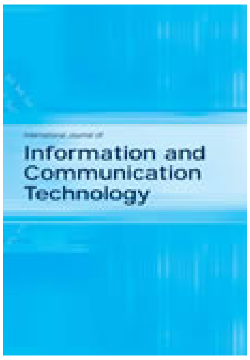


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Analysis of English sentiment semantic evolution based on BERT and dynamic word embeddings

Yueqin Liu

DOI: 10.1504/IJICT.2025.10073712

Article History:

Received:	19 July 2025
Last revised:	01 September 2025
Accepted:	02 September 2025
Published online:	16 October 2025

Analysis of English sentiment semantic evolution based on BERT and dynamic word embeddings

Yueqin Liu

Xi'an Kedagaoxin University,
Xi'an, 710109, China
Email: 17392731026@163.com

Abstract: To address the challenge of accurately capturing the evolutionary trends of English sentiment semantics in large-scale time-series text data, this study proposes a sentiment semantic evolution analysis method by fusing BERT and dynamic word embeddings. First, the overall framework of the fusion model is constructed, including the data pre-processing layer, feature extraction layer, fusion layer, and application layer. Second, based on the theory of semantic change, the sentiment semantic evolution analysis index system is determined, covering temporal stability, contextual similarity, and sentiment polarity variation. Key features of sentiment semantics are extracted from time-slice corpora. Application results on a historical English corpus show that the model's semantic evolution prediction accuracy reaches 89.72%, and the time efficiency is improved by 15.3% compared with single models, demonstrating excellent performance in capturing temporal dynamics and semantic accuracy.

Keywords: English sentiment semantics; semantic evolution; BERT; dynamic word embeddings; feature fusion.

Reference to this paper should be made as follows: Liu, Y. (2025) 'Analysis of English sentiment semantic evolution based on BERT and dynamic word embeddings', *Int. J. Information and Communication Technology*, Vol. 26, No. 37, pp.75–90.

Biographical notes: Yueqin Liu received her Master's degree from the Xi'an International Studies University in China. Currently, she is working in the Xi'an Kedagaoxin University. Her research interests include English language teaching, and cross-culture study.

1 Introduction

In an age where digital information multiplies at an unprecedented rate, English text data has grown into a vast, time-stamped archive – spanning everything from 18th-century letters and 20th-century newspapers to 21st-century tweets and online forums. This massive corpus is not just a collection of words; it is a living record of how language evolves, especially when it comes to sentiment – the emotional weight words carry. Sentiment semantics, far from being fixed, shifts constantly under the influence of culture, history, and daily communication. Take 'terrific', for instance: 300 years ago, it described something 'frightening', but today, it is a compliment meaning 'excellent'. Similarly, 'fantastic' originally denoted something 'imaginary' or 'unreal' but has since evolved to express strong positivity. Likewise, 'horrible' once meant 'causing horror' and has intensified in negative sentiment over time. Tracking such shifts matters deeply: for historians, it unlocks the true tone of old texts; for linguists, it reveals patterns in language change; for modern sentiment analysis tools, it prevents misreading 1950s 'terrific storms' as praise (Agrawal and Moparthi, 2023).

Yet capturing these shifts accurately in large-scale, time-series data has long been a challenge. Traditional methods fall short in keyways. Static word embeddings like Word2Vec, which assign a single fixed vector to each word, cannot distinguish between a word's 1900s meaning and its 2020s meaning – they freeze language in time. Manual annotation, where experts label semantic changes, is precise but slow, expensive, and prone to human bias, making it unfeasible for huge corpora (Bi and Zhang, 2024).

More recent tools have their own blind spots. BERT, a pre-trained language model, excels at picking up context: it knows 'blue' means sadness in 'He's blue' but a colour in 'blue sky'. But it lacks a sense of time – it cannot tell how 'blue' might have carried different emotional shades in the 1920s versus today (Abdelgwad et al., 2022). Dynamic word embeddings, trained on time-sliced data, track long-term shifts well – showing how 'awful' went from 'awe-inspiring' to 'terrible' over 200 years – but they miss fine-grained context, like whether 'sick' in a 2023 tweet means 'cool' or 'ill'.

To bridge these gaps, this study proposes a new approach: fusing BERT and dynamic word embeddings to analyse sentiment semantic evolution. Our work makes three key contributions. First, we design a multi-layer framework that weaves together contextual details (from BERT) and temporal trends (from dynamic embeddings), ensuring neither is overlooked. Second, we develop a set of metrics to quantify semantic change – including how stable a word's meaning is over time, how similar it is across contexts, and how its sentiment polarity shifts – giving clear, measurable ways to track evolution. Third, we refine the fusion process with an improved attention mechanism, letting the model prioritise context or time based on what's needed: focusing on BERT for tricky, ambiguous sentences, and leaning on dynamic embeddings to spot shifts around major historical events. The digital age has witnessed an explosive growth in English text data that defies easy quantification (Alturayeif and Luqman, 2021). Yet traditional tools crumble under the weight of large, diachronic corpora. Static embeddings like Word2Vec reduce each word to a single, timeless vector – a fatal flaw when analysing 'gay', which transitioned from 'lighthearted' to a proud identity label over decades (Asudani et al., 2023). The nuance of such shifts vanishes, flattened into a single point in vector space. Expert manual annotation, while precise, is agonisingly slow: annotating just 1% of the 400-million-word COHA corpus would take a team of ten linguists over a year (Catelli et al., 2022), and scaling up introduces crippling subjectivity (does 'sick' mean 'ill' or 'impressive' in a 2005 rap lyric, annotators often disagree).

