. 4. IFG Framework for Label Propagation

At the onset, the composed features from the language model L_1 are merged with the dictionary-based features as in (20).

$$L_2 = \alpha_0 \left(L_1\right) + \alpha_1 \left(WD\right) \tag{20}$$

$\alpha_0, \alpha_1 \rightarrow$ regression coefficients with values same as L_1 model. The formulated FVs (termed dialectic features in DFFGL) are fuzzified utilizing the adaptive Bezier curve-based membership function [22] characterized using the envelope of the data distribution curve rendering closed approximation. Later, the attained fuzzy FV set $\mu_{\mathcal{X}}\left(x\right)$ with $\mathcal{X} = \{x_1, x_2, x_3, ..., x_l, x_{l+1}, ..., x_n\}$, constitutes a dataset for the whole corpus. Herein, the first l FVs are labeled from the set $Y = \{1, 2, 3, ..., c\}$, while the remaining unlabeled ul data points in $\mathcal{X}$ meet the constraint $l \ll ul$ and are stated in (21) and (22).

$$\mathcal{X}_l = \{\left(x_1, y_1\right), \left(x_2, y_2\right), \left(x_3, y_3\right), ..., \left(x_l, y_l\right)\} \tag{21}$$

$$\mathcal{X}_{ul} = \{x_{l+1}, x_{l+2}, ..., x_{ul}\} \tag{22}$$

With $\mathcal{X} = \{\mathcal{X}_l, \mathcal{X}_{ul}\}$, the labels are propagated using GSSL from l labeled to ul unlabeled fuzzy FVs using a novel cosine similarity-based FDM as elaborated in Section III-C1.

1) Cosine Similarity-based Fuzzy Distance Matrix: GSSLs are bestowed with unique and inherent properties like simplicity, and scalability. Despite these qualities, its performance mainly depends on the choice of the constructed distance matrix ensuring local connectivity or affinity between FVs in $\mathcal{X}$ [28]. Also, TC and DI demand the recognition of subtle linguistic variations in the text documents for effectively distinguishing the closely related dialects or terms with different contexts. Accordingly, the well-known Euclidean metric is adopted by the majority of the schemes available in the literature. Generally, Euclidean quantifies the FVs' magnitude deviations and perhaps fails to establish their relationships, while in the text context [50], [51] their magnitude signifies the document length. This dimensional loss is resolved using the cosine similarity in this approach wherein the angle between the vectors represents their affinity. Traditional cosine similarity measure emphasizes the relative direction of FVs, neglecting their magnitudes, and is sensitive to document length which is observed in traditional TFIDF metrics [52]. This impacts TC efficiency as different features have varying importance for categorizing FVs. A partial mitigation of this issue is achieved by modifying the TF, particularly by assigning higher importance to less frequent terms based on their entropy.

To address these issues, the DFFGL insists on assigning more weightage to the dialect-specific terms for capturing nuanced linguistic variations in the text. Hence, magnitude differences are incorporated in the cosine similarity measure through a squared distance term. This interplay between angular similarity (Cosine Function) and magnitude (Squared difference as in Euclidean distance) significantly augments the discriminatory capacity and addresses the disparities in document lengths that are commonly encountered in TC. This effective seizure of the direction and magnitude of FVs assisted in the efficient distinction of closely related dialects and text categories where semantic relationships and syntactic patterns

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3421595

are paramount. The squared distance measure is scaled with $\frac{\pi}{2}$ to normalize its range to $[0\ 1]$ enabling a clear interpretation of the similarity between dialect-specific features extracted from the L_1 model and the dialect dictionaries (W_{AE} and W_{BE}). This normalization of similarity scores between 0 and 1 by FDM enables consistent interpretation across diverse datasets and feature spaces, ensuring domain independence of DFFGL, and facilitating meaningful comparisons which is essential for accurate classification. Also, FDM's hybridization of merging squared feature differences with cosine similarity ensures robustness against feature variability, making FDM a valuable asset for various tasks involving fine-grained language understanding. Accordingly, the fine deviations in the fuzzified FVs' are capitulated using the new FDM tuned by the cosine distance measure presented in (23).

$$d_{ij}^{\mathcal{X}} = 1 - \cos\left(\frac{\pi}{2}\left\{\frac{1}{m}\sum_{k=1}^{m}\left\|x_i(k) - x_j(k)\right\|^2\right\}\right); \quad \text{for } i, j = 1, 2, ..., n \tag{23}$$

The similarity values obtained using (23) are constrained as $0 \le S_{ij}^{\mathcal{X}} \le 1$ satisfying the fuzzy theory, whereas cosine similarity elongates the range to $-1 \le S_{ij}^{\mathcal{X}} \le 1$, making it sensitive to outliers. Nonetheless, FDM's sensitivity in acquiring subtle word variations warrants effective distinction of closely related dialects and is crucial for specific NLP tasks such as syntactic parsing or grammar checking. The ablation studies presented in Section IV-C quantitatively support these advantages in warranting consistent and interpretable predictions. Additionally, the validity of the introduced FDM is assessed using Propositions 1-5 furnished below (**with their proofs in Appendix - A**).

Proposition 1. *Let x_i, x_j are any two FVs in the dataset $\mathcal{X}$, then the constructed FDM $d_{ij}^{\mathcal{X}}(x_i, x_j)$, $for\ i, j = 1, 2, 3, ..., n$, should satisfy the condition in 24*

$$0 \le d_{ij}^{\mathcal{X}}(x_i, x_j) \le 1 \ \forall \ 0 \le x_i, x_j \le 1 \tag{24}$$

Proposition 2. *The constructed FDM satisfies the condition of separability as in (25) for any fuzzy FVs x_i, x_j in the dataset $\mathcal{X}$.*

$$d_{ij}^{\mathcal{X}}(x_i, x_j) = 0, \ \text{if and only if } x_i = x_j \tag{25}$$

Proposition 3. *For the fuzzy FVs x_i, x_j from the dataset $\mathcal{X}$, the obtained distance measure must meet the condition given in (26)*

$$d_{ij}^{\mathcal{X}}(x_i, x_j) = d_{ij}^{\mathcal{X}}(x_j, x_i) \tag{26}$$

Proposition 4. *If x_k is a fuzzy FV in and $\mathcal{X}$, and $x_i \subseteq x_j \subseteq x_k$ then then the FDM must satisfy the condition given in (27).*

$$d_{ij}^{\mathcal{X}}(x_i, x_k) \ge d_{ij}^{\mathcal{X}}(x_i, x_j) \ \& \ d_{ij}^{\mathcal{X}}(x_i, x_k) \ge d_{ij}^{\mathcal{X}}(x_j, x_k) \tag{27}$$

Proposition 5. *For any FVs x_i, x_j, x_k in the dataset $\mathcal{X}$, then $d_{ij}^{\mathcal{X}}(x_i, x_j) \le \max\left\{d_{ij}^{\mathcal{X}}(x_j, x_k), d_{ij}^{\mathcal{X}}(x_i, x_k)\right\}$.*

These propositions assist in generalizing the geodesic distance for fuzzy sets which is vital in formulating GSSL and supports identifying the shortest path between the nodes in the IFG. When dealing with unclear and ambiguous data prevailing in many real-world events, IFGs are quite beneficial, whereas their crisp counterparts offer poor support [53].

2) Similarity Learning: For similarity learning it is mandatory that two fuzzy FVs x_i, x_j in the dataset $\mathcal{X}$ are highly correlated if their $d_{ij}^{\mathcal{X}}$ is less and vice versa [54]. This condition is extended to the $S_{ij}^{\mathcal{X}}$ with the additional fulfillment of positivity, separability, symmetry, and triangular inequality properties presented in Propositions (1-5). Initially, $S_{ij}^{\mathcal{X}}$ is constructed by fuzzy complementing $d_{ij}^{\mathcal{X}}$ as shown in (28)

$$S_{ij}^{\mathcal{X}} = \cos\left(\frac{\pi}{2}\left\{\frac{1}{m}\sum_{k=1}^{m}\left\|x_i(k) - x_j(k)\right\|^2\right\}\right); \quad \text{for } i, j = 1, 2, ..., n \tag{28}$$

Also, the diagonal elements are zeroed to prevent the self-loops in IFG ($S_{ii}^{\mathcal{X}} = 0$) . IFG generalization requires the meeting of two constraints that are essential for optimization. 1) $S_{ij}^{\mathcal{X}}$ is constrained in the range $[0\ 1]$, denoting its equivalence to probabilistic bound and 2) the row-wise sum of $S_{ij}^{\mathcal{X}}$ must be equal to 1, which is met by deploying a SoftMax tuned optimization. This arrangement additionally ensures higher interclass deviations and simultaneously groups the intraclass FVs in the dataset $\mathcal{X}$ [55].

