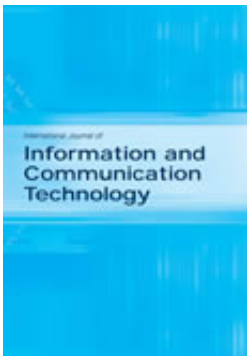
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

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

Machine learning-based multidimensional sentiment visualisation and analysis of digital media

Man Cao

DOI: 10.1504/IJICT.2025.10071633

Article History:

Received:	15 April 2025
Last revised:	28 April 2025
Accepted:	29 April 2025
Published online:	20 June 2025

Machine learning-based multidimensional sentiment visualisation and analysis of digital media

Man Cao

College of Information Engineering,
KaiFeng University,
KaiFeng 475004, China
Email: myc3066@163.com

Abstract: Digital media contains a huge amount of emotional information that needs to be mined. To solve the problem that existing models ignore the features of multi-dimensional emotional words, firstly, multi-dimensional emotional words are expanded based on improved Word2vec, and then the digital media comments are input into the pre-trained model to generate a multi-dimensional text emotional word vector. The modified term frequency-inverse document frequency (TF-IDF) method is used to obtain the representation of multidimensional emotion subject words. Then the global features are obtained by using the hybrid model of convolutional neural network (CNN) and gated recurrent unit (GRU). Multi-dimensional attention mechanism is used to interact global features and multi-dimensional emotion features, and multi-dimensional emotion classification results are output by full connection layer. The results show that the Marco-F1 of the proposed model is 91.17%, which can accurately classify the emotions of digital media.

Keywords: digital media visualisation; sentiment classification; Word2vec algorithm; TF-IDF method; multidimensional attention mechanism; convolutional neural network; CNN; gated recurrent unit; GRU.

Reference to this paper should be made as follows: Cao, M. (2025) 'Machine learning-based multidimensional sentiment visualisation and analysis of digital media', *Int. J. Information and Communication Technology*, Vol. 26, No. 21, pp.39–54.

Biographical notes: Man Cao received her Master's degree from the Henan University in 2013. She is currently a Lecturer at the College of Information Engineering, Kaifeng University. Her research interests include digital media, film and television directing, and communication.

1 Introduction

With the booming of social media, digital media content is growing at an unprecedented rate. Sentiment analysis, as an important branch in the field of natural language processing, aims at identifying, extracting, and quantifying the sentiment tendencies in texts, which is important for understanding public opinion, brand reputation management, and social event monitoring (Rodríguez-Ibánez et al., 2023). However, in the face of such a huge amount of data and complex sentiment expressions, traditional analysing methods appear to be overwhelmed. The introduction of machine learning techniques provides a

new solution for multidimensional digital media sentiment analysis (Dhaoui et al., 2017). By training models, machine learning is able to automatically learn and extract features from large amounts of text data to achieve accurate classification and quantification of emotional tendencies (Dhaoui et al., 2017). On this basis, the visual presentation of sentiment analysis results can show the sentiment distribution and change trends in an intuitive and easy-to-understand way, helping users to better understand and utilise this sentiment information (Kucher et al., 2020). Therefore, multi-dimensional digital media sentiment visualisation analysis based on machine learning has important theoretical and practical value for promoting the development of digital media field and enhancing user experience.