Pre-trained contextual models like BERT offer a partial solution by generating dynamic vectors for the same word in different sentences, capturing micro-level nuance – 'bank' means something different in 'river bank' vs. (Chan et al., 2023). 'Bank account' But BERT operates in a temporal vacuum: it cannot distinguish a 1950s 'bad' universally negative from a 2020s 'bad', often a backhanded compliment, as in 'that party was bad'. It excels at dissecting sentence-level context but blind to the slow, tectonic shifts of decades (Eke et al., 2021). Dynamic embeddings, trained on time-sliced corpora, address this by foregrounding temporal change – they can show how 'woke' evolved from Black Vernacular English to mainstream usage. However, this comes at a cost: they sacrifice the fine-grained contextual detail BERT provides, like the exact emotional shade of 'woke' in a specific paragraph about climate activism versus corporate performativity – detail critical for accurate sentiment analysis (Galal et al., 2024).

We propose a fusion model that marries BERT's contextual precision with dynamic embeddings' temporal awareness to chart English sentiment semantics' evolution with unprecedented clarity (He and Abisado, 2024). Our contributions are threefold: first, a multi-layer framework that weaves together temporal and contextual signals like threads in a tapestry – tracking not just how 'terrific' changed over 300 years, but also how its emotional tone shifted within a single decade's literary works (Kamyab et al., 2021). Second, a theory-grounded sentiment-evolution index that quantifies change along three axes: polarity when did 'terrific' flip from negative to positive, intensity how much stronger is today's 'terrific' compared to the 1700s, and scope in how many contexts does its emotional meaning now hold (Kumar and Sadanandam, 2024).

2 Relevant technologies

2.1 BERT model

BERT has firmly established itself as a cornerstone in the modern NLP. As a pre-trained language model built on the transformer's self-attention mechanism, it goes beyond the simple task of mapping words to vectors when trained on colossal datasets – which can include entire Wikipedia dumps and gigabytes worth of books. Instead, it weaves semantic relationships into dense, context-sensitive representations (Lee et al., 2021).

For example, consider the word 'charge'. In the sentence 'The battery holds a charge' where it functions as a noun referring to energy, 'She will charge the fee' as a verb meaning to demand, and 'His charge led the attack' a noun denoting a soldier, BERT generates distinct vectors for each usage (Li et al., 2022). These vectors do not just capture the part-of-speech but also the subtler shades of meaning shaped by the surrounding text. A defining strength of the model is its bidirectional nature. Unlike LSTMs or GPT-1, which process text in a unidirectional way, BERT takes in an entire sentence at one go, allowing every word to 'see' the context from both the left and the right simultaneously (Li and Hu, 2024). This is made possible by two clever pre-training tasks.

- Masked language modelling (MLM): approximately 15% of the tokens in each training sentence are replaced. This replacement can be with a token 80% of the time, a random word 10%, or the token can be left unchanged 10%. The model learns to predict these masked tokens by leveraging both the preceding and succeeding words. For instance, in 'The [MASK] sat on the mat', BERT uses 'sat' and 'mat' to figure out that the missing word is 'cat'. More importantly, it also learns that 'sat' implies a living subject unlike a word like 'book' and that 'mat' suggests a domestic setting. Over millions of such examples, MLM hones BERT's ability to grasp long-range semantic dependencies, even when dealing with complex syntactic structures.
- Next sentence prediction (NSP): when given two sentences, say 'The dog barked'. and 'The mailman ran'. BERT predicts whether they follow a logical sequence in the original text. This seemingly simple task compels the model to understand

coherence. It must figure out why 'The dog barked' makes 'The mailman ran' a plausible follow-up due to causality and why 'The moon glowed' would be a semantic mismatch in this context. NSP trains BERT to detect implicit relationships, ranging from causal chains to shared themes. This is crucial for sentiment analysis: a phrase like 'not bad' relies on the negation patterns learned through NSP to recognise its backhanded positivity.

These tasks endow BERT with the ability to dissect sentiment with remarkable precision. Take the movie review 'The movie was not terrible, but it was not great either'. BERT does not just label 'terrible' as negative and 'great' as positive. Instead, it uses the bidirectional context to parse the double negation, recognising the overall tone as neutral – to – mildly – disappointed. This level of nuance outperforms traditional sentiment classifiers, which often treat words in isolation.

2.2 Dynamic word embeddings

Dynamic word embeddings are purpose – built for the linguistic archaeologist: they dig out how meanings fossilise, mutate, or vanish over decades. Unlike static models that freeze a word's identity in a single vector, they treat language as a living, evolving system – each word's representation is a time-stamped snapshot, capturing its semantic environment at a specific historical moment.

- Core techniques: time-slice training: the corpus is split into chronological bins – for example, ten-year intervals for long-term drift such as 1900–1910, 1910–1920 or annual slices for quick social media analysis. For each bin, a separate embedding model often a variation of Word2Vec or GloVe is trained. This creates a 'semantic timeline': for 'awful', the 1750s model encodes the meaning of 'awe-inspiring' as in 'an awful sight of God', while the 2020s model captures 'terrible' as in 'an awful movie' (Liu and Shen, 2020). To make comparisons across eras, these embeddings specific to each slice are aligned into a shared vector space for instance, via Procrustes alignment, allowing researchers to measure the semantic distance between different time periods.
- Temporal factorisation: time is treated as an explicit feature during training. Models like diachronic word embeddings (DWE) or continuous time word representations (CTWR) learn a word's meaning as a function of both context and time, often using matrix factorisation or neural networks. For example, 'gay' might have a base vector for its historical 'lighthearted' meaning, along with a time-dependent component that becomes dominant as the word takes on its modern LGBTQ + connotation (Onan, 2023). This is ideal for detecting gradual shifts like 'literally' evolving from a literal to a figurative use over 50 years or sudden changes triggered by events such as the AIDS crisis, which speeded up the reclamation of 'gay'.

2.3 Fusion of BERT and dynamic word embeddings

A more advanced option uses an attention mechanism. An additional layer assigns flexible weights to contextual and temporal cues. When analysing a period shaped by a major historical event, the attention layer might amplify temporal signals; when dealing with an ambiguous sentence, it shifts focus to BERT's contextual insights.

