DB²Net: A Deep Learning Approach for Predicting Levels of Interest for Articles Posted on Social Forums

*Guan-Wei Chen, Cheng-Chin Chiang**

Department of Computer Science and Information Engineering,
National Dong Hwa University,
Taiwan
610921249@gms.ndhu.edu.tw, ccchiang@gms.ndhu.edu.tw

Abstract

Online social forums are crucial for communication and information sharing in today's digital era. This paper proposes the DB²Net (Dual BERT Decomposed Bilinear Layer Network), a deep learning model designed to forecast the levels of interest (LOIs) for articles posted on Taiwan's PTT social forum. Using dual BERT modules alongside a newly devised Decomposed Bilinear Layer (DB Layer), DB²Net can explore second-order inter-feature correlations within the textual features of articles. It effectively achieves a prediction accuracy of 98.54% for articles posted within a short-term span of one week, outperforming other BERT models and traditional machine learning models, including XGBoost and Decision Trees. The paper also compares BERT's performance with that of Bidirectional LSTM, further substantiating the efficacy of using BERT in LOI prediction.

Keywords: BERT, Bilinear layer, Social forum, Level of interest, Second-order correlation

1 Introduction

In the current digital era, characterized by the ubiquitous presence of smartphones and unlimited internet access, online forums have emerged as the main platforms for communication, knowledge sharing, and community building [1-2]. Online social forums fulfill our inherent need for social engagement and interaction. Consequently, articles in online social forums significantly influence readers' perspectives and inform their decisions. While some articles with emotional nuances can illuminate issues and guide readers to well-informed decisions, others may impair judgment and obstruct impartial reasoning. When articles resonate and become popular, they often contribute to extensive dialogues, intensifying emotional reactions and sometimes leading to unexpected outcomes. On the positive side, compelling articles can unify communities and promote shared goals. On the negative side, however, they can also provoke contentious discussions, incite discord, or lead to legal disputes, thereby disrupting social harmony. Therefore, the level of engagement with forum articles—reflected in their Levels of Interest (LOIs)—is intricately linked to these dynamics of public sentiment.

Forecasting the LOIs for articles on social forums is crucial due to its wide-ranging applications. For both public and private sectors entities, they often rely on forums to engage with their audiences and stimulate dialogues. Thus, forecasting the LOIs of their promotional and informational posts helps in assessing the impact. For forum administrators, who are responsible for careful content monitoring, gauging public reaction, and ensuring regulatory compliance, predicting the LOIs of posts can improve their regulatory efforts. The levels of engagement with forum articles usually depend on both their title and content. This study aims to develop novel models for predicting LOIs by analyzing these elements of articles posted on the Gossiping Board of Taiwan's PTT forum, a key player in shaping public opinion on various social issues since its inception at National Taiwan University in 1995.

The advent of deep learning has been transformative in the field of natural language processing (NLP). Within the specific application of LOI prediction for social forum articles, deep learning offers several substantial benefits:

1. Automated Feature Extraction: Deep learning architectures can automatically identify important textual features of article content, essential for gauging the level of discourse engagement.
2. Handling Unstructured Content: Deep learning techniques excel at interpreting unstructured social forum content, enabling them to draw out critical insights and themes.
3. Contextual Integration: Deep learning frameworks can capture contextual dependencies in articles, thereby improving the accuracy of LOI prediction.
4. Iterative Refinement: Deep learning models can adaptively improve their predicting performance by incorporating learning from more new articles.

Deep learning models excel in automatic feature learning, yet they often rely on linear layers that compute linear combinations of features from previous layers. However, these linear combinations do not account for complex, higher-order feature correlations that are crucial in addressing many classification problems. Consider the XOR problem of two Boolean variables, say x and y, as an example. This function represents a typical binary classification problem that is

*Corresponding Author: Cheng-Chin Chiang; E-mail: ccchiang@gms.ndhu.edu.tw
DOI: https://doi.org/10.70003/160792642024112506003

linearly non-separable and cannot be successfully learned by one single linear layer [3]. However, by introducing an additional second-order correlation term of xy into the input, this problem becomes linearly separable, enabling a single linear layer to learn it easily [4]. Yet, due to the many possible combinations among all input variables, a model would require many parameters to explore the effective high-order inter-feature correlations, leading to increased computational costs and a higher risk of overfitting. The aim of this paper is to propose a Decomposed Bilinear Layer (DB Layer) that can proficiently learn these second-order inter-feature correlations without resorting to an excessively large parameter set and a DB^2Net (Dual BERT Decomposed Bilinear Net) model that incorporates the DB Layer to enhance the prediction of LOIs for articles in social forums.

This paper introduces innovative techniques and contributions, primarily encompassing:

1. The design of a DB^2Net which incorporates a dual-stream feature extraction approach using BERT networks and integrates a newly devised DB Layer to automatically explore second-order inter-feature correlations for improving the LOI prediction.
2. A comparative study of the performance of LOI prediction using deep learning models versus traditional machine learning models.
3. A novel paradigm for predicting LOIs for forum articles that proposes a systematic and objective LOI grading scheme and pioneers a study on LOI prediction across time spans of different lengths.

The rest of this paper is organized as follows. Section 2 briefly surveys related works in the field. Section 3 delves into the proposed methodology, beginning with the mathematical rationale for the DB Layer, followed by an analysis of its computational complexity. Subsequently, the architecture of the DB^2Net model is detailed. Section 4 discloses the experimental findings, including specifics of the experimental setup, dataset, evaluation metrics, and comparative assessments. Additionally, this section deliberates on pivotal insights derived from the experiments, the impact of temporal spans of different lengths on LOI prediction, and the efficacy of BERT in LOI prediction. Finally, we conclude this paper with some remarks in Section 5.