Upon formulating IFG using the developed $d_{ij}^{\mathcal{X}}$ the $S_{ij}^{\mathcal{X}}$ is constructed by tuning the objective function (29) constrained by elastic net regularization that helps in grouping the correlated FVs [56].

$$\begin{aligned} J\left(S_{ij}^{\mathcal{X}}, \lambda_1, \lambda_2\right) = \min_{S_i} &\sum_{i,j=1}^{n} S_{ij}^{\mathcal{X}} d_{ij}^{\mathcal{X}} + \lambda_1 \sum_{j=1}^{n}\left\|S_{ij}^{\mathcal{X}}\right\|_2^2 \\ &+ \lambda_2 \sum_{i,j=1}^{n}\left|S_{ij}^{\mathcal{X}}\right|_1 \frac{\left|\sigma_{x_i} - \sigma_{x_j}\right|}{\sigma_{x_i}\sigma_{x_j}} \\ s.t.\ \forall i, S_i \in [0\ 1];\ S_{ii} = 0;\ \& &\sum_{j=1}^{n}\frac{e^{S_j}}{\sum_{i,j=1}^{n} e^{S_{ij}}} = 1 \end{aligned} \tag{29}$$

The regularization parameters λ_1 and λ_2 are related as $\lambda_2 = (1 - \lambda_1)$ thereby, becoming convex. The first term in (29)captures the FV similarity whilst, the second term evades trivial overfitting solutions and provides better disparity between interclass FVs [33]. The third term promises higher interclass FVs deviations [57], [58]. Later (29) is minimized with respect to S_i producing (30)

$$\frac{dJ\left(S_{ij}^{\mathcal{X}}, \lambda_1, \lambda_2\right)}{dS_i} = 0 \Rightarrow S_{ij}^{\mathcal{X}} + \frac{\lambda_2\sigma_{x_{ij}} + d_{ij}^{\mathcal{X}}}{2\lambda_1} = 0 \tag{30}$$

where, $\sigma_{x_{ij}} = \frac{\left|\sigma_{x_i} - \sigma_{x_j}\right|}{\sigma_{x_i}\sigma_{x_j}}$. Finally, the overall objective function is presented in (31) upon reorganizing (31) in vector form as in [28], [34], [59].

$$J = \min_{S_i}\left(\left\|S_i + \frac{\lambda_2\sigma_{x_{ij}} + d_{ij}^{\mathcal{X}}}{2\lambda_1}\right\|_2^2\right) \tag{31}$$

Later, (31) is minimized to determine the regularization parameters using the Lagrangian multipliers introduced in (32) **(with in detail derivation presented in Appendix - B)**

$$\begin{aligned} L\left(S_i, \eta_1, \beta\right) = \frac{1}{2}&\left\|S_i + \frac{\lambda_2\sigma_{x_{ij}} + d_{ij}^{\mathcal{X}}}{2\lambda_1}\right\|_2^2 \\ &- \eta_1\left(\sum_{i=1}^{n}\frac{e^{S_i}}{\sum_{i=1}^{n} e^{S_i}} - 1\right) - \beta_i^T S_i \end{aligned} \tag{32}$$

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3421595

$\eta_1,\ \beta \geq 0$ are the Lagrangian multipliers. Solving (32) with respect to $S_i^{\mathcal{X}},\ \ \eta_1,\ \beta$ results in (33)- (35).

$$\frac{dL}{dS_{ij}} = 0 \Rightarrow S_i + \frac{\lambda_2 \sigma_{x_{ij}} + d_{ij}^{\mathcal{X}}}{2\lambda_1} - \eta_1 \left(\sum_{j=1}^{n} \frac{e^{S_i}}{\sum_{j=1}^{n} e^{S_i}} \left(1 - \frac{e^{S_i}}{\sum_{j=1}^{n} e^{S_i}} \right) \right) - \beta = 0 \tag{33}$$

$$\frac{dL}{d\eta_1} = 0 \Rightarrow \sum_{j=1}^{n} \frac{e^{S_i}}{\sum_{j=1}^{n} e^{S_i}} = 1 \tag{34}$$

$$\frac{dL}{d\beta_i} = 0 \Rightarrow S_i = 0 \tag{35}$$

Finally, the optimal $S_{ij}^{\mathcal{X}}$ in (36) is determined upon satisfying the Karush–Kuhn–Tucker (KKT) conditions [28], [34], [59].

$$S_{ij}^{\mathcal{X}} = \left(-\frac{\lambda_2 \sigma_{x_{ij}} + d_{ij}^{\mathcal{X}}}{2\lambda_1} + \eta_1 \left(\frac{n-1}{n^2} \right) \right)_+ \tag{36}$$

The resulting dense $S_{ij}^{\mathcal{X}}$ is made sparse by selecting the first k closest neighbors from the unlabeled FVs while the remaining FVs are made zeros. To achieve this, the elements in $d_{ij}^{\mathcal{X}}$ are sorted in ascending order, and the top k elements are selected followed by LP in the later stages.This process is mathematically stated in (37), (38)

$$S_{ik}^{\mathcal{X}} = -\left(\frac{\lambda_2 \sigma_{x_{ik}} + d_{ik}^{\mathcal{X}}}{2\lambda_1} \right) + \eta_1 \left(\frac{n-1}{n^2} \right) > 0 \tag{37}$$

$$S_{ik+1}^{\mathcal{X}} = -\left(\frac{\lambda_2 \sigma_{x_{ik+1}} + d_{ik+1}^{\mathcal{X}}}{2\lambda_1} \right) + \eta_1 \left(\frac{n-1}{n^2} \right) \leq 0 \tag{38}$$

Subsequently η_1 is calculated as in (39) with the constraint that the row-wise sum is equal to 1.

$$\eta_1 = \frac{n^2}{k\,(n-1)} \left[1 + \frac{\sum_{j=1}^{k} \left(d_{ij}^{\mathcal{X}} + \lambda_2 \sigma_{x_{ij}} \right)}{2\lambda_1} \right] \tag{39}$$

Substituting (35) in ((33)) results in (40)

$$S_i = 0 \Rightarrow \eta_1 \left(\frac{n-1}{n^2} \right) - \left(\frac{\lambda_2 \sigma_{x_{ij}} + d_{ij}^{\mathcal{X}}}{2\lambda_1} \right) = 0 \tag{40}$$

Later substitution of η_1 in (40) yields the regularizing constant λ_1 in equation (41), with λ_2 evaluated using the relation $\lambda_2 = 1 - \lambda_1$ and is presented in (42).

$$\lambda_1 = \frac{k \left(d_{ij}^{\mathcal{X}} + \sigma_{x_{ij}} \right) - \sum_{j=1}^{k} \left(d_{ij}^{\mathcal{X}} + \sigma_{x_{ij}} \right)}{2 + k\sigma_{x_{ij}} - \sum_{j=1}^{k} \sigma_{x_{ij}}} \tag{41}$$

$$\lambda_2 = \frac{2 + 2\left(k\sigma_{x_{ij}} - \sum_{j=1}^{k} \sigma_{x_{ij}} \right) + \left(k d_{ij}^{\mathcal{X}} - \sum_{j=1}^{k} d_{ij}^{\mathcal{X}} \right)}{2 + k\sigma_{x_{ij}} - \sum_{j=1}^{k} \sigma_{x_{ij}}} \tag{42}$$

Finally, the optimal similarity matrix $S_{ij}^{\mathcal{X}}$ is realized in (43).

$$S_{ij}^{\mathcal{X}} = \begin{cases} \frac{\sum_{j=1}^{k} d_{ij}^{\mathcal{X}} - k d_{ij}^{\mathcal{X}} + \lambda_2 \left(\sum_{j=1}^{k} \sigma_{x_{ij}} - k\sigma_{x_{ij}} \right) + 2\lambda_1}{2k\lambda_1}, & \text{for } j \leq k \\ 0, \text{ elsewhere} \end{cases} \tag{43}$$

The information of labeled nodes along with the membership values of graph edges is captured in the $S_{ij}^{\mathcal{X}}$. The labels for unlabeled FVs are derived from transferring label information from labeled data based on their closest immediate neighbors, followed by the secondary neighbors using the constructed IFG.