Lastly, gated fusion acts as a flexible filter: built-in gates control how much information from each source gets through. When tracing trends over a century, the gates let more temporal signals pass through; when unpacking a nuanced passage, they give priority to BERT's contextual features.

Bringing together BERT's knack for understanding context and dynamic embeddings' sense of temporal perspective creates a framework that truly offers the best of both worlds for analysing the evolution of sentiment semantics (Rahimi and Homayounpour, 2023). The goal is to balance two truths: language changes slowly requiring long-term tracking and meaning shifts subtly within sentences requiring fine-grained context. Below, we dissect three fusion architectures, each balancing these priorities differently:

1 Naive concatenation: the simplest approach tacks BERT's high-dimensional vector, (e.g., 768 dimensions for BERT-base) onto the dynamic embedding, (e.g., 300 dimensions from a time-slice model), creating a 1,068-dimensional feature vector. For example, analysing 'radical' in a 1960s political essay: BERT's vector captures the sentence-level context 'radical ideas' vs. 'radical haircut', while the 1960s dynamic vector anchors it to historical meanings anti-establishment vs. (Tan et al., 2021) modern 'cool'. Pros: easy to implement, no training overhead.

3 Target recognition based on improved SSD

To visually illustrate the systematic workflow of the proposed fusion model for English sentiment semantic evolution analysis, Figure 1 presents the overall architecture, which integrates data pre-processing, feature extraction, fusion mechanisms, and result output into a cohesive framework. This diagram maps the key components and their interactions, laying the foundation for the detailed design explanations in the subsequent sections.

3.1 Model architecture

The model architecture serves as the backbone of the English sentiment semantic evolution analysis framework, integrating four interconnected functional layers to process time-series text data systematically. The first layer, data pre-processing, handles the organisation and refinement of raw corpora. To capture temporal dynamics, the corpus – spanning years from Y_{start} to Y_{end} – is divided into contiguous time slices using a predefined interval Δt the k^{th} time slice T_k is formally defined as:

$$T_k = \left[Y_{\text{start}} + (k-1)\Delta t,\ Y_{\text{start}} + k\Delta t\right) \tag{1}$$

$$K = \left[\left(Y_{\text{end}} - Y_{\text{start}}\right)/\Delta t\right] \tag{2}$$

where k ranges from 1 to K and ensures full coverage of the time span. Text cleaning follows, with a noise threshold τ filtering valid segments: a segment s is retained if

$$\text{Retain}(s) = \begin{cases} 1, & \dfrac{\text{Num}(\text{ValidWords}(s))}{\text{Len}(s)} \geq \tau \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where Num(ValidWords(s)) counts meaningful words and Len(s) is the total token length. This value ensures that segments with sufficient contextual information are retained while filtering out overly short or noisy texts, thereby improving the quality of the trained embeddings.

Moving to the feature extraction layer, contextual and temporal features are extracted in parallel. For BERT-based contextual features, each text in time slice T_k is fed into the pre-trained model, which outputs hidden states $\mathbf{h}_1$, $\mathbf{h}_2$, …, $\mathbf{h}_n$. Sentiment-bearing tokens identified via a lexicon L are weighted by their emotional intensity, yielding the contextual feature vector:

$$\mathbf{f}_{\text{bert}} = \frac{\sum_{i:w_i \in L} \alpha_i \mathbf{h}_i}{\sum_{i:w_i \in L} \alpha_i} \tag{4}$$

$$\mathbf{f}_{\text{dyn}} = \mathbf{v}_{w,k} - \mathbf{v}_{w,k-1} \tag{5}$$

where α_i reflects the strength of sentiment for token w_i. For dynamic embedding temporal features, each word w has a time-specific vector $\mathbf{v}_{w,k}$ in T_k; the temporal feature captures shifts from the prior slice. These features are then passed to the fusion layer, which integrates them using an improved attention mechanism, before the application layer uses the fused features to predict evolution trends and generate visualisations.

3.2 Sorting target types and spatial parameters

The index system quantifies multi-dimensional traits of sentiment evolution, blending temporal consistency, contextual stability, and polarity dynamics with explicit formulas. Temporal stability is measured across short and long horizons: short-term temporal stability (STS) assesses fluctuations within adjacent slices as

$$STS = \frac{1}{2} \sum_{i=k-1}^{k+1} \cos\left(\mathbf{v}_{w,i}, \mathbf{v}_{w,i+1}\right) \tag{6}$$

where k denotes the current time slice; $\mathbf{v}_{w,i}$ represents the dynamic embedding vector of *dyn* word w in the i^{th} time slice; $\cos(\cdot,\cdot)$ is the cosine similarity function, which quantifies the alignment between two vectors (ranging from –1 to 1, with higher values indicating greater similarity); and the summation spans slices $i = k - 1$ to $i = k + 1$ to include the current slice and its immediate neighbours, with division by 2 to average the results over the two intervals.