2 Related Work

Predicting the LOIs for articles in social forums is a specific application of text classification. Other applications of text classification include sentiment analysis [5-7], topic labeling [8], customer support [9], news aggregation [10], social media monitoring [11], and so on. According to the literature [12-16], this task involves extracting key textual features that carry semantic significance and contextual information. To allow computational models to analyze such data, the textual data must be transformed into numerical vectors. Some traditional methods for this task include the Bag-of-Words (BoW) [17], Term Frequency-Inverse Document Frequency (TF-IDF) [18], N-Grams [19], and so on. For instance, using TF-IDF, Kathy Lee et al. [20] employed a C5.0 decision tree [21] to categorize Twitter articles into 18 groups using their titles, achieving an accuracy of 70.96% on 768 randomly chosen articles. Some studies also exploit other metadata associated with the textual data for text classification. Sejal Bhatia [22] tested several traditional machine learning models on 39,797 news articles to determine their popularity (popular and non-popular) using metadata attributes such as number of words, publication day, article polarity, and number of internal links of an article. Among the tested models, the XGBoost [23] was found to be the most accurate, with accuracies between 70% and 75%. An important limitation of these traditional models is that once the feature vectors are transformed, they cannot be readjusted to better resonate with the learning of the classification model.

The emergence of deep learning has introduced methods that more effectively convert textual data into numerical feature vectors, aligning better with the needs of classification models. For instance, embeddings from the language model (ELMo) [24] use a deep, bidirectional LSTM [25] model to derive word representations that fully consider the context in which a word is used. As a result, the same word can yield different embeddings based on its contextual use, thereby enhancing classification accuracy. Similarly, BERT [26] represents a significant advancement in deep learning for generating contextual word embeddings by integrating both preceding and following context at every layer of the network. Both the ELMo and BERT facilitate the learning of representations through end-to-end learning process, allowing for the adaptive extraction of effective numerical feature vectors from textual inputs.

Existing research on text classification using deep learning often employs BERT, LSTM, or their derivatives. Atul Anand [27] conducted a study on the efficacy of BERT for automated movie rating using the IMDB movie review database. This approach, however, does not capture inter-feature interactions, which could compromise accuracy. Hongxia Wei et al. [28] investigated the news-comments relevance analysis, which is framed as a classification problem based on BERT. They assessed the correlation between news narratives and user comments on several news platforms, categorizing relevance into four levels. Their comparative study with models like Doc2Vec [29], Siamese LSTM [30], BERT, and BERTRNC found that BERTRNC achieved the highest accuracy of 91.5%, highlighting BERT's effectiveness. In Taiwan, Lin et al. [31] developed a sentiment classifier for local indigenous texts employing an LSTM model trained on dialog transcripts with elderly seniors. A notable limitation of LSTM is its diminishing efficiency with longer data sequences. Lan et al. [32] developed a 'Taipei City QA Chatbot' using an ALBERT model to direct citizens' inquiries to their appropriate government agencies. However, the imbalanced volume of labeled data for different agency departments posed a risk of overfitting. Wang's investigation [33] into the PTT HatePolitics Board using an DFTC [34] model to predict their future popularity. Articles with 0~23, 24~174, and over 175 comments were labeled as low, moderate, and high popularity, respectively. The data volumes for these three levels are 73.37%, 25.56%, and 1.07%, respectively. After trained with the data spanning over six months, their model achieves the accuracies 95.7%, 89.9%, and 91.6% for the three popularity levels, respectively. However, due to the rapid evolution of forum topics, this extended data collection period may not be appropriate for accurate popularity prediction.

3 Problem Solution

3.1 Decomposed Bilinear Layer

Text classification by deep learning models typically involves tokenizing and encoding textual features, which are then processed by a multilayer perceptron (MLP) [35] for classifying the textual features. The MLP utilizes one or several fully-connected layers (FCL) [36] to perform the linear combination of the encoded features. These features are then transformed by a non-linear activation function. Subsequently, the softmax function then converts the activation outputs into probabilities of different categories for categorizing the text. However, this approach overlooks the high-order inter-feature correlations between features. Incorporating a mechanism that captures the intricate higher-order inter-feature correlations could significantly improve the model's classification performance.

The Bilinear Layer (BL) [37] is an enhanced neural network layer designed to compute the correlation terms between two feature vectors. Given two vectors $\mathbf{x} = [x_1 \ x_2 \ ... \ x_m]^T \in \mathbb{R}^m$ and $\mathbf{y} = [y_1 \ y_2 \ ... \ y_n]^T \in \mathbb{R}^n$, each neuron in a BL calculates its output using a learnable parameter matrix $W_k \in \mathbb{R}^{m \times n}$ by

$$z_k = \mathbf{x}^{\mathbf{T}} W_k \mathbf{y}, \ k = 1, \dots, K \ , \quad \textbf{(1)}$$

where K denotes the number of neurons in the BL. In Eq. (1), each element from $\mathbf{x}$ and $\mathbf{y}$ undergoes pairwise multiplication, yielding the second-order correlation terms $\{x_i y_j | 1 \leq i \leq m, \ 1 \leq j \leq n\}$. Each correlation term is multiplied with the corresponding weight $w_{i,j}$ in W_k to yield z_k. Such a mechanism underscores the BL's capability to explore the intricate second-order correlations between elements of $\mathbf{x}$ and $\mathbf{y}$. However, a notable concern about the BL is its intensive parameters and computational demand. The dimensionality of the weight matrix W_k is $m \times n$, leading to a total parameter count of $K \times m \times n$ for a BL with K neurons. In situations where input feature vectors are long, the abundance of parameters risks overfitting and excessive computational load.