IV. Results and Discussions

The efficacy of the DFFGL model is studied using the $F1-$score (Harmonic mean of precision and recall) by conducting tests on different English text datasets.

A. Experimental Setup and Dataset Description

The DFFGL framework is evaluated for DI and LP to determine their applicability and reliability utilizing diverse text corpora. Accordingly, various datasets utilized in this work along with the number of documents and categories are tabulated in Table II. Herein, the first six corpora are

TABLE II
Dataset Description for DI and TC

Text Collection	Files	Classes
News on Web (NOW)	768	4
Global Web-based English (GloWbE)	440	4
Corona Virus	1560	4
iWeb	1040	5
TV	540	5
The Corpus of Contemporary American English (COCA) and British National Corpus (BNC)	960	8
Computer Science Technical Reports (CSTR)	299	4
Foreign Broadcast Information Service (FBIS)	2463	17
Ohsumed 0 (Oh0)	1003	10
Ohsumed 5 (Oh5)	918	10
Ohsumed 10 (Oh10)	1050	10
Ohsumed 15 (Oh15)	913	10
Ohsumed	7400	23
Reuters 0 (Re0)	1504	13
Reuters 1 (Re1)	1657	25
Reuters 8 (R8)	7674	8
Reuters 52 (R52)	9100	52
Syskill Webbert	334	4
Text Retrieval Conference 11 (TREC 11)	414	9
Text Retrieval Conference 12 (TREC 12)	313	8
Text Retrieval Conference 21 (TREC 21)	336	6
20 Newsgroups	20000	20
Movie Review (MR)	10662	2

from English text corpora [60] with two different varieties of English specifically utilized to test the proposed DI model. The remaining 17 text corpora obtained from diverse domains

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3421595

encompassing various categories are employed for LP investigations. In the context of LP, the experiments are carried out by varying the number of labeled FVs as $\{10,\ 20,\ 30,\ 40,\ 50\}$ for 20 newsgroup collections and $\{1, 5, 10, 20, 30\}$ for the remaining corpora with k (nearest neighbors) varying with the unlabeled FVs. Also, the regularization parameters are optimally determined using the Lagrangian function making this model more adaptive than the existing LP models [36].

For relative composition with K-Means and KNND [36], the thresholds are manually fixed with k ranging between values ranging from $[1\ 17]$ and for LP-PUL it is set to $3 + (4 \times p)\ \forall p \in [0,\ 1,\ ....,\ 7]$ with the hyperparameter α sequentially varying as $\{0.1,\ 0.3,\ 0.5,\ 0.7,\ 0.9\}$. Experiments were conducted by recording the averaged $F1-$score of 10 runs over 5 random splits selecting different labeled documents in each trail.

B. Baselines

To test the applicability of DFFGL, the framework utilizes the following baselines consisting of both black-box and transparent models for relative analysis:

- HG-BERT [61]: Novel hierarchical graph-based TC integrated with BERT, excelled in contextual embedding and providing complex relationships.
- TextFCG+BERT [62]: Merged TextFCG with BERT by fusing contextual information for TC.
- TextFCG [62]: The constructed comprehensive text graph enhanced GNNs efficiency by addressing transductive learning limits.
- CGA2TC [63]: Captured word relationships via contrastive learning, and adaptive augmentation to escalate model's robustness.
- TextSSL [64]: The sparse structure learning model addressed inductive TC challenges.
- LP-PUL [36]: Graph-based PUL excelled in semi-supervised TC and outperformed vector space models.
- TextQGNN [65]: Learnt graph representations in Quaternion Space, and performed TC using TextING.
- BERT + GAT [66]: Blended BERT with GAT that captured word relationships using graphs for TC.
- T-VGAE [67]: Captured structural information of text by leveraging topic models and variational autoencoders.
- DHTG [68]: Constructed a dynamic hierarchical topic graph, based on document topic modeling.
- HyperGAT [69]: Hypergraphs employed dual attention for efficient feature extraction assisting TC.
- TextING [70]: Performed inductive TC via dynamic graph construction with sliding windows.
- TextING-M [70]: Unlike TextING, this modified framework constructed a large graph for the whole corpus and generated the word embeddings.
- K-Means [36]: Non-parametric, clustering-based, boundary method for graph structuring and LP.
- K-NN Density [36]: Non-parametric, density-based, nearest-neighbor approach for graph structuring and LP.
- TextGCN [71]: Utilized pre-trained word embeddings, and applied GCN layers for graph structuring with node information.
- PU-LP [36]: A graph-based PU learning with Katz index, excelled with limited positive labels.
- RC-SVM [36]: A novel approach for classifying unlabeled documents using Rocchio and SVM.

C. Relative Analysis

The proposed framework is evaluated in two scenarios for DI and LP on various datasets. Research on DI or Dialect Classification has extensively explored local languages or regional dialects, while studies specifically focusing on AE and BE dialects within International English are scarce. This work addressed this gap by introducing dialectic features and its self-analysis in Terms of ROC parameters is presented in Table III.

TABLE III
AVERAGED ROC METRICS OF THE PROPOSED DI MODEL ON 6 DIVERSE ENGLISH CORPORA

Corpus	# Files	Precision	Recall	Accuracy	$F1-$Score
NOW	768	82.17	87.81	92.67	84.90
GloWbE	440	93.27	86.02	95.83	89.50
Coronavirus	1560	96.97	90.18	97.53	93.45
iWeb	1040	94.15	93.35	98.91	93.75
TV	540	91.27	94.23	96.68	92.73
COCA-BNC	960	89.34	90.13	95.56	89.73
Average		89.20	90.79	96.19	90.74

As seen in Table III, it is clear that the introduced DI model distinguishes the two English dialects efficiently as witnessed in the averaged $F1-$score greater than 90%. These higher metrics are owed to the developed language model that efficiently merges the syntactic and semantic structure of the text into the FV. However, this self-analysis is further supported by a relative analysis with two other DI approaches in Table IV.

TABLE IV
RELATIVE ANALYSIS OF DI ON BNC & COCA CORPORA

Method	Features	Classifier	Accuracy
Method 1 [7]	TFIDF + Dictionary	SVM	92.1
Method 2 [30]	$\mathcal{N}$ − grams	MNB	79.4
Method 3 (Proposed)	L_1 Model (Eqn. 16)	DFFGL	**95.56**

Bold corresponds to best.

From Table IV, it is evident that DFFGL efficiency supersedes its predecessors which are mainly due to the usage of the language model L_1 and the DFFGL. Particularly, MTFIDF and PoS-tagged $\mathcal{N}$ − grams in feature extraction are highly responsible for the accuracy improvement over the conventional TFIDF and $\mathcal{N}$ − grams-based features. Overall, upon observing Table III and Table IV it is evident that DFFGL promises distinction of ABE varieties accurately.

Similarly, the experimental evaluation of the DFFGL for class labeling is compared with different LP algorithms [36]. The experimental results for the proposed model along with its predecessors are relatively visualized and displayed in Fig. 5 (the quantitative analysis of the same using a Table is presented in Supplementary Materials).