Most of the research on digital media sentiment visualisation and analysis involves sentiment classification of social platform texts and visualisation of the classification results. Traditional studies are based on corpus-based methods for visualising and analysing sentiment lexicons. Duric and Song (2012) visualised prefix relations of textual lexicons at the syntactic level and conducted visualisation experiments. Jindal and Aron (2021) accomplished sentiment analysis based on their own basic sentiment lexicon and knowledge corpus, and used node-link graphs to represent the relationships between semantics in unstructured text. Jain et al. (2023) analysed the sentiment analysis of text information of user comments on digital media platforms based on rule-based methods and used a tree diagram method to portray the similarity between texts, but the number of texts was too large leading to confusing results. Chu et al. (2022) used an unsupervised learning method to classify user comments into positive and negative polarities for sentiment classification and to determine the sentiment category of the text by counting words with different lexical properties. Li et al. (2020) combined the dimensional features of word vectors and plain Bayes to construct a learning model, and obtained better classification results by visualising and analysing the experiments in multiple datasets. Urologin (2018) extracted sentiment words and adverbs of degree from user comments to construct sentiment phrases, combined bag-of-words model and support vector machine algorithm to create a learning model for sentiment analysis, and performed visual analysis through SentiView. Park et al. (2017) constructed a joint model of sentiment words and topic words for sentiment analysis based on the LDA topic model by extracting sentiment words and topic words in the text, but the classification accuracy is not high.

The results obtained by relying only on the sentiment dictionary are relatively homogeneous, with a strong dependence on the dictionary and a relatively low accuracy rate. Machine learning-based sentiment visualisation and analysis methods can automatically extract features from a large amount of data and improve classification efficiency. Yu et al. (2016) used BERT to embed text word vectors and extracted text features by convolutional neural network (CNN), output the sentiment classification results by using the fully connected layer, and finally displayed them visually based on the matrix view. Tembhurne and Diwan (2021) utilised long short-term memory (LSTM) and attention mechanism can better mine the semantic information in text. A single sentiment classification can no longer meet the needs of industrial development, and multi-dimensional sentiment analysis can more comprehensively and accurately mine the user's emotional tendencies. Chang et al. (2020) achieved good results in multidimensional sentiment visualisation analysis by constructing multiple implicit layers to capture contextual semantic information of sentiment words. Hou et al. (2023) used two attention mechanisms to improve the learning efficiency of the model in order to

fully capture the contextual semantic information and better incorporate dimension words, and finally demonstrated the visualisation results of multidimensional sentiment classification via T-SNE.

According to the above analysis of the current state of research, it can be seen that the single-dimension digital media sentiment classification model cannot comprehensively and accurately mine the emotional tendency of users, resulting in unsatisfactory classification results. For this reason, this paper proposes a multi-dimensional digital media sentiment classification method based on machine learning, and carries out a multi-dimensional visualisation analysis. Firstly, based on the improved Word2vec, the multidimensional emotion words of the basic emotion dictionary are expanded to construct a multidimensional emotion dictionary for digital media, and the pre-processed digital media comment text is inputted into the pre-training model to generate a multidimensional text emotion word vector to more accurately capture the semantic relationship between words. The improved TF-IDF method is utilised to create dense clustering to obtain the characterisation of multidimensional text emotion topic words. Subsequently, CNN is used to realise the initial extraction of local semantic characteristics of the text, and gated recurrent unit (GRU) strengthens the attention to the contextual semantic information, and obtains the global semantic features through the splicing operation. The extracted multidimensional topic word representations are used as auxiliary features, the global and auxiliary features are fully semantically interacted using a multidimensional attention mechanism, and finally a fully connected level is used to project the fused characteristics into the multidimensional sentiment polarity space to obtain sentiment classification results. Visual analysis outcome implies that the classification accuracy of the proposed model is improved by 5.39%–28.64%, which provides a strong support for digital media sentiment classification.