To circumvent the inherent issues of the BL, we propose the Decomposed Bilinear Layer (DB Layer). Suppose that $\mathbf{x}$ and $\mathbf{y}$ in Eq. (1) have the same dimensionality of n. If their dimensionalities are not the same, we can transform them into two vectors of the same dimensionality n using linear layers. We can compose the weight matrix W_k by multiplying two weight vectors, $\mathbf{w}_{k,1} \in \mathbb{R}^{n \times 1}$ and $\mathbf{w}_{k,2} \in \mathbb{R}^{n \times 1}$, expressed as:

$$W_k = \mathbf{w}_{k,1} \mathbf{w}_{k,2}^T$$

Accordingly, z_k in Eq. (1) becomes:

$$\begin{aligned} z_k &= \mathbf{x}^T \mathbf{w}_{k,1} \mathbf{w}_{k,2}^T \mathbf{y}, \ k = 1, 2, ..., K \\ &= (\mathbf{w}_{k,1}^T \mathbf{x})(\mathbf{w}_{k,2}^T \mathbf{y}) \\ &= v_{k,1} v_{k,2.} \end{aligned} \quad \textbf{(2)}$$

From Eq. (2), we can see that the BL's output can be equivalently computed as the product of two separate linear transformations of $\mathbf{x}$ and $\mathbf{y}$, i.e., $v_{k,1}$ $(= \mathbf{w}_{k,1}^T \mathbf{x})$ and $v_{k,2}$ $(= \mathbf{w}_{k,2}^T \mathbf{x})$, by decomposing W_k into $\mathbf{w}_{k,1}$ and $\mathbf{w}_{k,2}$. When expanded, the product $v_{k,1} v_{k,2}$ yields

$$v_{k,1} v_{k,2} = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{k,1,i} w_{k,2,j} x_i y_j ,$$

which also reveals the second-order correlation $x_i y_j$ between x_i and y_j, showcasing the potential to learn the significance of their correlations through the product of $w_{k,1,i}$ and $w_{k,2,j}$. For a DB Layer with K outputs, we can streamline the computation by multiplying the corresponding output elements of two linear layers with K neurons, i.e.,

$$\begin{aligned} \mathbf{z} &= \left[z_1, z_2, ..., z_K \right]^T \\ &= [v_{1,1}, v_{1,2}, v_{2,1}, v_{2,2}, ..., v_{K,1} v_{K,2}]^T \\ &= [v_{1,1}, v_{2,1}, ..., v_{K,1}]^T \odot [v_{1,2}, v_{2,2}, ..., v_{K,2}]^T \\ &= \mathbf{v}_1 \odot \mathbf{v}_2 = (W_1^T \mathbf{x}) \odot (W_2^T \mathbf{y}) \in \mathbb{R}^K \end{aligned} \quad \textbf{(3)}$$

where vectors $\mathbf{v}_1 = [v_{1,1}, \ v_{2,1}, \ \dots, \ v_{K,1}]^T$ and $\mathbf{v}_2 = [v_{1,2}, \ v_{2,2}, \ \dots, \ v_{K,2}]^T$ are obtained by the element-wise multiplication of the outputs from two K-output linear layers with weight matrices $W_1 \in \mathbb{R}^{n \times K}$ and $W_2 \in \mathbb{R}^{n \times K}$, respectively. The symbol $\odot$ represents the Hadamard product operation, which multiplies the corresponding elements of two vectors.

Upon comparing the DB Layer with the BL, both configured to produce K outputs, the DB Layer requires $2 \times K \times n$ weight parameters, while the BL requires $K \times n^2$ weight parameters. This great reduction in weight parameters makes the DB Layer notably less prone to overfitting, especially when the available training data is sparse. In terms of computational complexity, the DB Layer requires $(2n + 1)K$ multiplication operations, whereas the BL demands $(n^2 + n)K$ multiplication operations. This indicates a significantly lower computational expense for the DB Layer compared to the BL. A comparative analysis of the linear layer, BL, and DB Layer is presented in Table 1. To regulate the output values post-Hadamard product in the DB Layer and ensure they remain within a normalized range of -1 and +1, the outputs are passed through a hyperbolic tangent function (tanh()). The architecture of a DB Layer is illustrated in Figure 1.

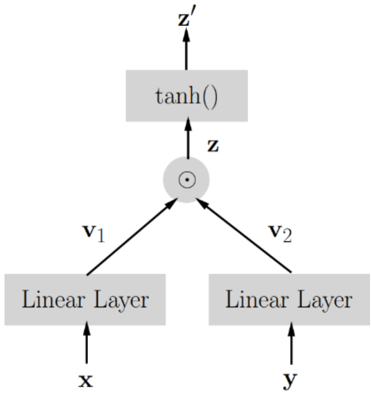

Figure 1. The Decomposed Bilinear Layer

3.2 DB²Net

The DB²Net is specifically designed for predicting the Levels of Interest (LOIs) in forum articles, capitalizing on the effectiveness of the newly introduced DB Layer. As illustrated in Figure 2, the architecture of DB²Net begins with encoding textual inputs using the BERT tokenizer. These inputs are then processed through a pair of pre-trained BERT models, each producing a 768-dimensional feature vector to capture textual features. The core of the network, the DB Layer, takes these two feature vectors and explores second-order inter-feature correlations to enhance prediction accuracy. To reduce the risk of overfitting, a dropout layer follows the DB Layer. The final step involves the use of an MLP, which employs one or more fully connected layers to transform the features into categorical probabilities. These probabilities correspond to four predefined levels of interest, completing the LOI prediction process for forum articles.

4 Experimental Results

4.1 Environment and Dataset

Table 2 details the software and hardware used in the experiments. For our experiments, the 'bert-base-chinese' model from Hugging Face, which includes both a tokenizer and a pre-trained model suitable for processing Chinese text, was employed (https://huggingface.co/bert-base-chinese). A web crawler was developed specifically to harvest data from the PTT forum's Gossiping Board. Data collected for each article included include the title, posting time, content, comments, upvotes, downvotes, and article URL. Given the typically ephemeral nature of discussions on social forums, our data collection spanned a period from November 2nd to December 6th, 2022, yielding a dataset of 48,648 articles. The jieba package (available on https://github.com/fxsjy/jieba) was utilized to perform Chinese word segmentation, a crucial step for identifying prominent keywords within the articles. For the purpose of benchmarking the effectiveness of deep learning approaches against conventional machine learning techniques, the Scikit-learn library for Python (https://scikit-learn.org/stable/index.html) serves as the primary tool for implementing the conventional machine learning models.