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3421595

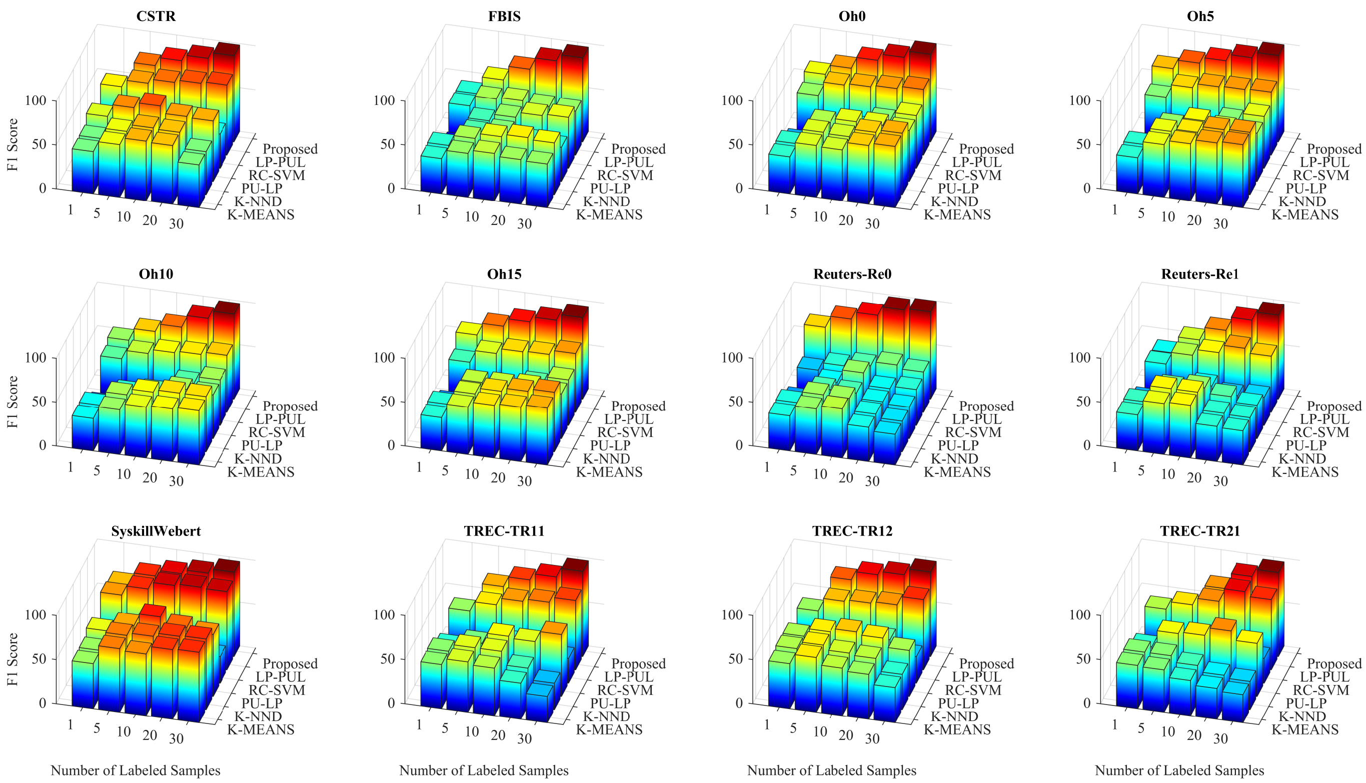


Fig. 5. Performance Evaluation of DFFGL for Class Labeling on Diverse Corpora

DFFGL outperforms its predecessors except for a few text collections as witnessed in Fig. 5, even when the the labeled FVs are 1. This dominance is owed to the MTFIDF-based FVs that were assigned higher weights to the less frequent terms when compared to frequent terms. As noticed in Fig. 5, peers achieved less $F1-$score if the number of labeled samples is one whereas DFFGL dominates even in this case for almost 10 text datasets. Moreover, this superiority sustains for the cases comprising the number of labeled FVs greater than one which is attributed to the elastic net-based regularization that had correlated similar data categories thereby escalating LP's performance. Furthermore, $d_{ij}^{\mathcal{X}}$ assists in decorrelating the interclass text thereby, recording supreme accuracies in comparison with its competitors. To demonstrate DFFGL's flexibility and reliability, the analysis is extended to large corpora [61], [71] and is compared with graph networks, and BERT models to demonstrate DFFGL's efficacy. Accordingly, the average accuracy of five random cross-foldings of unlabeled data is reported in Table VI.

It is apparent from Table VI, that DFFGL's accuracy is consistent in large corpora and in both binary and multiclass, when compared with recent graph representational learning utilizing convolutional networks, and transformer-based models. This is attributed to the structuring of IFG utilizing the introduced cost function regularized by elastic nets. Also, the dialectic features extracted utilizing the introduced word dictionary helped in distinguishing the interclass FVs with ease and made the framework surpass its competitors.

TABLE V
PERFORMANCE EVALUATION OF DFFGL FOR CLASS LABELING ON LARGE CORPORA

Method	Year	Ohsumed	R8	R52	MR	20 NG
DFFGL (Proposed)	-	**76.5**	93.6	**95.3**	**91.6**	**95.02**
HG-BERRT	2023	74.87	**98.75**	94.11	89.3	93.71
Text-FCG + BERT	2023	74.08	95.43	91.63	87.19	85.67
Text-FCG	2023	66.51	84.36	82.91	78.68	78.46
CGA2TC	2022	70.62	97.76	94.47	77.8	79.94
TextSSL	2022	70.16	90.34	86.54	82.94	81.5
TextQGNN	2021	69.93	97.02	94.45	78.93	-
BERT + GAT	2021	71.09	92.12	89.04	85.31	83.49
T-VGAE	2021	68.46	88.08	85.15	81.88	80.28
DHTG	2020	68.8	97.33	93.93	77.21	-
HyperGAT	2020	68.87	88.41	85.44	81.97	80.36
TextING	2020	70.35	90.53	86.57	83.23	81.61
TextING-M	2020	70.44	90.88	86.94	83.52	81.62
TextGCN	2019	68.36	97.07	93.56	76.74	86.34

Bold corresponds to best.

Overall, from Table III to Table VI, it is inferred that DFFGL, consistently categorized the text documents from diverse characteristic text corpora is owed to the introduced three-fold feature extraction. Moreover, the introduced FDM and the formulated objective function with elastic net regularization helped in correlating the intraclass features with distanced interclass FVs. To support these statements an ablation study is conducted to validate the feature extraction with familiar SVM classifier and graph-based models presented in Table VII.

To showcase the impact of dialectic feature actions and their absence on DFFGL performance the above ablation study was conducted and the results are tabulated in Table VII,

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3421595

TABLE VI
ABLATION STUDY FOR LP WITH DIFFERENT FEATURE EXTRACTION METHODS ($F1-$SCORE)

Dataset	SVM			LP via CG			LP via FG		
	TFIDF	MTFIDF	Dialectic	TFIDF	MTFIDF	Dialectic	TFIDF	MTFIDF	Dialectic
FBIS	85.5	87.4	90	87	88.6	90.7	86.2	91.2	**95.8**
Oh0	90.9	92.7	95.8	89.6	86.8	87.3	85.1	94.5	**99.8**
Re0	84.2	87.3	92.8	87.9	89.7	93.7	88.3	93	**99.5**
Re1	80.4	78.3	88.9	85.2	89.4	92.7	86.7	93.9	**95.7**
Re8	92.4	93.6	**94.1**	88.8	92.5	93	82.4	92.3	93.6
Re52	90.6	91.2	92.8	85.4	87.8	91.6	89.2	**96.2**	95.3
TR11	77	85.9	87.3	84.5	88.3	92.5	89.4	94.1	**97**
TR21	70.4	80.8	89.2	81.9	91	95.2	84.6	92.9	**95.7**
20NG	93	88.2	91.5	87.3	85.1	90.2	85.1	89.4	**93.2**
MR	**93.5**	86.6	88.4	78.7	84.6	89.2	87.3	88.3	90.6

Bold corresponds to best.

wherein the $F1-$scores of 3 different classifiers are validated with conventional TFIDF, the proposed MTFIDF and Dialectic features. From Table VII, it is hierarchically witnessed that Dialectic features superseded the MTFIDF followed by TFIDF across 10 datasets, except for the movie review dataset. This is likely because reviews primarily express sentiment rather than dialect, thereby witnessing a slight decline in the recorded $F1-$score by DFFGL. Also, the $F1-$score obtained by DFFGL using FG is ahead of the crisper version irrespective of the features, which is attributed to the novel cosine-similarity-based FDM and the elastic net-tuned optimization in fuzzy space.