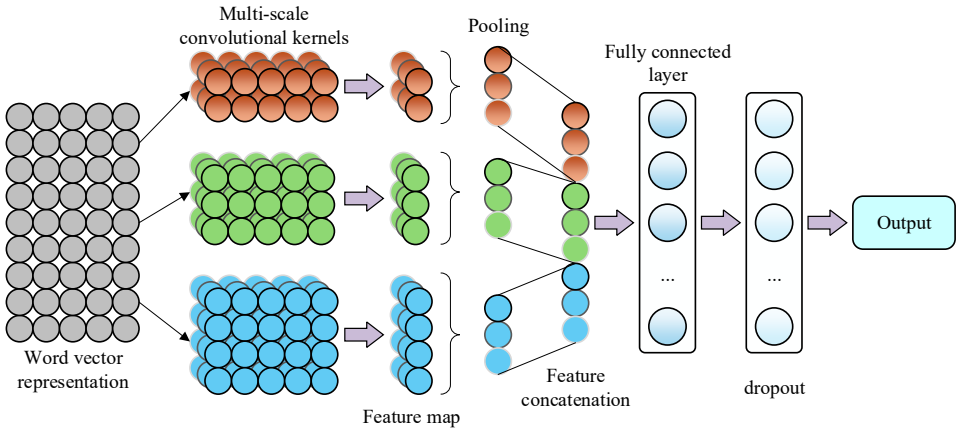
2 Relevant technologies

2.1 Convolutional neural network

CNNs can efficiently process high-dimensional data and automatically capture useful characteristics through their unique structure and convolutional operations (Kuo, 2016). Compared to neural network models such as RNN, they are characterised by high parameter efficiency, robustness and scalability. CNN consists of an input layer, a convolutional level, a pooling level, a fully-connected level, and an output level, and its structure is shown in Figure 1, with the functions of each level clearly defined and easy to expand.

Convolution and pooling are the two most important steps in CNN operation. The convolution process mainly uses the convolution kernel as a feature extractor to traverse the word vectors for feature extraction, while the pooling process mainly aims at reducing the number of features through feature selection to reduce the amount of computation and avoid overfitting. CNN has a clear advantage in text feature extraction and can greatly improve the efficiency and accuracy of sentiment classification (Liu et al., 2016).

Figure 1 The structure of CNN (see online version for colours)



2.2 Gated recurrent unit

GRU is a simplification of LSTM, the biggest difference between it and traditional neural network is that there is a temporal order relationship in GRU network, so that the text sequence can be regarded as an ordered sequence, and then carry out feature extraction. It controls the passing and cutting of information through gates, which initially achieved excellent results in statistical machine translation (Khodabandelou et al., 2020). There are many applications in dealing with serialised data, the formulas for reset gate R_t and update gate Z_t in GRU are as follows.

$$R_t = \delta\left(X_t W_{xr} + h_{t-1} W_{hr} + b_r\right) \tag{1}$$

$$Z_t = \delta\left(X_t W_{xz} + h_{t-1} W_{hz} + b_z\right) \tag{2}$$

$$\tilde{h}_t = \tanh\left(X_t W_{xh} + \left(R_t \odot h_{t-1}\right) W_{hh} + b_h\right) \tag{3}$$

$$h_t = Z_t \odot h_{t-1} + \left(1 - Z_t\right) \odot \tilde{h}_t \tag{4}$$

where X_t is the input data, W is the weight, b is the bias, h_t is the hidden state at time step t, and δ is the sigmoid function.

In this paper, BiGRU will be taken to capture the global characteristics of the text, due to the structural features of BiGRU, the inputs of the cells inside it are the inputs of the current moment and the obscured state of the previous moment, and each grid cell passes the information to the next cell and establishes the relationship between the cells before and after, so that there is a historical memory throughout the network.

2.3 TF-IDF algorithm

The TF-IDF algorithm can effectively evaluate the importance of words in a document by combining word frequency and inverse document frequency. Its simplicity, efficiency, and differentiation have made TF-IDF broadly adopted in the field of text mining and natural language processing. The principle of TF-IDF is that the more frequently the text

appears in the document, the more it represents the central idea of the whole document (Liang and Niu, 2022). The formula for the inverse document frequency IDF is as follows, where x is the amount of occurrences of a word in the text, y is the total amount of words, and df_x is the amount of documents containing a word.

$$IDF = \log\left(\frac{N}{df_x}\right) \tag{5}$$

IDF has a higher ability to recognise a term in a document if it occurs a large number of times in that document, but rarely, if ever, in other documents. The TF-IDF value is higher for the higher frequency words in a document and the lower document frequency of the word in the document set. Therefore, TF-IDF can filter out all the general words and leave only the important words.