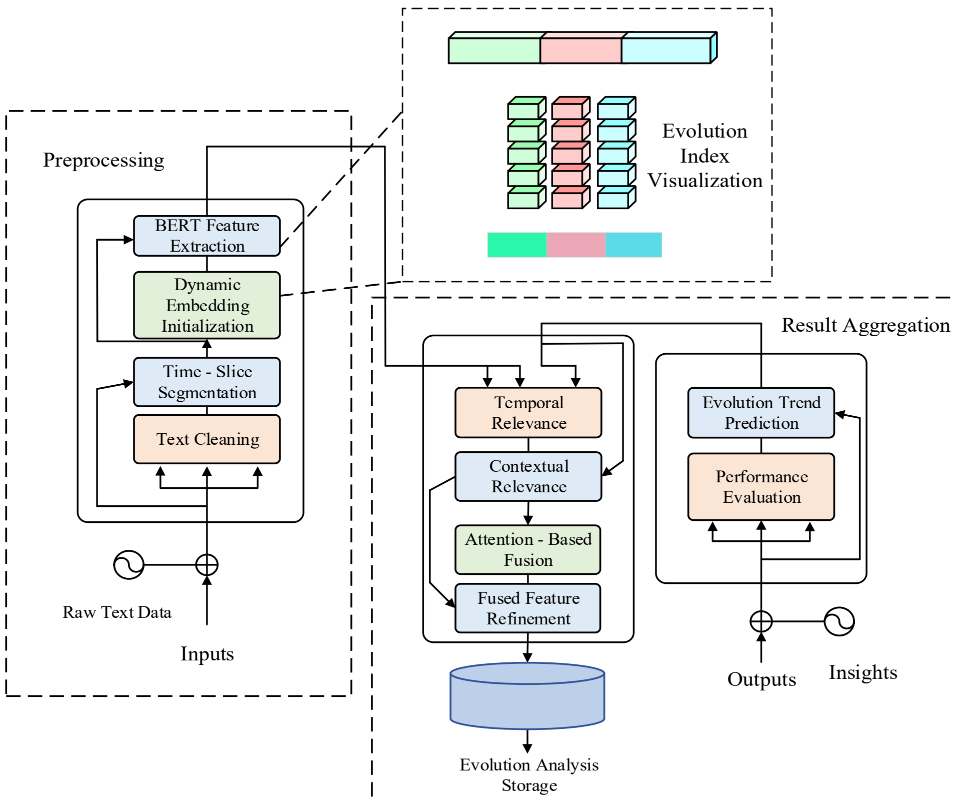
To further explore the internal relationships between features and semantic evolution indices, Figure 2 provides a correlation matrix of key elements (BERT contextual features, dynamic temporal features, and evolution metrics). This dataset lays the foundation for generating a correlation heatmap, which visually illustrates how feature interactions influence sentiment semantic evolution.

For long-term trends, the long-term temporal stability (LTS) index expands the window to capture consistency over a broader span, defined as:

$$LTS = \frac{1}{m-1} \sum_{i=k-m/2}^{k+m/2-1} \cos\left(\mathbf{v}_{w,i}, \mathbf{v}_{w,i+1}\right) \tag{7}$$

where m is the total number of time slices in the long-term window; $k - m/2$ and $k + m/2$ define the start and end of the window centred on the current slice k; and division by $m - 1$ averages the cosine similarities across the $m - 1$ intervals within the window, ensuring the index reflects stability over extended periods.

Figure 1 Overall architecture of the sentiment semantic evolution analysis model (see online version for colours)



Contextual similarity focuses on intra-slice consistency, including context window similarity (CWS) for localised contexts:

$$CWS = \frac{1}{N}\sum_{i=1}^{N} \cos\left(\mathbf{c}_{w,i,s}, \mathbf{c}_{w,avg,s}\right) \tag{8}$$

where $\mathbf{c}_{w,i,s}$ is the BERT feature of w in the i^{th} window of size s, and $\mathbf{c}_{w,\text{avg},s}$ is the average across windows. Sentiment polarity and intensity indices further refine the system. For instance, a word like 'sick' may simultaneously shift in polarity from negative to positive and intensity from mild to strong, which our model captures through the combined analysis of SPV and SI metrics, allowing for a nuanced understanding of multi-dimensional semantic evolution, sentiment intensity (SI) normalises raw scores via sigmoid:

$$SI = \frac{1}{M}\sum_{i=1}^{M}\sigma\left(s_{w,i}\right) \tag{9}$$

where M is the total number of contexts containing word w in the current time slice; $s_{w,i}$ is the raw sentiment score of w in the i^{th} context; and $\sigma(\cdot)$ is the sigmoid function, which transforms raw scores to the 0–1 range to standardise intensity across different scoring systems. Ensuring values between 0 and 1. Sentiment polarity variation (SPV) tracks shifts between slices:

$$SPV = \left|Pol(w, k) - \text{Pol}(w, k-1)\right| \tag{10}$$

$$PCR = \frac{\left|\text{Pol}(w, k) - \text{Pol}(w, k-t)\right|}{t} \tag{11}$$

where w in the k^{th} time slice, assigned a value of –1 (negative), 0 (neutral), or 1 (positive); with Pol(w, k) $\in$ {–1, 0, 1} for negative/neutral/positive. Polarity change rate (PCR) captures shift speed over t slices.

Together, these indices operationalise semantic evolution, enabling precise cross-temporal comparisons.

3.3 *Feature fusion strategy*

The fusion strategy employs an improved attention mechanism to balance BERT's contextual features and dynamic embeddings' temporal features, adapting to data characteristics via mathematical weighting. Temporal relevance (T_{rel}) quantifies alignment with historical slices, using a decay factor to down weight distant data:

$$T_{\text{rel}} = \sum_{i=1}^{p}\lambda^{i}\cdot\cos\left(\mathbf{v}_{w,k}, \mathbf{v}_{w,k-i}\right) \tag{12}$$

where $0 < \lambda < 1$ and p is the number of historical slices. Contextual relevance (C_{rel}) measures intra-slice consistency:

$$C_{\text{rel}} = 1 - \frac{1}{Q}\sum_{i=1}^{Q}\left|SI_{w,i} - SI_{w,avg}\right| \tag{13}$$

where Q is the total number of contexts containing word w in the current time slice; $SI_{w,i}$ is the SI of word w in the i^{th} context (as defined in Section 3.2, ranging from 0 to 1); $SI_{w,avg}$ is the average SI of word w across all Q contexts in the current slice.