In addition to extracted features, the proposed similarity measure also influences classifier performance. To analyze this impact, a DFFGL graph is constructed for each dataset containing 30 known labels per class using the proposed and traditional similarity measures, and the accomplishments are stated in Table VIII.

TABLE VII
ABLATION STUDY FOR LP WITH DIFFERENT SIMILARITY MEASURES ($F1-$SCORE)

Datasets	Cosine Similarity	FDM
FBIS	90.7	**95.8**
Oh0	93.4	**99.8**
Re0	91.8	**99.5**
Re1	90.3	**95.7**
Re8	86.2	**93.6**
Re52	91.6	**95.3**
TR11	94.5	**97**
TR21	93	**95.7**
20NG	88.2	**93.2**
MR	**93.7**	90.6

Bold corresponds to best.

It is witnessed from Table VIII, that FDM enhanced the performance of DFFGL by 4.5% when compared with cosine similarity-based DFFGL. This is attributed to the squared distance incorporated in the distance metric that ensured the closeness of nearest neighbors in the vicinity and distanced the non-neighbors thereby, enhancing its discerning capacity.

Further, to evaluate the performance improvements between model combinations (feature extractors with different classifiers), the Wilcoxon signed-rank test with a significance level of $\alpha = 0.05$ is performed for pair-wise analysis. This test identifies statistically significant differences between the performance of the two classifiers based on their paired results which are tabulated in Table IX.

From the statistical analysis, presented in Table IX, it is evident that the proposed DFFGL framework achieves the best performance. Also, when operated using the traditional cosine similarity it is ranked 5^{th}. This highlights the positive impact of the proposed cosine-based fuzzy similarity measure. Additionally, all DFFGL-based combinations outperform TFIDF-based classifiers, demonstrating the combined effectiveness of the proposed feature extraction and fuzzy-based similarity measure.

Also, to justify the adoption of 5-fold cross-validation, herein the tests are conducted on 4 diverse datasets by varying $\mathcal{K}$ values $\{3, 5, 7, 9, 10\}$ and is presented in Table X.

Upon analysis from the above table, it is witnessed that the DFFGL's performance is consistent for different folds. However, the change in the value of '$\mathcal{K}$' impacts time complexity. Also, each test is run 10 times to ensure consistency in $F1-$scores even in smaller datasets. Based on the investigations and to have a better trade-off between models' performance and computational burden 5-fold is chosen as optimum.

V. COMPLEXITY ANALYSIS

The realization complexity of DFFGL is better understood by noting the incurred computations along the time, and space dimensions using the Big $\mathcal{O}$ notation. Generally, the total time and space complexities are arrived at by summing up all the intermediate computations performed at different stages involved in the DFFGL formulation in DI and TC as stated in (44).

$$T_{total} = \sum_{i=1}^{n} T_i \simeq \max\left(T_i\right) \tag{44}$$

where T_i is the computational time required for each block and T_{total} is approximated to the maximum time required in all the individual stages.

A. Time Complexity

The time complexity of DFFGL is analytically modeled by decomposing it into individual steps. Initially, a few key parameters are assumed to characterize the dataset size (n), word counts in text files (ω) with average sentence length (L), and word dictionary size (d) .

1) Preprocessing: Case folding, stop/short/long word removal, and tokenization, exhibit linear complexity based on the text length given as $\mathcal{O}\left(\omega\right)$when considering the entire corpus the complexity rises to $5 \times \mathcal{O}\left(n\omega\right)$. Additionally, PoS tagging incurs $\mathcal{O}\left(n\omega L\right)$ operations owing to the influence of L. Therefore, the preprocessing complexity is the sum of the time required across six stages and is given in (45).

$$T_{Preprocess} \simeq \max\left\{5 \times O\left(n\omega\right), O\left(n\omega L\right)\right\} \simeq O\left(n\omega L\right) \tag{45}$$

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3421595

TABLE VIII
WILCOXON'S SIGNED-RANK TEST

	SVM TFIDF	SVM MTFIDF	SVM Dialectic	CG TFIDF	CG MTFIDF	CG Dialectic	FG TFIDF	FG MTFIDF	FG Dialectic (DFFGL)	FG Cosine Similarity	Rank
SVM TFIDF	1	0.908	0.984	0.541	0.821	0.949	0.581	0.982	0.995	0.967	7
SVM MTFIDF	0.111	1	0.998	0.063	0.620	0.979	0.380	0.996	0.997	0.978	6
SVM Dialectic	0.021	0.003	1	0.003	0.026	0.620	0.005	0.889	0.997	0.581	4
CG TFIDF	0.500	0.949	0.998	1	0.979	0.997	0.730	0.998	0.998	0.996	10
CG MTFIDF	0.207	0.419	0.979	0.026	1	0.998	0.077	0.997	0.998	0.979	8
CG Dialectic	0.063	0.026	0.419	0.004	0.003	1	0.003	0.730	0.998	0.390	3
FG TFIDF	0.459	0.658	0.996	0.305	0.939	0.998	1	0.998	0.998	0.998	9
FG MTFIDF	0.023	0.005	0.131	0.003	0.004	0.305	0.003	1	0.997	0.063	2
FG Dialectic (DFFGL)	**0.007**	**0.005**	**0.004**	**0.003**	**0.003**	**0.003**	**0.003**	**0.004**	**1**	**0.007**	**1**
FG Cosine Similarity	0.042	0.029	0.459	0.005	0.026	0.663	0.003	0.949	0.995	1	5

TABLE IX
ABLATION FOR $\mathcal{K}$-FOLD VALIDATIONS

Datasets	$\mathcal{K}=3$	$\mathcal{K}=5$	$\mathcal{K}=7$	$\mathcal{K}=9$	$\mathcal{K}=10$
FBIS	93.7363	95.7296	96.783	96.5198	95.8977
Oh0	97.7807	99.2765	95.9004	94.6609	93.2624
Re8	99.2527	99.3332	99.4914	99.3749	99.3933
TR11	96.7055	97.8688	99.7899	99.6867	99.7135
Average CPU runtime (in sec)	30.36	52.49	98.6	143.5	190.7

2) Language Model (L_1): The language model L_1 fuses the unique FVs from MTFIDF and hybrid $\mathcal{N}_h-$grams vectorizers. Initially, MTFIDF segregates words into HF and LF terms requiring $\mathcal{O}\left(\omega+b\right)$ time for each file, and when considering the corpus it scales to $\mathcal{O}\left(n\omega+b\right)$, where b is the unique word count. Further, LF is decomposed into LF1 and LF2 which requires $\mathcal{O}\left(nl_f\right)$, operations with l_f being the number of LF terms. On the whole MTF extraction requires $\mathcal{O}\left(n\left(\omega+b\right)\right)+\mathcal{O}\left(n\left(l_f\right)\right)$ time. Finally, MTFIDF results in the complexity given in (46).

$$T_{MTFIDF}=\mathcal{O}\left(n\left(\omega+b+l_f\right)\times\log\left(n\left(\omega+b+l_f\right)\right)\right) \tag{46}$$

Similarly, the complexity in obtaining hybrid $\mathcal{N}-$grams is the sum of the time required for extracting $\mathcal{N}-$grams and PoS-tagging as presented in (47)

$$T_{N_h-\text{grams}}=\mathcal{O}\left(n\omega\right)+\mathcal{O}\left(n\omega L\right) \tag{47}$$

Overall, the feature extraction time complexity using the language model L_1 is given in (49).