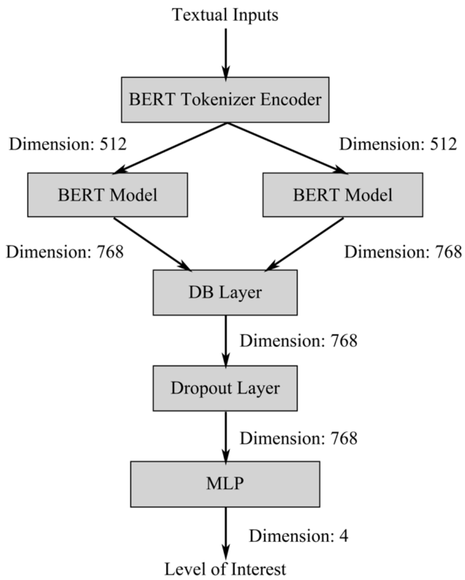

Figure 2. The proposed Dual BERT Decomposed Bilinear Net (DB²Net)

The LOI for each article in the dataset is determined based on the aggregate of its comments, upvotes, and downvotes. Articles are ranked by these metrics and divided into quartiles that represent different interest levels: the top 0%-25% are labeled as 'Highly Interested', 25%-50% as 'Interested', 50%-75% as 'Lowly Interested', and the bottom 75%-100% as 'Non-Interested'. This categorization method allows for an automated labeling process, significantly reducing the need for intensive labor. Moreover, by ensuring an even distribution of articles across these four categories, the issue of model overfitting can be mitigated.

Table 2. The environment configurations for experiments

Hardware	CPU	Intel i7-10700F
	CPU RAM	32 GB
	GPU	NVIDIA GeForce RTX 3070
	GPU RAM	16 GB
Software	OS	Windows
	DL Framework	PyTorch
	jieba	0.42.1
	Scikit-learn	1.0.2

Table 1. Comparison of Linear Layer, Dual Linear Layer and DB Layer

Properties	Linear Layer	Bilinear Layer	DB Layer
Order of feature correlation	1st-order	2nd-order	2nd-order
Ouput computation	$z_k = w^T\mathbf{x}$	$z_k = \mathbf{x}^T W \mathbf{y}$	$z_k = (\mathbf{w}_1^T\mathbf{x})(\mathbf{w}_2^T\mathbf{y})$
Number of weights for K output neurons (give that input dimension = n)	$K \times n$	$K \times n^2$	$2 \times K \times n$
Computation complexity for K output neurons	$K \times n$	$(n^2+n)K$	$(2n+1)K$

To validate the effectiveness of this LOI labeling approach, the jieba package was used for word segmentation on the collected articles to identify common keywords within each interest level. The eight most frequent keywords from articles in each level are depicted in word clouds, as shown in Figure 3. These keywords were further analyzed using Google Trends to evaluate their global search popularity during the week of November 20th to 26th, 2022. The results, presented in a bar chart in Figure 3, demonstrate a positive correlation between the search popularity of these keywords and their respective LOI levels, thus substantiating the credibility of our labeling scheme for LOIs.

Metrics like comments, upvotes, and downvotes are indicative of an article's interaction level and public engagement. When correlated with their publication dates, these metrics present a dynamic picture of the article's LOI, reflecting the transient nature of interest over time. To address this temporal variation, three distinct datasets have been curated: DS_1w, which includes articles within one week for highlighting immediate engagement; DS_2w, encompassing a two-week period to capture more extended interaction trends; and DS_1m, for a full month for observing long-lasting interest. The specifics of these datasets are outlined in Table 3. Each dataset is partitioned into training (70%), validation (10%), and testing (20%) subsets.

4.2 Performance Evaluation and Comparison

4.2.1 Learning Curves

The DB^{2}Net architecture is distinctively characterized by its ability to identify second-order inter-feature correlations from features extracted by two streams of BERT backbones. This capability prompts an investigation into whether the DB^{2}Net can expedite the learning process more effectively than a single BERT model. For comparative analysis, two foundational BERT models are established:

1. Base BERT model I: This configuration utilizes a single BERT architecture as its core, incorporating a dropout layer with a dropout rate of 0.2 for feature extraction. Following this, it employs a straightforward multilayer perceptron (MLP) without any hidden layers to conduct the prediction task.
2. Base BERT model II: Similar to the first, this model employs a single BERT backbone and a dropout layer with an identical dropout rate. However, it extends its structure with an MLP that includes a hidden layer composed of 768 neurons, aiming to enhance prediction capabilities.

By contrasting DB^{2}Net's performance with those of these two base models, particularly through the analysis of their respective learning curves, the effectiveness and accelerated learning proficiency attributed to the DB Layer can be assessed.

Figure 4 presents the learning curves for the dataset DS_1w, specifically for the time period from November 2nd to November 8th. The depicted learning curves illustrate a notable observation: Base BERT Model I and Base BERT Model II exhibit significant fluctuations in their loss reduction trajectory. Moreover, there is an apparent trend of increasing loss on the validation dataset over time for these models. In contrast, DB^{2}Net demonstrates a more rapid convergence compared to the two base models. Beyond mere speed, the convergence pattern of DB^{2}Net is markedly more stable, lacking the pronounced fluctuations seen in the two base BERT models. This indicates that DB^{2}Net not only learns faster but also with greater consistency, highlighting its superior learning dynamics.