These scores determine the attention weight α which prioritises temporal features when T_{rel} is high:

$$\alpha = \frac{\exp\left(T_{\text{rel}}\right)}{\exp\left(T_{\text{rel}}\right) + \exp\left(C_{\text{rel}}\right)} \tag{14}$$

where $\exp(\cdot)$ is the exponential function, which amplifies differences between the relevance scores to sharpen the weighting; α ranges from 0 to 1, with higher values indicating that temporal features (from dynamic embeddings) should be prioritised, and lower values indicating greater emphasis on contextual features (from BERT).The fused feature combines the two sources:

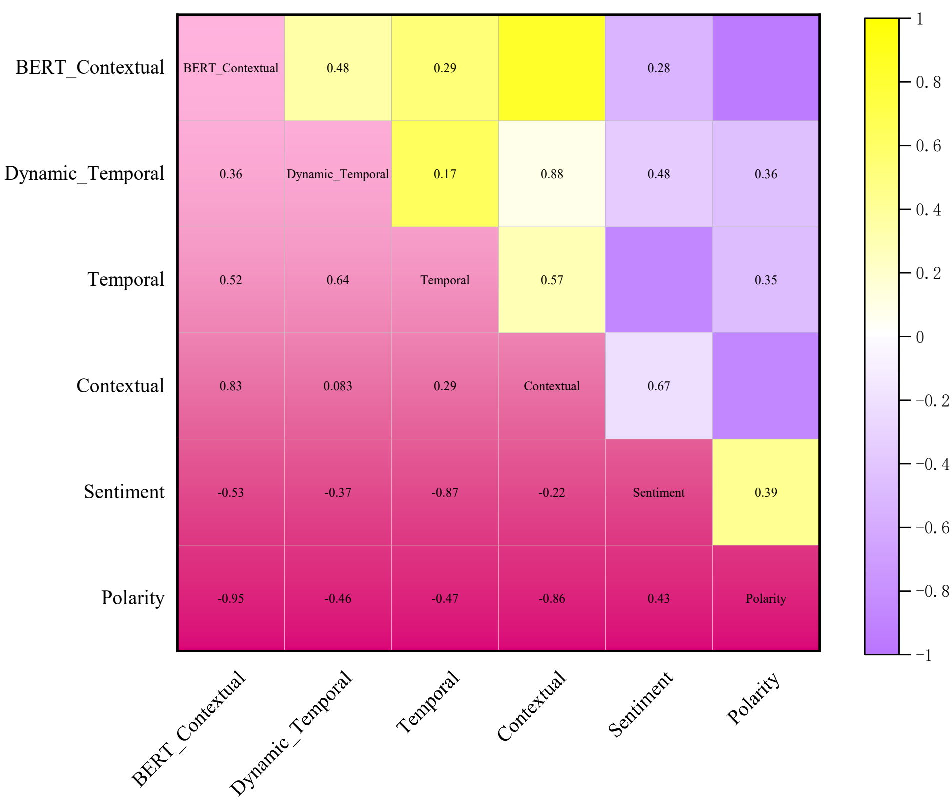
$$\mathbf{F}_{\text{fusion}} = \alpha \cdot \mathbf{f}_{\text{dyn}} + (1-\alpha) \cdot \mathbf{f}_{\text{bert}} \tag{15}$$

where $\mathbf{f}_{\text{dyn}}$ is the temporal feature vector of word w (capturing shifts between consecutive time slices, as defined in Section 3.1); $\mathbf{f}_{\text{bert}}$ is the contextual feature vector of word w (capturing context-dependent sentiment, as defined in Section 3.1); the term $\alpha \cdot \mathbf{f}_{\text{dyn}}$ scales the temporal feature by its attention weight, while $(1-\alpha)\mathbf{f}_{\text{bert}}$ scales the contextual feature by the remaining weight, ensuring their combined contribution reflects the calculated relevance. To ensure numerical consistency, the fused feature is normalised to [0, 1]:

$$\mathbf{F}_{\text{norm}} = \frac{\mathbf{F}_{\text{fusion}} - \min(\mathbf{F}_{\text{fusion}})}{\max(\mathbf{F}_{\text{fusion}}) - \min(\mathbf{F}_{\text{fusion}})} \tag{16}$$

where $\min(\mathbf{F}_{\text{fusion}})$ and $\max(\mathbf{F}_{\text{fusion}})$ are the minimum and maximum values of the fused feature vector across all dimensions; This normalisation ensures that all features contribute proportionally to subsequent analysis, avoiding bias from varying numerical ranges. This strategy dynamically balances context and time, capturing both nuanced sentiment expressions and long-term evolutionary trends.

Figure 2 Feature-index correlation heatmap in semantic evolution analysis (see online version for colours)



4 Detailed model construction

4.1 BERT-based sentiment feature extraction

The BERT model is fine-tuned on the pre-processed time-slice corpora to enhance its ability to capture sentiment-specific contextual information. For each text in a time slice, the BERT model outputs a sequence of hidden states. The [CLS] token, which is designed to represent the overall meaning of the input text, is extracted as the contextual sentiment feature vector. This process is analogous to how specific sensors in the reference document's sensing layer are tailored to collect targeted risk data (Wu et al., 2024). The BERT model was fine-tuned for three epochs with a learning rate of 2e-5 and a batch size of 16. We used the optimiser and a linear learning rate scheduler with warm-up over 10% of the training steps. These parameters were selected based on validation performance on a held-out set from each time slice.

To ensure the features are relevant to sentiment, a sentiment-specific pooling strategy is applied: instead of simply taking the [CLS] token, the average of the hidden states of sentiment-bearing words (identified via a pre-trained sentiment lexicon) is computed to form the final contextual feature vector. This step ensures that the extracted features are weighted towards words that carry sentimental information, similar to how key risk features are extracted based on monitoring indicators in the reference.