$$T_{L_1}=\max\left\{T_{MTFIDF},\,T_{N_h-\text{grams}}\right\} \tag{48}$$

$$T_{L_1}=\mathcal{O}\left(n\left(\omega+b+l_f\right)\times\log\left(n\left(\omega+b+l_f\right)\right)\right) \tag{49}$$

3) DI & Dialectic Feature Extraction: Upon extraction of L_1 features, DI is performed by comparing them with features derived from a predefined word dictionary, resulting in a complexity as in (50).

$$T_{DI}\simeq\max\left\{T_{L_1},T_{WD_{AE}},T_{WD_{AE}},T_{comparison}\right\} \tag{50}$$

At the onset, the dictionary feature extraction incurs a time complexity of $\mathcal{O}\left(n\omega Ld\right)$, for the entire corpus and it doubles for AE and BE dictionaries.Subsequently, the extracted dictionary features are compared with L_1 features that incur a quadratic time complexity of $\mathcal{O}\left(n^2\right)$. Hence, the overall complexity involved in DI is presented in (51).

$$T_{DI}=\mathcal{O}\left(n^2\right) \tag{51}$$

Upon DI, the dialectical feature extraction using the language model L_2 incurs a computational time stated in (52).

$$\begin{gathered}T_{L_2}\simeq\max\left\{T_{L_1},T_{dict}\right\}\\T_{L_2}=\mathcal{O}\left(n\left(\omega+b+l_f\right)\times\log\left(n\left(\omega+b+l_f\right)\right)\right)\end{gathered} \tag{52}$$

4) FG construction & LP: Upon dialectic feature extraction, the process constructs the initial graph using the distance metric to assess the FV similarity which requires $\mathcal{O}(n^2m)$ ($m\ll n$ - representing the number of FV elements)computations for the entire dataset $\mathcal{X}$ and is reduced to $\mathcal{O}(k^2m)$ by considering the optimized k nearest FVs. Accordingly, overall time complexity involved in graph structuring is presented in (53).

$$T_{FG}\simeq\mathcal{O}(n^2m)+\mathcal{O}(k^2m)=\mathcal{O}(n^2m) \tag{53}$$

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3421595

Therefore the total time complexity involved in DFFGL is the sum of all the individual computations performed at each stage and is mathematically presented in (54).

$$T_{total} \simeq \max\left\{T_{Preprocess}, T_{L_1}, T_{L_2}, T_{DI}, T_{FG}\right\} = \mathcal{O}\left(n^2 m\right) \tag{54}$$

Further to showcase DFFGL's implementation simplicity, relative time complexity analysis with the latest contemporaries is presented in Table XI.

TABLE X
COMPARISON OF TIME COMPLEXITIES OF DIFFERENT METHODS

Methodology	Computational Complexity
NNSG [35]	$\mathcal{O}\left(\begin{array}{c}\left(\begin{array}{l}n^3+\max\left\{n^3, n^2 c\right\}\\+iN^3\end{array}\right)t+\\(ndc)+\left(\begin{array}{c}2d^2 n+d^3\\+n^2 d+n^3\end{array}\right)\end{array}\right)$
LPSGL [33]	$\mathcal{O}\left(\max\left\{n^2, k^3\right\} t\right)$
GNMFLD [33]	$\mathcal{O}\left(t\left(mnk\right)+n^2 m+lk\right)$
PUL [36]	$\mathcal{O}(n^2 \log n)$
SSC-NGC [32]	$\mathcal{O}\left(t_2\left(\begin{array}{c}N^3+N^2 n+Nn^2\\+n^3+t_1 nc^2\end{array}\right)\right)$
DFFGL (Proposed)	$\mathcal{O}\left(n^2 m\right); m \ll n$

The complexities revealed in Table X justify DFFGL's simplicity and with flexibility surpasses its current and contemporary competitors thereby outlining the suitability of this model for real-time TC.

B. Space Complexity

Similar to the time complexity, the space complexity in realizing DDFGL is also evaluated across all stages. In the first stage, the feature extraction process requires a maximum of $\mathcal{O}\left(n\omega\right)$ space. Likewise, for determining TFIDF features, the same amount of space is required and finally in the third stage for storing FDM, $\mathcal{O}\left(n^2\right)$ space is required. So, the overall space requirement for the proposed work is given in (55)

$$S_{total} = \mathcal{O}(n\omega) + \mathcal{O}\left(n^2\right)$$

$$S_{total} \approx \mathcal{O}\left(n^2\right) \tag{55}$$

Even though the assessed time and space complexity in (54) and (55) are in quadratic order, they are considerably smaller than their counterparts. Alongside the computational efficiencies, the technique does not risk classification accuracy, making it more suitable for real-time data categorization. Overall from the analyzed complexities it is understood that DFFGL is relatively simpler in realization incurring minimal time and space when compared with its competitors.

C. Interpretability Analysis

Finally, the model's fairness in concluding the arrived decisions is assessed in terms of the interpretability I index [22], [28] on diverse text corpora given in (56).

$$I = 1 - C_F \tag{56}$$

Where $C_F = \frac{C_{rules}+C_{FS}+C_{inputs}}{3}$ represents the overall fuzzy complexity which is completely dependent on FG structuring, characterized by the number of fuzzy rules and sets reducing C_F to (57)

$$C_F = \frac{C_{rules} + C_{FS}}{2} \tag{57}$$

Wherein $C_{rules} = \frac{1}{T_{rules}}$ (with $T_{rules} = \frac{n(n-1)}{2}$) is the complexity associated with the number of rules executing sequentially, and $C_{FS} = \frac{fs_i}{T_{fs}}$ (with $T_{fs} = n$) is the fraction of the total number of fuzzy sets called at a time. By considering these factors, and utilizing 2 fuzzy sets at a time, C_F is finally reorganized in (58)

$$C_F = \frac{\frac{2}{n^2-n} + \frac{2}{n}}{2}$$
$$C_F = \frac{n^2 - n + n}{n^3 - n^2} \tag{58}$$

To better understand FGLPs reliability and expressiveness, the developed interpretability measure is examined on 4 datasets and presented in Table XII.

TABLE XI
INTERPRETABILITY ANALYSIS

S. No.	Dataset	# FVs	CF	I
1	CSTR	299	0.175	0.825
2	Oh0	1003	0.090	0.909
3	Oh5	918	0.033	0.966
4	Oh10	1050	0.013	0.986

From Table XII, it is evident that the interpretability involved in FG structuring is highly dependent on dataset dimensions. Furthermore, the elastic net-based optimization ensured optimal similarity learning by considering the nearest k FVs and neglecting $n-k$ FVs that add to DFFGL's interpretability that enhanced the fairness in LP.

VI. CONCLUSION

This research adopted a simple yet adaptable FG-based model for DI and LP. One of the key features of this contribution is the feature extraction methodology that helps in formulating a language model predicting the dialect of the given text. Also, the MTFIDF coupled with the cosine-based weighting scheme generated the best textual features. Moreover, the construction of the FDM detects graph edges that remain a significant contribution of this LP model in structuring the IFG. Later, for interclass separability, the newly framed elastic net-based cost function employing the SoftMax activation function renders broader data deviations and facilitates TC. Rigorous experiments on English text corpora reveal that the intended DI model efficiently distinguishes ABE dialects owing to the constructed word dictionary. Specifically, DFFGL $F1-$Score improvement of 14% over the recent competitors' transformer/deep learning models examined on diverse text corpora discloses its superior efficiency and consistency. Furthermore, the complexity and interpretability analysis strongly supports its suitability for real-time applications without sacrificing its performance at a maximum fairness of 98%. Despite

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3421595

the remarkable performance, the DFFGL $F1-$score declines in cases where the inputted text has typos leading to either out-of-vocabulary words or changes to different contexts hindering the effectiveness of the extracted features. These issues can be addressed by correcting the spelling by proper usage of PoS-Tagged $\mathcal{N}-$grams in the pre-processing stage. Likewise, longer documents with increased word count lead to higher FV length, impacting computational complexity. Hence, necessitating the exploration of suitable Dimensionality reduction or feature selection, or graph embedding techniques and remains the future scope of this work.