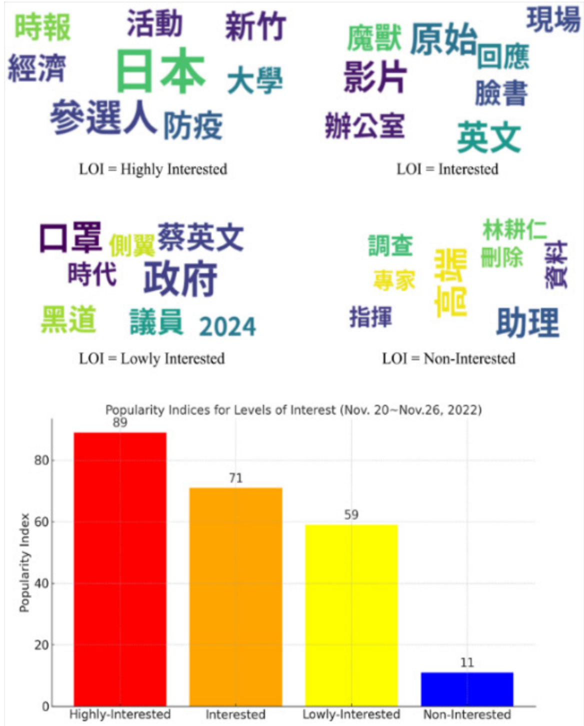

Figure 3. On Nov. 20-26, 2022, Google Trends measured the popularity of the top eight keywords from articles across four interest levels (Popularity scores range from 0-100.)

Table 3. The three types of datasets corresponding to time spans of 1 week, half a month, and a month

DS_1w		DS_2w		DS_1m	
Duration	Set size	Duration	Set size	Duration	Set size
Nov02-Nov08	9054	Nov02-Nov15	17966	Nov02-Dec06	48648
Nov09-Nov15	8912	Nov09-Nov22	18122		
Nov16-Nov22	9210	Nov16-Nov29	21741		
Nov23-Nov29	12531	Nov23-Dec06	21472		
Nov30-Dec06	8941				

4.2.2 Comparative Studies on Prediction Accuracies

In addition to evaluating the learning dynamics in terms of speed and stability, the efficacy of the features learned by the models is assessed through their prediction accuracy. Prediction accuracy is quantified as the ratio of correct predictions to the total number of test samples, multiplied by 100 to yield a percentage, i.e., Prediction Accuracy = $\left(\frac{\text{\#of correct predictions}}{\text{\#of test samples}}\right)\times 100\%$.

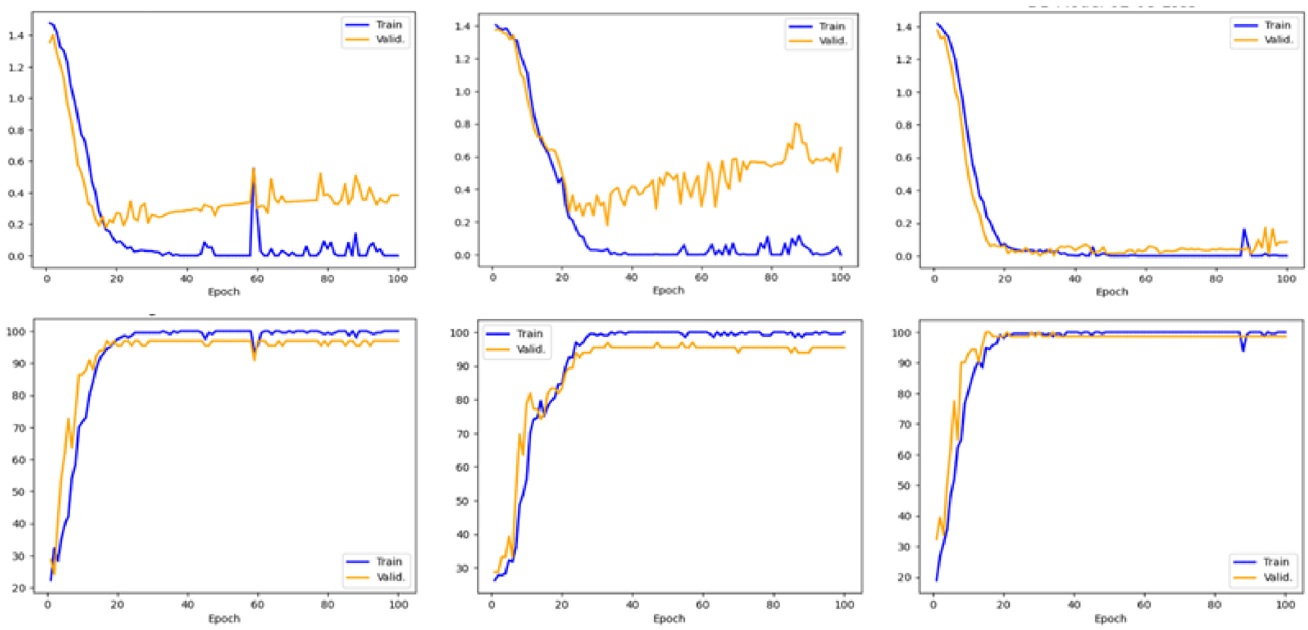

Figure 4. The learning curves of the three compared deep learning models
(Column 1, Column 2, and Column 3 are curves of Base BERT model I, Base BERT model II, and DB^{2}Net, respectively. The top row shows the curves of loss and the bottom row shows the accuracy.)

Table 4 consolidates the prediction accuracies for the DB^{2}Net and the two foundational BERT base models. These models are evaluated against the three datasets—DS_1w, DS_2w, and DS_1m. The results reveal that DB^{2}Net consistently achieves higher accuracy across all datasets when compared to the two base BERT models. This empirical evidence underscores the advanced capability of DB^{2}Net to extract more effective features for predicting the LOIs of forum articles and confirms DB^{2}Net's superior performance compared to the BERT's base models.

Additionally, to compare the prediction accuracies of deep learning models and traditional machine learning models, we implement the XGBoost and the Decision Tree using the Scikit-learn library. Both models use features extracted by our DB Layer. The accuracies of each model on the datasets are also presented in Table 4. As shown in this table, all BERT-based deep learning models achieve higher prediction accuracies than the two traditional machine learning models. The results emphasize the advanced learning and generalization capabilities of deep learning models over their traditional counterparts in this specific text classification task.