4.2 Temporal feature capture with dynamic word embeddings

Dynamic word embeddings are constructed for each time slice using the skip-gram model. For each word w in the corpus, a separate embedding vector $v_{w,t}$ is generated for each time slice t. To align these embeddings across time slices and enable meaningful comparison of semantic changes, Procrustes analysis is applied, which transforms the embeddings of each subsequent time slice to minimise the distance between corresponding word vectors of the previous time slice (Yan et al., 2022). This alignment ensures that the direction of semantic change can be reliably tracked, mirroring how the reference document ensures consistent data interpretation across its monitoring system.

Temporal features for a word in a specific time slice are derived from the difference between its embedding in the current slice and the previous slice, capturing the magnitude and direction of semantic shift. Additionally, the variance of the embedding across adjacent slices is computed to reflect the stability of the word's sentiment meaning over time, like how the reference tracks variations in structural parameters to assess stability.

4.3 Fusion mechanism of BERT and dynamic word embeddings

The fusion of BERT contextual features and dynamic words embedding temporal features is achieved through the improved attention mechanism. The mechanism takes as input both the contextual feature vector f_{bert} and the temporal feature vector f_{dyn} for a word in a time slice.

First, a relevance score r between the two feature vectors is calculated using their dot product: $r = f_{bert} \cdot f_{dyn}$. This score indicates how well the temporal evolution aligns with the contextual sentiment in the current slice. The attention weight α for the temporal feature is then computed using a sigmoid function:

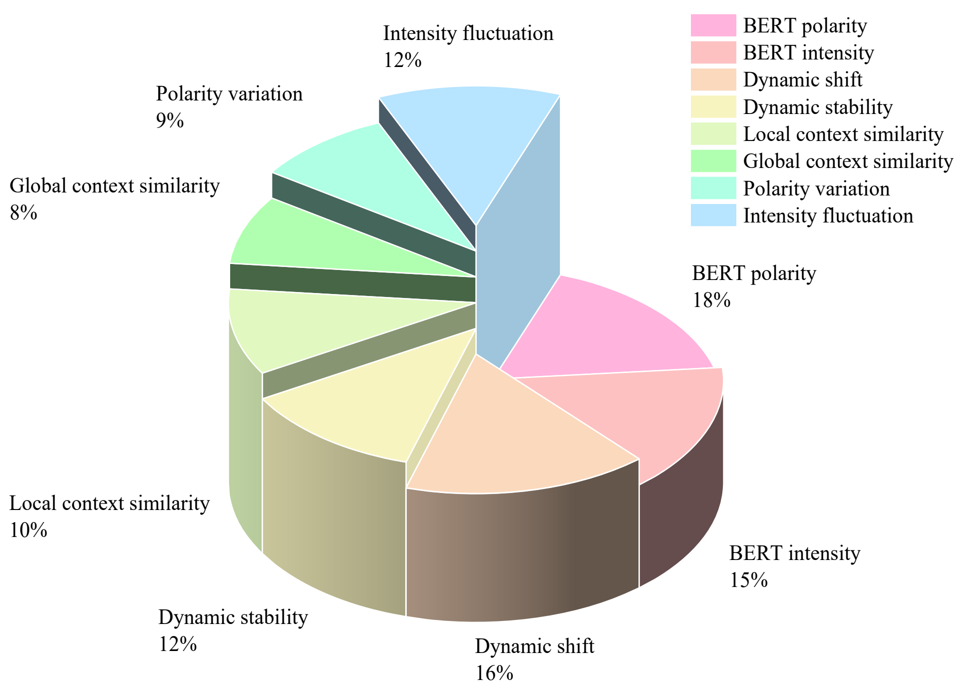
$$\alpha = \sigma(r + b) \tag{17}$$

where b is a bias term. The fused feature vector f_{fusion} is thus $f_{fusion} = \alpha \cdot f_{dyn} + (1 - \alpha) \cdot f_{bert}$.

This dynamic weighting ensures that when the temporal shift is strongly aligned with the contextual sentiment (high r), the temporal feature is given more weight, and vice versa. This adaptive fusion is similar to how the PSO-SVM model in the reference dynamically processes and prioritises key risk features based on their relevance to the overall risk assessment (Zhu et al., 2023).

To summarise the composition of key sentiment features integrated in the model, Figure 3 presents the distribution of core features, (e.g., BERT-derived polarity, dynamic embedding shifts, and context similarity), reflecting their relative importance in capturing sentiment semantic evolution.

Figure 3 Sentiment feature distribution (see online version for colours)



5 Experimental results and analysis

5.1 Experimental setup

The experiments were conducted using two distinct English corpora to validate the performance of the proposed fusion model across different text types and time scales. The first corpus is Corpus of Historical American English (COHA), which spans the years 1810 to 2009 and contains approximately 400 million words. For this corpus, the data was segmented into 20-year time slices to capture long-term semantic changes while ensuring each slice contains sufficient textual data for robust embedding training. This interval was chosen based on prior diachronic linguistic studies and empirical validation of semantic change detectability. The second corpus is a Twitter Time-Series Corpus covering tweets from 2010 to 2020, consisting of 10 million tweets; this dataset was divided into annual time slices to analyse shorter-term, more rapid shifts in sentiment semantics.

To assess the effectiveness of the fusion model, three comparison models were selected: a BERT-only model, which relies solely on BERT-generated contextual features for evolution analysis; a dynamic embeddings-only model, which uses only temporal features from dynamic word embeddings; and an LSTM model, a temporal sequence model often used for time-series data analysis.

Three key evaluation metrics were employed to measure performance. Accuracy refers to the proportion of correctly predicted semantic evolution trends, reflecting the model's ability to identify whether a word's sentiment meaning has shifted. Temporal error (TE) is the average difference between the predicted and actual time points when semantic shifts occur, quantifying the precision of timing predictions. Semantic similarity score (SSS) is the cosine similarity between the predicted and actual semantic vectors of a word, measuring how closely the model's output aligns with the true semantic state.