REFERENCES

[1] Chen, L., Jiang, L. & Li, C, "Using modified term frequency to improve term weighting for text classification," *Engineering Applications Of Artificial Intelligence*. vol. 101, May 2021, Art. no. 104215.

[2] Chen, L., Jiang, L. & Li, C, "Modified DFS-based term weighting scheme for text classification," *Expert Systems With Applications*, vol. 168, April 2021, Art. no. 114438.

[3] Hassonah, M., Al-Sayyed, R., Rodan, A., Ala'M, A., Aljarah, I. & Faris, H, "An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter," *Knowledge-Based Systems*, vol. 192, March 2020, Art.no. 105353.

[4] Gangavarapu, T., Jaidhar, C. & Chanduka, B, "Applicability of machine learning in spam and phishing email filtering: review and approaches," *Artificial Intelligence Review*, vol. 53, pp. 5019-5081, Feb 2020.

[5] Silva, R., Santos, R., Almeida, T. & Pardo, T, "Towards automatically filtering fake news in Portuguese," *Expert Systems With Applications*, vol. 146, May 2020, Art. no. 113199.

[6] Wikipedia, (2021, July): *Comparison of American and British English*, Accessed: 2022 Nov 02, [Online]. Available:https://en.wikipedia.org/wiki/Comparison of American and British English

[7] Utomo, M. & Sibaroni, Y, "Text classification of british english and american english using support vector machine," in *Proc.(ICoICT)2019 7th International Conference On Information And Communication Technology*, pp. 1-6, 2019.

[8] Shi, D., Zhu, L., Li, Y., Li, J. & Nie, X, "Robust Structured Graph Clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4424-4436, Nov. 2020.

[9] Ienco, D. & Pensa, R, "Enhancing Graph-Based Semisupervised Learning via Knowledge-Aware Data Embedding," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 5014-5020, Nov. 2020.

[10] Zhang, Y., Ji, S., Zou, C., Zhao, X., Ying, S. & Gao, Y, "Graph Learning on Millions of Data in Seconds: Label Propagation Acceleration on Graph using Data Distribution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 1835-1847, Feb. 2023.

[11] Vapnik, V. & Izmailov, R, "Rethinking statistical learning theory: learning using statistical invariants," *Machine Learning*, vol. 108, no. 3, pp. 381-423, 2019.

[12] Dornaika, F., Baradaaji, A. & Traboulsi, Y, "Joint Label Inference and Discriminant Embedding," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4413-4423, Sept. 2022.

[13] Yan, R., Zhang, J., Yang, J. & Hauptmann, A, "A discriminative learning framework with pairwise constraints for video object classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, pp. 578-593, April 2006.

[14] He, F., Wang, R. & Jia, W, "Fast semi-supervised learning with anchor graph for large hyperspectral images," *Pattern Recognition Letters*, vol. 130 pp. 319-326, Feb. 2020.

[15] Song, Z., Yang, X., Xu, Z. & King, I, "Graph-Based Semi-Supervised Learning: A Comprehensive Review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 8174-8194, Nov. 2023

[16] Zhang, C., Lin, Y., Chen, C., Yao, H., Cai, H. & Fang, W, "Fuzzy Representation Learning on Graph," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 10, pp. 3358-3370, Oct. 2023.

[17] Salton, G. & McGill, M, *Introduction to modern information retrieval*, 2nd ed, USA: McGraw-Hill, 1983.

[18] Salton, G. & Buckley, C, "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*. vol. 24, no. 5, pp. 513-523, 1988.

[19] Zhou, K., Ethayarajh, K., Card, D. & Jurafsky, D. "Problems with Cosine as a Measure of Embedding Similarity for High Frequency Words," in *Proc. The 60th Annual Meeting Of The Association For Computational Linguistics*, vol. 2, pp. 401-423, May 2022.

[20] Xuan, N. & Quang, H, "A new improved term weighting scheme for text categorization," *Knowledge And Systems Engineering*, pp. 261-270, 2014.

[21] Tjoa, E. & Guan, C, "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4793-4813, Nov. 2021.

[22] Madhu, C. & Sudhakar, M.S., "Adaptive Bezier Curve-based Membership Function formulation scheme for Interpretable Edge Detection," *Applied Soft Computing*. vol. 133, Jan. 2023, Art. no. 109968.

[23] Guillaume, S, "Designing fuzzy inference systems from data: An interpretability-oriented review," *IEEE Trans. Fuzzy Syst.*, vol. 9, no. 3, pp. 426-443, June 2001.

[24] Razak, T., Garibaldi, J., Wagner, C., Pourabdollah, A. & Soria, D., "Toward a Framework for Capturing Interpretability of Hierarchical Fuzzy Systems - A Participatory Design Approach," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 5, pp. 1160-1172, May 2021.

[25] Mikut, R., Jäkel, J. & Gröll, L., "Interpretability issues in data-based learning of fuzzy systems," *Fuzzy Sets And Systems*, vol. 150, pp. 179-197, Mar. 2005.

[26] Alcalá, R., Casillas, J., Cordón, O. & Herrera, F., "Building fuzzy graphs: Features and taxonomy of learning for non-grid-oriented fuzzy rule-based systems," *Journal Of Intelligent And Fuzzy Systems*. vol. 11, pp. 99-119, Jan. 2001.

[27] Mordeson, J., Mathew, S. & Gayathri, G. Fuzzy Graph Theory: Applications to Global Problems. (Springer Nature,2023)

[28] Madhu, C. & Sudhakar, M.S. An Interpretable Fuzzy Graph Learning for Label Propagation Assisting Data Classification. *IEEE Transactions On Fuzzy Systems*. **32**, 1331-1345 (2024)

[29] Hasni, S. & Faiz, S., "Word embeddings and deep learning for location prediction: tracking Coronavirus from British and American tweets," *Social Network Analysis And Mining*. Vol. 11, July 2021, Art. no. 66.

[30] Lui, M. & Cook, P. Classifying English documents by national dialect. *Proceedings Of The Australasian Language Technology Association Workshop 2013 (ALTA 2013)*. pp. 5-15 (2013)

[31] Dogan, T. & Uysal, A, "Improved inverse gravity moment term weighting for text classification," *Expert Systems With Applications*. vol. 130 pp. 45-59, Sept. 2019.

[32] Yu, Z., Ye, F., Yang, K., Cao, W., Chen, C., Cheng, L., You, J. & Wong, H., "Semisupervised Classification with Novel Graph Construction for High-Dimensional Data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 1, pp. 75-88, Jan. 2022.

[33] Wang, F., Zhu, L., Xie, L., Zhang, Z. & Zhong, M., "Label propagation with structured graph learning for semi-supervised dimension reduction," *Knowledge-Based Systems*. vol. 225, Aug. 2021, Art. no. 107130.

[34] Wang, L., Chan, R. & Zeng, T., "Probabilistic Semi-Supervised Learning via Sparse Graph Structure Learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 853-867, Feb. 2021.

[35] Fang, X., Xu, Y., Li, X., Lai, Z. & Wong, W., "Learning a nonnegative sparse graph for linear regression," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2760-2771, Sept. 2015.

[36] Carnevali, J., Rossi, R., Milios, E. & De Andrade Lopes, A., "A graph-based approach for positive and unlabeled learning," *Information Sciences*, vol. 580, pp. 655-672, Nov. 2021.