Table 4. Comparison of prediction accuracies achieved by five compared models on datasets of different time spans

Dataset		Base BERT model I		Base BERT model II		XGBoost		Decision Tree		DB^{2}Net	
		Valid. (%)	Test. (%)	Valid. (%)	Test. (%)	Valid. (%)	Test. (%)	Valid. (%)	Test. (%)	Valid. (%)	Test. (%)
DS_1w	Nov02-Nov08	96.97	95.38	95.45	96.92	77.77	76.92	81.48	80.77	**98.59**	**98.57**
	Nov09- Nov15	95.38	95.31	95.38	96.87	81.48	80.77	74.07	73.07	**98.57**	**98.55**
	Nov16-Nov22	98.51	96.97	95.52	96.97	75.01	74.07	71.43	70.37	**98.61**	**98.59**
	Nov23-Nov29	95.38	96.34	96.97	96.34	82.86	82.35	77.14	76.47	**98.87**	**98.45**
	Nov30-Dec06	95.38	98.43	97.59	98.43	81.48	80.77	85.19	84.62	**98.87**	**98.55**
Average		96.32	96.47	96.18	97.11	79.72	78.98	77.86	77.06	**98.70**	**98.54**
DS_2w	Nov02- Nov16	77.21	76.36	78.24	76.34	65.46	64.28	69.09	67.85	**78.57**	**80.35**
	Nov09- Nov22	76.78	76.35	76.77	76.35	65.45	64.28	63.64	62.50	**80.35**	**75.01**
	Nov16- Nov29	73.43	73.01	72.50	71.90	63.50	63.49	64.06	63.49	**76.56**	**74.60**
	Nov23- Nov06	73.78	72.03	71.29	70.31	57.14	57.06	55.56	55.45	**74.60**	**75.81**
Average		75.30	74.44	74.7	73.73	62.88	62.28	63.09	62.32	**77.52**	**76.44**
DS_1m	Nov02-Dec06	67.28	60.25	67.13	60.11	59.71	58.64	57.83	57.04	**67.44**	**61.42**

4.2.3 Discussions

A. Interpretations and Findings on the Results

Upon examining the learning curves of the two base BERT models and the DB^2Net depicted in Figure 4, the lower stability and the slower speed of learning could be attributed to the linear layer used in the MLP, which potentially constrains the ability to assimilate higher-order inter-feature correlations. Conversely, DB^2Net demonstrates proficiency in harnessing second-order inter-feature correlations which make the learning process easier and smoother.

Table 4 illustrates that the prediction accuracies of the three deep learning models surpass those of the two traditional machine learning models, XGBoost and Decision Tree. Although the two models exploit the features distilled by DB^2Net, these features are fixed and cannot dynamically adapt to resonate with the parameter estimation of the two classification models. This exposes a gap between the acquisition of effective features and the parameter estimation of classification models which can limit the further enhancement in prediction accuracy. In contrast, the deep learning models are capable of simultaneously refining features and predictions through comprehensive end-to-end training, leading to superior prediction accuracy in the experiments.

According to our labeling scheme of LOI, articles of adjacent levels often show some degree of overlap in article topics. Our observations on inaccurately predicted test cases reveals that a majority of these misclassifications involve articles with similar topics but assigned to different LOIs. In addition, some cases are related to sudden changes in the social dynamics due to particular events. For example, during the 2022 FIFA World Cup, the rivalry between Japan and South Korea garnered significant public attention before December 1. Subsequently, Japan's elimination on December 1 led to a precipitous decline in attention. Given that our model does not account for such abrupt transitions, the misclassifications are understandable. A potential solution is to shorten the time span of training data, such as moving from weekly to daily or even hourly intervals, allowing the model to better assimilate rapid temporal variations.

B. Impact of Time Spans

Extending the duration from one to two weeks, we noted in Table 4 a substantial decline in accuracy. The DS_2w dataset's average prediction accuracy fell to 76.44%, a 22.1% decrease from the DS_1w dataset's 98.54% accuracy. When the duration expanded to a month, accuracy further reduces to 61.42%. This stands to reason since an article's LOI can vary over time. As previously discussed, particular events may prompt abrupt change in social dynamics. With elongated durations, the likelihood of such occurrences also increases, particularly for most articles typically discussing ephemeral topics. These abrupt change in social dynamics often occur at the end of articles' life intervals, thereby diminishing the prediction accuracy.

Another factor that impedes the prediction accuracy of LOI over longer time spans is that longer time spans would gather more topics in articles, thereby introducing more sophisticated temporal dependencies. Without increasing the data, the network model can fail to learn these temporal dependencies completely. Therefore, to avoid the gradual decrease in prediction accuracy, collecting a large amount of data is necessary. However, the dataset size is limited by the total number of posted articles on the forum board. To address this problem, devising a way to augment the dataset would be crucial.

C. Efficacy of BERT

To examine how effectively BERT extracts key textual features, we create a DBLSTM network by replacing each BERT in DB^2Net with a single-layer bidirectional LSTM of 768 hidden neurons and compare its performance with that of DB^2Net. Table 5 shows the prediction accuracies of DBLSTM and DB^2Net. The results demonstrate that DB^2Net outperforms DBLSTM on all tested datasets, verifying that BERT has a superior ability to characterize textual features for text classification.

Table 5. Comparison of prediction accuracies of DBLSTM and DB^2Net

		DBLSTM	DB^2Net
		Test. (%)	Test. (%)
DS_1w	Nov02-Nov08	73.08	**98.57**
	Nov09- Nov15	88.89	**98.55**
	Nov16-Nov22	77.78	**98.59**
	Nov23-Nov29	73.53	**98.45**
	Nov30-Dec06	76.92	**98.55**
Average		78.04	**98.54**
DS_2w	Nov02- Nov16	67.27	**80.35**
	Nov09- Nov22	70.91	**75.01**
	Nov16- Nov29	70.36	**74.60**
	Nov23- Nov06	75.80	**75.81**
Average		71.09	**76.44**
DS_1m	Nov02-Dec06	57.27	**61.42**

Another important observation from the results is the drop in prediction accuracy when the time span is lengthened from one week to one month. The DB^2Net experiences a higher drop of 37% (degraded from 98% to 61%), while the DBLSTM reveals a lower drop of 16% (degraded from 73% to 57%). Although BERT is better at capturing long-term contextual dependencies, it requires more parameters than LSTM. Thus, fine-tuning DB^2Net would require more training samples than does the DBLSTM. As mentioned previously, due to the limited number of articles available from the PTT forum in one month, the DS_1m dataset may not be sufficient for effectively fine-tuning our DB^2Net. Therefore, with the same amount of training data, DB^2Net is more prone to overfitting and experience a larger drop in prediction accuracy.