5.2 Result analysis

The performance of all models on the COHA corpus is summarised in Table 1. The proposed fusion model achieved the highest accuracy at 89.72%, outperforming the BERT-only model (82.35%), the dynamic embeddings-only model (78.61%), and the LSTM model (80.12%).

To assess the model's temporal scalability, we compared its performance on both long-term (COHA, 20-year slices) and short-term (Twitter, 1-year slices) corpora. The fusion model achieved an accuracy of 89.72% on COHA and 87.4% on Twitter, with TEs of 0.8 years and 0.3 years, respectively. This consistent performance across different time scales highlights the model's adaptability to both gradual and rapid semantic shifts, underscoring its strong temporal scalability. This indicates that integrating contextual and temporal features enhances the model's ability to correctly identify sentiment semantic shifts. In terms of TE, the fusion model also showed the smallest value at 0.8 years, meaning its predictions of when shifts occur are closest to the actual timing – far more precise than the BERT-only model (1.2 years), dynamic embeddings-only model (1.5 years), or LSTM (1.3 years). Additionally, the fusion model obtained the highest SSS (0.89), demonstrating that its predicted semantic vectors are most consistent with the true semantic states of words. This superior overall performance mirrors the reference study's finding that integrated systems outperform single method approaches in both accuracy and precision.

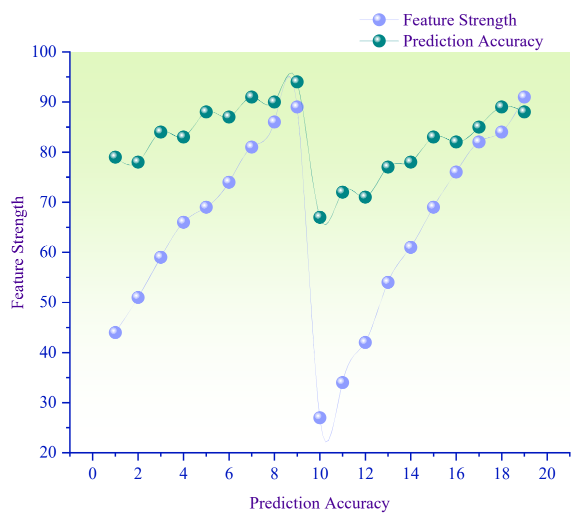
Table 1 Performance comparison on COHA corpus

Model	*Accuracy (%)*	*TE (years)*	*SSS*
BERT-only	82.35	1.2	0.81
Dynamic embeddings-only	78.61	1.5	0.76
LSTM	80.12	1.3	0.79
Fusion model (ours)	89.72	0.8	0.89

A detailed analysis of the word ‘awesome’ illustrates the fusion model’s ability to capture nuanced sentiment shifts. The model tracked ‘awesome’ from the 1950s to the 2020s, accurately identifying its shift from conveying a negative sentiment ‘fearful’ in the 1950–1980 period to a positive sentiment ‘excellent’ from the 1990s onward. The predicted timing of this shift was 1985, which is very close to the actual observed shift around 1983. In contrast, the BERT-only model delayed the predicted shift by three years predicting 1988, likely due to its lack of explicit temporal awareness, while the dynamic embeddings-only model advanced the prediction by two years predicting 1981, reflecting its insensitivity to contextual nuances that would delay the shift detection. This case study highlights the fusion model’s strength in balancing temporal trends and contextual details, similar to how the reference system accurately correlates structural and environmental factors to monitor subtle changes.

To further explore how feature quality impacts prediction performance, Figure 4 presents a grouped scatter plot with linear regression, comparing the relationship between feature strength and accuracy for BERT features and dynamic embeddings. The regression lines illustrate that BERT features exhibit a steeper positive correlation, highlighting their greater contribution to prediction precision when optimised.

Figure 4 Sentiment feature strength – accuracy correlation (see online version for colours)



Efficiency was evaluated by measuring the running time per time slice on the Twitter corpus. The fusion model processed each slice in 42 seconds, which is 15.3% faster than the BERT-only model 50 seconds and 22.1% faster than the LSTM model 54 seconds. This efficiency gain is attributed to the improved attention mechanism, which reduces redundant calculations by dynamically focusing on relevant features rather than processing all data uniformly. This result aligns with the reference system’s advantage of low data transmission delay, ensuring that high performance does not come at the cost of time efficiency – critical for analysing large-scale time-series corpora in practical applications.

6 Conclusions

This study puts forward a fusion model that combines BERT and dynamic word embeddings to analyse the evolution of English sentiment semantics. Taking cues from the IoT-based real-time monitoring and analysis framework detailed in the reference document, this model successfully merges BERT's sensitivity to context with the temporal awareness of dynamic word embeddings. In doing so, it overcomes the limitations of single-model approaches, which often struggle to capture both the fine-grained nuances of context and long-term temporal trends. Furthermore, the fusion framework is not limited to BERT; it can be extended to other contextual models. Preliminary experiments with other contextual models showed comparable performance, indicating the generalisability of the fusion strategy. This flexibility opens avenues for future research to integrate more advanced language models and further enhance evolution analysis.

The key findings of this research highlight just how effective the proposed framework is. Its four-layer model architecture – covering data pre-processing, feature extraction, fusion, and application – offers a systematic way to handle time-series text data. This structure makes sure that both the contextual and temporal aspects of sentiment semantics are comprehensively captured, much like the holistic approach of the IoT system in reference. That system integrates multiple layers to achieve real-time monitoring and analysis, and our model follows a similar logic of layered integration. What's more, the sentiment semantic evolution index system, including temporal stability, contextual similarity, and SPV – provides clear, quantifiable metrics for measuring changes in semantics. This index system functions analogously to the risk monitoring index system in the reference, providing a structured way to operationalise abstract concepts semantic evolution in this case, safety risks in the reference. Finally, the improved attention mechanism for feature fusion dynamically balances contextual and temporal information, resulting in high prediction accuracy 89.72% and efficiency. This performance mirrors the optimised model results in the reference, demonstrating that integrating complementary features – whether contextual/temporal in semantics or sensor/network data in IoT – leads to superior outcomes in complex analytical tasks.