3 Construction of a multidimensional digital media sentiment dictionary based on improved Word2vec modelling

Expanding the multidimensional emotion words in the field of digital media. The multidimensional emotion words in the basic emotion lexicon are expanded to construct a multidimensional emotion lexicon in the field of digital media, so that the resulting emotion lexicon can realise the emotion orientation in a deeper level. Word2vec models are trained with large-scale corpora to learn generic semantic representations and have strong generalisation capabilities compared to other bag-of-words models (Cahyani and Patasik, 2021). Based on such a sentiment lexicon, the sentiment tendency of a sentence can be judged more accurately. The improved Word2vec equation is as follows.

$$h = x^T W = W_{(k)} x_k = V_{wI}^T \tag{6}$$

where the h vector is computed entirely from the k^{th} row of the W matrix, and V_{WI} is the vector representation of the input word W_I. The above equation shows the relationship from the input layer to the obscured level, and from the obscured level to the output level a new $V * N$ weight matrix W' can be obtained, and using W' the score U_j can be calculated for each word as shown in equation (7).

$$U_j = V_{wj}'^T * h \tag{7}$$

In this paper, h is varied as follows when more than one word occurs in the context.

$$h = \frac{1}{C} W * (x_1 + x_2 + ... + x_C) \tag{8}$$

where C is the number of context words, the purpose is to find the mean of multiple input vectors and multiply it with the weight W from the input level to the implicit level.

The multidimensional sentiment lexicon is then extended by acquiring new words for digital media content reviews through the improved Word2vec method described above. Firstly, the basic sentiment lexicon of existing studies (Pellert et al., 2022) is organised to remove duplicate sentiment words, process and remove useless symbols using stuttering participles and third-party deactivation, and then perform lexical statistics to select nouns, verbs, adjectives and adverbs. The cleaned corpus dataset is trained by Word2vec, and

the word vectors of the words in the corpus are obtained. The word vectors are used to calculate the cosine values between the words and the positive and negative seed words, and n words are selected as candidate words in descending order. Finally, the similarity between positive candidate words and the base words such as 'happy' and 'love' in the basic dictionary is calculated by word vector, and the similarity between negative candidate words and the base words such as 'anger', 'sadness', 'fear' and 'surprise' in the basic dictionary is calculated, so as to expand the emotion words in various dimensions, thus laying the foundation for the subsequent multidimensional emotion classification.

4 Multidimensional sentiment and LDA-based text topic mining for digital media content reviews

4.1 Text multidimensional emotion word vector embedding and dimensionality reduction

To address the limitations of existing studies in semantic relationships and contextual associations, this paper inputs digital media content review texts into the improved Word2vec model to generate multidimensional document word vectors in order to more accurately capture the semantic relationships among multidimensional words. The overall process is shown in