5 Concluding Remarks

In this paper, we design a deep learning network model DB^2Net as the core technology to predict the LOI of forum articles. We verify the performance of our model using the well-known PTT forum in Taiwan as a case study. From this study, we draw the following conclusions:

1. We propose an LOI labeling scheme, which not only dramatically reduces the time and effort of manual data labeling but also avoids the issue of data imbalance and the subjectivity that may be introduced by human perception.
2. By exploiting the dual BERT modules to extract textual features from article texts, the proposed DB^2Net demonstrates its superior ability to characterize long-term temporal dependencies compared to the bidirectional LSTM.
3. We propose the DB Layer to explore the second-order correlations between elements of two feature vectors. Moreover, the DB Layer significantly reduces the number of learnable parameters and computation costs compared to the BL.
4. With the proposed DB Layer, the DB^2Net effectively improves the prediction accuracy over other base BERT base models.
5. Through a comparison of experimental results, the deep learning models outperform the traditional machine learning models, including XGBoost and Decision Trees.
6. We also verify the superior ability of BERT in characterizing long-term contextual dependencies compared to Bidirectional LSTM.

These outcomes not only exhibit the robust capabilities of DB^2Net but also underscore the potential of advanced neural network architectures in improving the analytical prowess of text classification models.

References

[1] P. S. Fleming, B. Ludwig, Orthodontic groups on social media: Creating and maintaining an educational forum, *Seminars in Orthodontics*, Vol. 29, No. 4, pp. 346–352, December, 2023. doi: 10.1053/j.sodo.2023.07.001

[2] D. A. Rivera, M. Frenay, V. Swaen, The Learning Design of MOOC Discussion Forums: An Analysis of Forum Instructions and Their Role in Supporting the Social Construction of Knowledge, *Technology, Knowledge and Learning*, Vol. 29, No. 2, pp. 585–619, June, 2024. doi: 10.1007/s10758-023-09670-w

[3] M. Minsky, S. A. Papert, *Perceptrons: An Introduction to Computational Geometry*. The MIT Press, 2017. doi: 10.7551/mitpress/11301.001.0001

[4] S. Stefanovic, PyTorch - Solving the famous XOR problem using Linear classifiers with PyTorch, Master Data Science. Accessed: 2023. [Online]. Available: https://datahacker.rs/006-solving-the-xor-problem-using-neural-networks-with-pytorch/

[5] S. S. Roy, A. Roy, P. Samui, M. Gandomi, A. H. Gandomi, Hateful Sentiment Detection in Real-Time Tweets: An LSTM-Based Comparative Approach, *IEEE Transactions on Computational Social Systems*, Vol. 11, No. 4, pp. 5028–5037, August, 2024. doi: 10.1109/TCSS.2023.3260217

[6] D. Liu, X. L. Feng, F. Ahmed, M. Shahid, J. Guo, Detecting and Measuring Depression on Social Media Using a Machine Learning Approach: Systematic Review, *JMIR Mental Health*, Vol. 9, No. 3, Article No. e27244, March, 2022. doi: 10.2196/27244

[7] A. Ahmed , S. Aziz, C. T. Toro, M. Alzubaidi, S. Irshaidat, H. A. Serhan, A. A. Abd-Alrazaq, M. Househ, Machine learning models to detect anxiety and depression through social media: A scoping review, *Computer Methods and Programs in Biomedicine Update*, Vol. 2, Article No. 100066, 2022. doi: 10.1016/j.cmpbup.2022.100066

[8] K. N. Lam, L. T. Truong, J. Kalita, Using Topic Models to Label Documents for Classification, *Proceedings of Future Data and Security Engineering. Big Data, Security and Privacy, Smart City and Industry 4.0 Applications (FDSE): 7th International Conference*, Quy Nhon, Vietnam, 2020, p. 443–451. doi: 10.1007/978-981-33-4370-2_32

[9] P. S. Parmar, P. K. Biju, M. Shankar, N. Kadiresan, Multiclass Text Classification and Analytics for Improving Customer Support Response through different Classifiers, *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Bangalore, India, 2018, pp. 538–542. doi: 10.1109/ICACCI.2018.8554881

[10] C. Li, G. Zhan, Z. Li, News Text Classification Based on Improved Bi-LSTM-CNN, *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*, Hangzhou, China, 2018, pp. 890–893. doi: 10.1109/ITME.2018.00199

[11] P. Hajibabaee, M. Malekzadeh, M. Ahmadi, M. Heidari, A. Esmaeilzadeh, R. Abdolazimi, Offensive Language Detection on Social Media Based on Text Classification, *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, NV, USA, 2022, pp. 0092–0098. doi: 10.1109/CCWC54503.2022.9720804

[12] T. Xie, C. Wu, K. Zheng, A Forwarding Prediction Model of Social Network based on Heterogeneous Network, *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Chongqing, China, 2021, pp. 960–964. doi: 10.1109/IAEAC50856.2021.9390634

[13] Y. Liu, W. Han, Y. Tian, X. Que, W. Wang, Trending topic prediction on social network, *2013 5th IEEE International Conference on Broadband Network & Multimedia Technology*, Guilin, China, 2013, pp. 149–154. doi: 10.1109/ICBNMT.2013.6823933

[14] H.-Y. Chen, C.-T. Li, Predicting and Analyzing Privacy Settings and Categories for Posts on Social Media, *2022 IEEE International Conference on Big Data (Big Data)*, Osaka, Japan, 2022, pp. 5692–5697. doi: 10.1109/BigData55660.2022.10020677

[15] P. Chen, X. Qiao, Z. Liu, X. Tian, The design of architecture, workflow, algorithm on grid system for Social Network context prediction analysis, *2009 IEEE*

International Conference on Network Infrastructure and Digital Content, Beijing, China, 2009, pp. 452–455. doi: 10.1109/ICNIDC.2009.5361003.