[37] Kumar, S., Kumar, N., Dev, A. & Naorem, S. Movie genre classification using binary relevance, label powerset, and machine learning classifiers. *Multimedia Tools And Applications*. **82**, 945-968 (2023)

[38] Bolívar, S., Nieto-Reyes, A. & Rogers, H. Statistical Depth for Text Data: An Application to the Classification of Healthcare Data. *Mathematics*. **11**, 1-20 (2023)

[39] Poczeta, K., Płaza, M., Michno, T., Krechowicz, M. & Zawadzki, M. A multi-label text message classification method designed for applications in call/contact centre systems. *Applied Soft Computing*. pp. 110562 (2023)

[40] Huang, Y., Dai, X., Yu, J. & Huang, Z. SA-SGRU: Combining Improved Self-Attention and Skip-GRU for Text Classification. *Applied Sciences*. **13**, 1296 (2023)

[41] Jaiswal, A. & Milios, E. Breaking the Token Barrier: Chunking and Convolution for Efficient Long Text Classification with BERT. *ArXiv Preprint ArXiv:2310.20558*. (2023)

[42] Abburi, H., Suesserman, M., Pudota, N., Veeramani, B., Bowen, E. & Bhattacharya, S. Generative ai text classification using ensemble llm approaches. *ArXiv Preprint ArXiv:2309.07755*. (2023)

[43] Saifullah, S., Dreżewski Rafałand Dwiyanto, F., Aribowo, A., Fauziah, Y. & Cahyana, N. Automated Text Annotation Using a Semi-Supervised Approach with Meta Vectorizer and Machine Learning Algorithms for Hate Speech Detection. *Applied Sciences*. **14**, 1078 (2024)

[44] Lotfy, A., "Fuzzy logic and the calculi of fuzzy rules, fuzzy graphs, and fuzzy probabilities," *Computers & Mathematics with Applications*, vol. 37, no. 11-12, pp. 35, June 1999.

[45] Etman, A. & Beex, A., "Language and Dialect Identification: A survey.," presented at *2015 SAI Intelligent Systems Conference (IntelliSys)*. pp. 220-231, 2015.

[46] Prabhu, K. *Window functions and their applications in signal processing*, CRC press, Sept. 2018.

[47] Joachims, T., "A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization," in *Proc.ICML '97* Fourteenth International Conference on Machine Learning, pp. 143–151, July 1997.

[48] Yüksel, A., Uügurlu, B. & Koç, A., "Semantic change detection with gaussian word embeddings," *IEEE/ACM Trans. Audio, Speech, And Language Processing*, vol. 29, pp. 3349-3361, 2021.

[49] Oxford Languages, (2023, October 17). "Semantic English Language Database," distributed by Oxford, https://languages.oup.com/products/semantic-english-language-database/

[50] Sohangir, S. & Wang, D., "Improved sqrt-cosine similarity measurement," *Journal Of Big Data*, vol. 4, July 2017.

[51] Singh, R. & Singh, S., "Text similarity measures in news articles by vector space model using NLP," *Journal of The Institution of Engineers (India): Series B*. vol. 102, pp. 329-338, Nov. 2021.

[52] Faisal Rahutomo, Teruaki Kitasuka & Masayoshi Aritsugi Semantic Cosine Similarity. *The 7th International Student Conference On Advanced Science And Technology ICAST*. **4** (2012)

[53] Sitara, M., Akram, M. & Yousaf Bhatti, M., "Fuzzy graph structures with application," *Mathematics*. vol. 7, no. 63, Jan. 2019.

[54] Jordan, M. & Weiss, Y., "On spectral clustering: Analysis and an algorithm," *Advances In Neural Information Processing Systems*. vol. 14 pp. 849-856, 2002.

[55] Luo, Y., Wong, Y., Kankanhalli, M. & Zhao, Q., "$\mathcal{G}$ -softmax: improving intraclass compactness and interclass separability of features," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 2, pp. 685-699, Feb. 2020.

[56] Wang, L., Zhu, J. & Zou, H., "The doubly regularized support vector machine," *Statistica Sinica*, vol. 16, no. 2, pp. 589-615, April 2006.

[57] Jayalakshmi, T. & Santhakumaran, A., "Statistical normalization and back propagation for classification," *International Journal Of Computer Theory And Engineering*. vol. 3, no. 1, pp. 1793-8201, Feb 2011.

[58] Shin, J. & Jeong, J., "Multiclass classification of hemodynamic responses for performance improvement of functional near-infrared spectroscopy-based brain–computer interface," *Journal Of Biomedical Optics*. vol. 19, no. 6, June 2014, Art. no. 67009.

[59] Nie, F., Shi, S. & Li, X., "Semi-supervised learning with auto-weighting feature and adaptive graph," *IEEE Trans. Knowledge And Data Engineering*, vol. 32, no. 6, pp. 1167-1178, 1 June 2020.

[60] English Corpora: most widely used online corpora. Billions of words of data. free online access. Available: https://www.english-corpora.org/corpora.asp

[61] Onan, A., "Hierarchical graph-based text classification framework with contextual node embedding and BERT-based dynamic fusion," *Journal Of King Saud University - Computer And Information Sciences*, vol. 35, no. 7, July 2023, Art. no. 101610.

[62] Wang, Y., Wang, C., Zhan, J., Ma, W. & Jiang, Y., "Text FCG: Fusing Contextual Information via Graph Learning for text classification," *Expert Systems With Applications*, vol. 219, June 2023, Art. no. 119658.

[63] Yang, Y., Miao, R., Wang, Y. & Wang, X., "Contrastive Graph Convolutional Networks with adaptive augmentation for text classification," *Information Processing & Management*. vol. 59, no. 4, July 2022, Art. no. 102946.

[64] Piao, Y., Lee, S., Lee, D. & Kim, S., "Sparse structure learning via graph neural networks for inductive document classification," in *Proc.The AAAI Conference On Artificial Intelligence*. vol. 36, pp. 11165-11173, 2022.

[65] Nguyen, T., Phung, D. et. al., "Quaternion graph neural networks," presented in *Asian Conference On Machine Learning*. pp. 236-251, 2021.

[66] Lin, Y., Meng, Y., Sun, X., Han, Q., Kuang, K., Li, J. & Wu, F., "Bertgcn: Transductive text classification by combining gcn and bert, *ArXiv Preprint ArXiv:2105.05727*, 2021.

[67] Xie, Q., Huang, J., Du, P., Peng, M. & Nie, J., "Inductive topic variational graph auto-encoder for text classification, in *Proc. The 2021 Conference Of The North American Chapter Of The Association For Computational Linguistics: Human Language Technologies*, pp. 4218-4227, 2021.

[68] Wang, Z., Wang, C., Zhang, H., Duan, Z., Zhou, M. & Chen, B., "Learning dynamic hierarchical topic graph with graph convolutional network for document classification," presented at *International Conference On Artificial Intelligence And Statistics*, pp. 3959-3969, 2020.

[69] Ding, K., Wang, J., Li, J., Li, D. & Liu, H., "Be more with less: Hypergraph attention networks for inductive text classification," *ArXiv Preprint ArXiv:2011.00387*, 2020.

[70] Zhang, Y., Yu, X., Cui, Z., Wu, S., Wen, Z. & Wang, L., "Every document owns its structure: Inductive text classification via graph neural networks," *ArXiv Preprint ArXiv:2004.13826*, 2020.

[71] Yao, L., Mao, C. & Luo, Y., "Graph convolutional networks for text classification," in *Proc. The AAAI Conference On Artificial Intelligence*, vol. 33, pp. 7370-7377, 2019.

Cherukula Madhu (Member, IEEE) has completed his B. Tech in Electronics and Communication Engineering from Siddarth Institute of Engineering and Technology, affiliated to JNTUA Ananthapur, Andhra Pradesh, India in 2009 and M. Tech in Communication Systems from Sri Venkateswara University, Tirupati, Andhra Pradesh, India in 2011.

Currently, he is a Research Scholar with School of Electronics Engineering, Vellore Institute of Technology, Vellore, Tamilnadu, India. His research interests include data analytics and interpretable machine learning.

Sudhakar MS received the PhD degree in information and communication engineering from Madras Institute of Technology, Anna University, India, in 2014. He is currently serving the School of Electronics Engineering, Vellore Institute of Technology, India as an Associate professor with. His research interests include machine learning and its applications, such as computer vision, data mining, image processing, information retrieval, and pattern recognition. He has several years of industrial and R&D experience in tier-1 organizations and has published sufficient number of research articles in Science Citation Index Journals. He is currently serving as the PI for a research project with Indian Space Research Organization (ISRO).