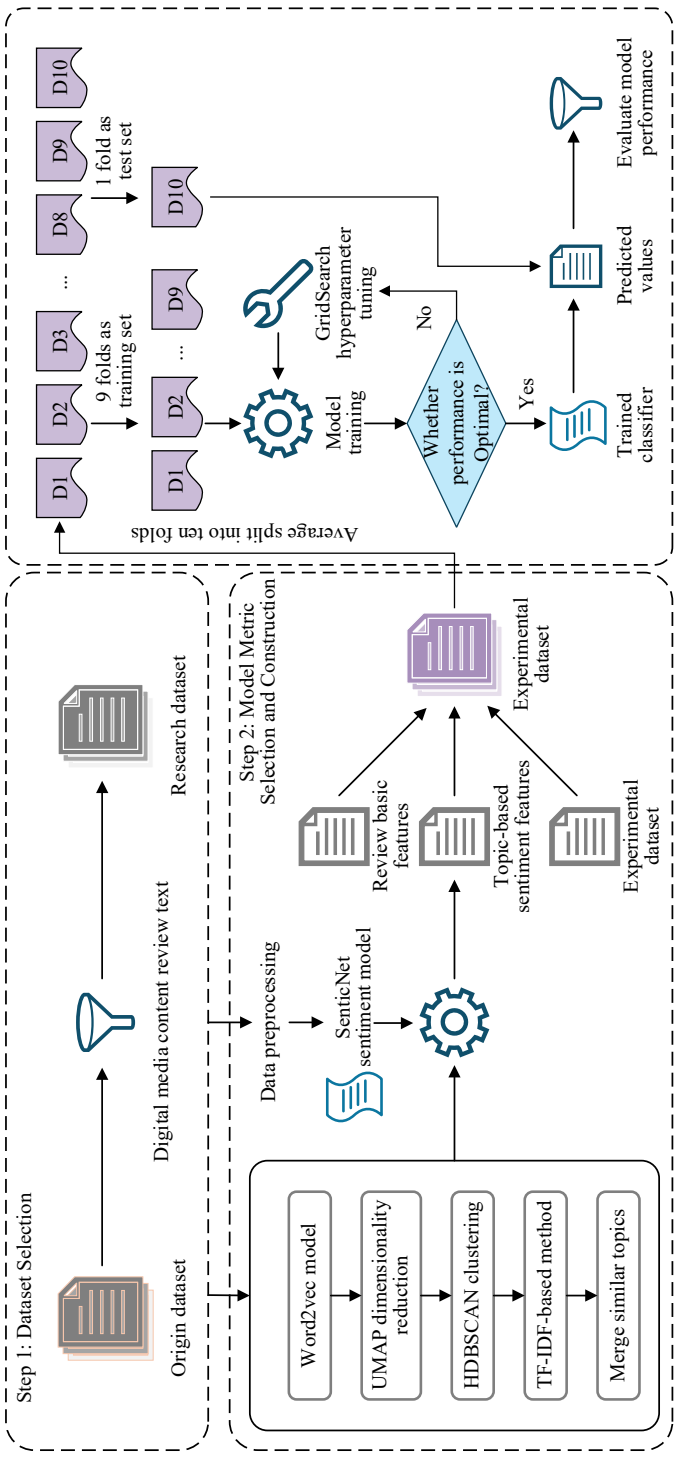
The input matrix of the model is constructed by training multidimensional sentiment word vectors with improved Word2vec. Firstly, the text data is de-duplicated to remove the words with low semantic information content; then the Jieba word separation toolkit in python is used to complete the text word separation process, and customised dictionaries are loaded through the Jieba.load_userdict function to improve the accuracy of the word separation results. The word2vec tool is then used to train the word segmentation results to get the trained word vector file, according to the word vector file, the text data will be processed into the format required by the model input.

Due to the high dimensionality of the embedding, this paper uses uniform manifold approximation and projection (UMAP) (Healy and McInnes, 2024) to downsize the data, which is faster than other downsizing techniques, and retains more of the global features of the high-dimensional data in lower projection dimensions during downsizing. For word vectors after dimensionality reduction, low-dimensional data clustering is needed to extract topics. In this paper, we use density clustering DBSACN (Hahsler et al., 2019) to cluster the word vectors and automatically generate the optimal clustering results, each category in the clusters is treated as a topic, and equation (9) is used to compute the distances between words a, b after transforming them into vectors.

$$d_{mreach-k}(a,b) = \max\left\{core_k(a), core_k(b), d(a,b)\right\} \tag{9}$$

where $d(a, b)$ is the original distance between a and b, k is the distance from the data point to the k^{th} nearest point.