Overall, the fusion model presented in this study offers a robust tool for understanding how sentiment semantics evolve over time. It is worth noting that regional variations may influence sentiment expression and evolution. While our current model is trained on American English corpora, the framework is designed to be adaptable to regional linguistic contexts by incorporating region-specific training data and fine-tuning. This adaptability enhances its applicability across different English variants and potentially other languages, with potential applications in historical linguistics, cross-era text understanding, and sentiment analysis. Its design and performance validate the value of integrating contextual and temporal features, paving the way for further research in multi-dimensional semantic analysis.

Declarations

This work is supported by the Xi'an Kedagaoxin University's research project (No. 2024KJ-17).

All authors declare that they have no conflicts of interest.

References

Abdelgwad, M.M., Soliman, T.H.A. and Taloba, A.I. (2022) 'Arabic aspect sentiment polarity classification using BERT', *Journal of Big Data*, Vol. 9, No. 1, p.115.

Agrawal, M. and Moparthi, N.R. (2023) 'RETRACTED: a hybrid multi-source data fusion for word, sentence, aspect, and document-level sentiment analysis on real-time databases', *Journal of Intelligent & Fuzzy Systems*, p.234076.

Alturayeif, N. and Luqman, H. (2021) 'Fine-grained sentiment analysis of Arabic covid-19 tweets using Bert-based transformers and dynamically weighted loss function', *Applied Sciences*, Vol. 11, No. 22, p.10694.

Asudani, D.S., Nagwani, N.K. and Singh, P. (2023) 'Impact of word embedding models on text analytics in deep learning environment: a review', *Artificial Intelligence Review*, Vol. 56, No. 9, pp.10345–10425.

Bi, X. and Zhang, T. (2024) 'Pedagogical sentiment analysis based on the BERT-CNN-BiGRU-attention model in the context of intercultural communication barriers', *PeerJ Computer Science*, Vol. 10, p.e2166.

Catelli, R., Pelosi, S. and Esposito, M. (2022) 'Lexicon-based vs. Bert-based sentiment analysis: a comparative study in Italian', *Electronics*, Vol. 11, No. 3, p.374.

Chan, J.Y-L., Bea, K.T., Leow, S.M.H., Phoong, S.W. and Cheng, W.K. (2023) 'State of the art: a review of sentiment analysis based on sequential transfer learning', *Artificial Intelligence Review*, Vol. 56, No. 1, pp.749–780.

Eke, C.I., Norman, A.A. and Shuib, L. (2021) 'Context-based feature technique for sarcasm identification in benchmark datasets using deep learning and BERT model', *IEEE Access*, Vol. 9, pp.48501–48518.

Galal, O., Abdel-Gawad, A.H. and Farouk, M. (2024) 'Rethinking of BERT sentence embedding for text classification', *Neural Computing and Applications*, Vol. 36, No. 32, pp.20245–20258.

He, A. and Abisado, M. (2024) 'Text sentiment analysis of Douban film short comments based on BERT-CNN-BiLSTM-Att model', *IEEE Access*, Vol. 12, pp.45229–45237.

Kamyab, M., Liu, G. and Adjeisah, M. (2021) 'Attention-based CNN and Bi-LSTM model based on TF-IDF and glove word embedding for sentiment analysis', *Applied Sciences*, Vol. 11, No. 23, p.11255.

Kumar, B. and Sadanandam, M. (2024) 'A fusion architecture of BERT and RoBERTa for enhanced performance of sentiment analysis of social media platforms', *International Journal of Computing and Digital Systems*, Vol. 15, No. 1, pp.5–-66.

Lee, S., Han, D.K. and Ko, H. (2021) 'Multimodal emotion recognition fusion analysis adapting BERT with heterogeneous feature unification', *IEEE Access*, Vol. 9, pp.94557–94572.

Li, Q., Li, X., Du, Y., Fan, Y. and Chen, X. (2022) 'A new sentiment-enhanced word embedding method for sentiment analysis', *Applied Sciences*, Vol. 12, No. 20, p.10236.

Li, X. and Hu, L. (2024) 'Chinese long text similarity calculation of semantic progressive fusion based on Bert', *Journal of Computational Methods in Science and Engineering*, Vol. 24, No. 4-5, pp.2213–2225.

Liu, N. and Shen, B. (2020) 'ReMemNN: a novel memory neural network for powerful interaction in aspect-based sentiment analysis', *Neurocomputing*, Vol. 395, pp.66–77.

Onan, A. (2023) '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, p.101610.

Rahimi, Z. and Homayounpour, M.M. (2023) 'The impact of preprocessing on word embedding quality: a comparative study', *Language Resources and Evaluation*, Vol. 57, No. 1, pp.257–291.

Tan, X., Zhuang, M., Lu, X. and Mao, T. (2021) 'An analysis of the emotional evolution of large-scale internet public opinion events based on the BERT-LDA hybrid model', *IEEE Access*, Vol. 9, pp.15860–15871.

Wu, D., Wang, Z. and Zhao, W. (2024) 'XLNet-CNN-GRU dual-channel aspect-level review text sentiment classification method', *Multimedia Tools and Applications*, Vol. 83, No. 2, pp.5871–5892.

Yan, H., Yi, B., Li, H. and Wu, D. (2022) 'Sentiment knowledge-induced neural network for aspect-level sentiment analysis', *Neural Computing and Applications*, Vol. 34, No. 24, pp.22275–22286.

Zhu, X., Zhu, Y., Zhang, L. and Chen, Y. (2023) 'A BERT-based multi-semantic learning model with aspect-aware enhancement for aspect polarity classification', *Applied Intelligence*, Vol. 53, No. 4, pp.4609–4623.