[16] A. K. Pandey, S. S. Roy, Natural Language Generation Using Sequential Models: A Survey, *Neural Processing Letters*, Vol. 55, No. 6, pp. 7709–7742, December, 2023, doi: 10.1007/s11063-023-11281-6.

[17] D. Yan, K. Li, S. Gu, L. Yang, Network-Based Bag-of-Words Model for Text Classification, *IEEE Access*, Vol. 8, pp. 82641–82652, April, 2020. doi: 10.1109/ACCESS.2020.2991074

[18] C. Liu, Y. Sheng, Z. Wei, Y.-Q. Yang, Research of Text Classification Based on Improved TF-IDF Algorithm, *2018 IEEE International Conference of Intelligent Robotic and Control Engineering (IRCE)*, Lanzhou, China, 2018, pp. 218–222. doi: 10.1109/IRCE.2018.8492945

[19] L. Zhu, M. Huang, M. Chen, W. Wang, An N-gram based approach to auto-extracting topics from research articles, arXiv, September, 2021. doi: 10.48550/arXiv.2110.11879

[20] K. Lee, D. Palsetia, R. Narayanan, Md. M. A. Patwary, A. Agrawal, A. Choudhary, Twitter Trending Topic Classification, *2011 IEEE 11th International Conference on Data Mining Workshops*, Vancouver, BC, Canada, 2011, pp. 251–258. doi: 10.1109/ICDMW.2011.171

[21] J. R. Quinlan, Induction of decision trees, *Machine Learing*, Vol. 1, No. 1, pp. 81–106, March, 1986. doi: 10.1007/BF00116251

[22] S. Bhatia, Application and evaluation of Machine Learning for news article popularity prediction, *2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, Chennai, India, 2021, pp. 1–5. doi: 10.1109/ICSES52305.2021.9633817

[23] T. Chen, C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, Arxiv, June, 2016. https://arxiv.org/abs/1603.02754

[24] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, *Deep contextualized word representations*, arXiv, March, 2018. doi: 10.48550/arXiv.1802.05365

[25] G. Van Houdt, C. Mosquera, G. Nápoles, A review on the long short-term memory model, *Artificial Intelligence Review*, Vol. 53, No. 8, pp. 5929–5955, December, 2020. doi: 10.1007/s10462-020-09838-1

[26] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv, May, 2019. doi: 10.48550/arXiv.1810.04805

[27] Kaggle, *BERT testing on IMDB dataset : Extensive Tutorial*, Accessed: Nov. 09, 2023. [Online]. Available: https://kaggle.com/code/atulanandjha/bert-testing-on-imdb-dataset-extensive-tutorial

[28] H. Wei, W. Zheng, Y. Xiao, C. Dong, News-Comment Relevance Classification Algorithm Based on Feature Extraction, *2021 International Conference on Big Data Analysis and Computer Science (BDACS)*, Kunming, China, 2021, pp. 149–152. doi: 10.1109/BDACS53596.2021.00040

[29] P. Karvelis, D. Gavrilis, G. Georgoulas, C. Stylios, Topic recommendation using Doc2Vec, *2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, Brazil, 2018, pp. 1–6. doi: 10.1109/IJCNN.2018.8489513

[30] A. Kamineni, M. Shrivastava, H. Yenala, M. Chinnakotla, Siamese LSTM with Convolutional Similarity for Similar Question Retrieval, *2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, Pattaya, Thailand, 2018, pp. 1–7. doi: 10.1109/iSAI-NLP.2018.8692937

[31] J.-W. Lin, *Use NLP to Build an LSTM Model for Sentiment Analysis- Taking Conversation Data from the Elderly as an Example*, Master Thesis, National Yunlin University of Science and Technology, Yunlin, Taiwan, 2021.

[32] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, R. Soricut, ALBERT: A Lite BERT for Self-supervised Learning of Language Representations, *International Conference on Learning Representations*, Addis Ababa, Ethiopia, 2020, pp. 1–17. [Online]. Available: https://openreview.net/forum?id=H1eA7AEtvS

[33] F.-L. Wang, *Design and Implementation of an Online Article Popularity Immediately Prediction System*, Master Thesis, Tunghai University, Taichung, Taiwan, 2021.

[34] D. Liao, J. Xu, G. Li, W. Huang, W. Liu, J. Li, Popularity Prediction on Online Articles with Deep Fusion of Temporal Process and Content Features, *Proceedings of AAAI Conference on Artificial Intelligence*, Honolulu, Hawaii, USA, 2019, pp. 200–207. doi: 10.1609/aaai.v33i01.3301200

[35] D. E. Rumelhart, J. L. McClelland, Learning Internal Representations by Error Propagation, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, MIT Press, 1987, pp. 318–362. [Online]. Available: https://ieeexplore.ieee.org/document/6302929

[36] S. H. S. Basha, S. R. Dubey, V. Pulabaigari, S. Mukherjee, Impact of Fully Connected Layers on Performance of Convolutional Neural Networks for Image Classification, *Neurocomputing*, Vol. 378, pp. 112–119, February, 2020. doi: 10.1016/j.neucom.2019.10.008

[37] PyTorch, *Bilinear Layer in PyTorch*, The Linux Foundation. [Online]. Available: https://pytorch.org/docs/stable/generated/torch.nn.Bilinear.html

Biographies

Guan-Wei Chen received his M.S. degree in Computer Science and Information Engineering from National Dong Hwa University, Hualien, Taiwan, R.O.C. in 2023.

Cheng-Chin Chiang received his Ph.D. degree in Computer Science and Information Engineering from National Chiao-Tung University, Hsinchu, Taiwan, R.O.C. in 1993. He is currently a professor of the Department of Computer Science and Information Engineering at National Dong Hwa University, Hualien, Taiwan. His research interests include deep learning, pattern recognition, and human-machine interactions.