Figure 2 Multidimensional emotion theme word mining process (see online version for colours)



4.2 Improved TF-IDF-based feature representation for text topic words

For text modelling for each cluster after clustering, a topic is assigned to each cluster and the distribution of words in the cluster generalises a topic. Therefore, to gain a more accurate topic representation, the number of redundant clustered topics is reduced by class-based TF-IDF to express the importance of multidimensional sentiment words to the topic. From Section 2.3, in the TF-IDF process, TF is the word frequency and IDF is the indicated inverse text frequency, $TF - IDF = TF * IDF$, where $TF = x/y$, $IDF = \log(N/df_x)$, and x are the amount of occurrences of multidimensional sentiment words in the document. Extending the above process to the class level, all documents in the cluster are treated as individual documents, as shown below.

$$W_x = \left\| tf_x \right\| \cdot \log\left(1 - \frac{A}{tf_x}\right) \tag{10}$$

where $\|tf_{x,c}\|$ is the frequency of word x in cluster A. By calculating the TF-IDF values of words under each topic category, similar topics are merged based on the cosine similarity calculation of TF-IDF vectors between topics, as shown below, to eliminate redundant clustered topic representations.

$$similarity = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}} \tag{11}$$

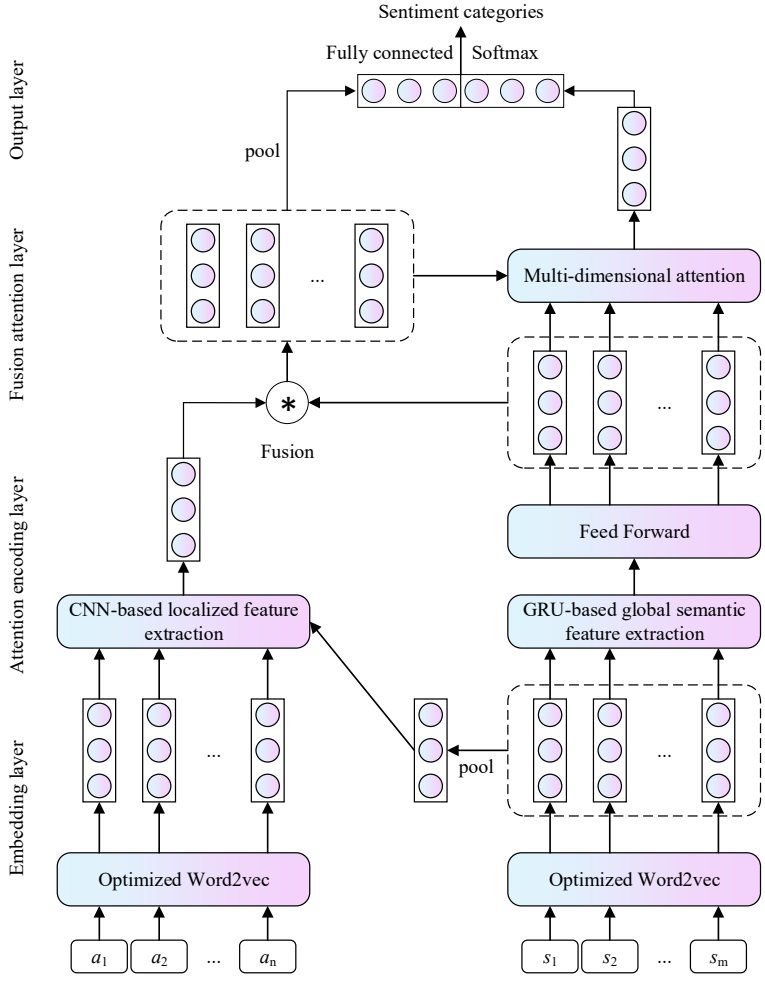
Themes are represented by a number of words that are most representative of the topic. Multidimensional sentiment subject terms are used as auxiliary feature inputs for subsequent models, thus reducing the redundancy of model inputs and improving computational efficiency.

5 Machine learning-based sentiment classification for multidimensional digital media

5.1 Comment text feature extraction based on CNN and BiGRU

To address the issue of incomplete feature extraction of single-dimension sentiment classification methods, this paper firstly utilises CNN to realise the preliminary extraction of local semantic features of digital media comment text, and GRU strengthens the attention to the contextual semantic information, and obtains the in-depth semantic fusion representation through the splicing operation. Subsequently, the extracted multidimensional topic word representations are used as auxiliary features, and the global and auxiliary features are fully semantically interacted using the multidimensional attention mechanism to enhance the critical features, and finally a fully connected level is used to project the fused characteristics into the multidimensional sentiment polarity space to obtain multidimensional digital media sentiment classification results. The proposed classification model is shown in Figure 3.

Figure 3 The proposed multidimensional digital media sentiment classification model (see online version for colours)



CNN initially obtains the local characteristic vector of the text through the convolution operation. Assuming that the sentence S contains A total of s terms after pre-processing and the dimension of the word vector is e, the input layer needs to convert S into a sentence vector matrix of $e \times s$ and input it into the convolution layer according to the improved Word2vec. The convolution layer performs convolution operations on input vectors by means of a convolution kernel of size $e * h$. Assume that W_h is the weight matrix corresponding to different sizes of convolution kernels, b is the bias term, $\oplus$ is the convolution operator, $f(\cdot)$ is the activation function, and the result of each convolution operation is C_i.

$$C_i = f\left(W_h \otimes X_{i:i+h-1} + b\right) \tag{12}$$

Each convolutional kernel traverses the input vector matrix from top to bottom to obtain a feature map $M = [m_0, m_1, m_2, \ldots, m_{s-h}]$ containing $s - h + 1$ feature vectors, and the merge level joins the output characteristic maps from the convolutional levels to form the final localised feature vector C. The merge level is then used as a base for the feature maps of the convolutional levels.

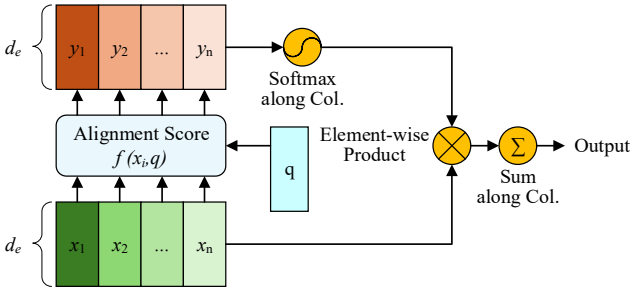
After obtaining C, this paper captures the semantic features of the comment text by BiGRU, which has a more concise network structure compared to LSTM network and can improve the model training efficiency. BiGRU is calculated as shown in equation (13)–equation (15), where the specific calculation of h_t is shown in Section 2.2.

$$\vec{h}_t = GRU\left(x_t, \vec{h}_{t-1}\right) \tag{13}$$

$$\overleftarrow{h}_t = GRU\left(x_t, \overleftarrow{h}_{t+1}\right) \tag{14}$$

$$H_t = \left[\vec{h}_t, \overleftarrow{h}_t\right] \tag{15}$$

Figure 4 The structure of the multidimensional attention mechanism (see online version for colours)



5.2 *Text feature fusion based on multidimensional attention mechanism*

In this paper, the generated feature vector C of the CNN channel and the generated feature vector H of the BiGRU channel are connected to obtain the global feature vector K of the model, as shown in equation (16). Subsequently, F and the topic representation f of the text are subjected to attention computation, which accomplishes the interaction between the context and the topic words in the model, and generates the final representation with more interactive sensory information.

$$K = C \oplus H \tag{16}$$

The attention computation in this layer uses the multidimensional attention mechanism, which is described more in the literature (Li and Huang, 2023). The structure of the multidimensional attention mechanism is shown in Figure 4. First, the importance level α_i of the i^{th} word in the contextual representation of the sentence is calculated according to equation (17) and equation (18).

$$F_s\left(K_i, f_i^s\right) = W^T \tanh\left(\left[K_i, f_i^s\right] \cdot W_1\right) \tag{17}$$

$$\alpha_i = \frac{\exp\left(F_s\left(K_i, f_i^s\right)\right)}{\sum_{j=1}^{m} \exp\left(F_s\left(K_j, f_j^s\right)\right)} \tag{18}$$

Then, the final multidimensional text feature vector representation R is obtained by element-by-element weighted summation of the sentence context representation h_i^s using the obtained weight matrix $\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ and equation (19).

$$R = \sum_{i=1}^{m} \alpha_i \odot h_i^s \tag{19}$$

5.3 Multidimensional digital media sentiment categorisation

After the multidimensional attention level, this paper obtains the deep semantic fusion representation K and the final representation R of the context after completing the attention interaction. First, an average pooling operation is performed on the fusion representation K, and it is connected to the final representation R of the context to gain the final characteristic representation R_{final} that will be sent to the output level for classification.

$$K_{avg} = \sum_{i=1}^{m} K_i / m \tag{20}$$

$$R_{final} = \left[K_{avg}, R\right] \tag{21}$$

The final representation R_{final} is then projected into the multidimensional affective polarity space using a fully connected level as follows.

$$x = W^T R f_{inal} + b \tag{22}$$

Finally, a softmax function is adopted to compute the probability distribution of multidimensional sentiment polarity, and the category with the highest probability is the final predicted sentiment category.

$$\hat{y} = \frac{\exp(x)}{\sum_{k=1}^{C} \exp(x)} \tag{23}$$

The model uses cross entropy loss (CE) (Zhou et al., 2019) combined with L2 regular term as the loss function as shown below.

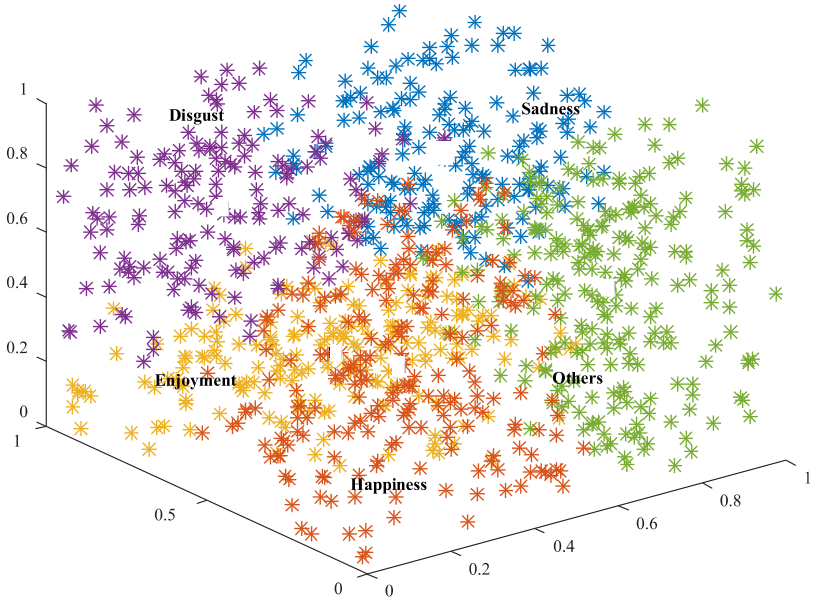
$$\mathcal{L}(\theta) = -\sum_{i=1}^{C} y_i \log \hat{y}_i + \lambda \|\theta\|^2 \tag{24}$$

where C is the number of polarity for sentiment classification, y_i is the current true sentiment polarity, $\hat{y}_i$ is the probability of occurrence of the current predicted sentiment polarity, and the sum of the predicted probabilities of all sentiment polarities is 1, λ is the L2 regularisation coefficients, and θ is a set of parameter matrices.

6 Experimental results and analyses

The experimental dataset is a textual dataset of reviews of digital menus and dish contents in online dining platforms provided by the AI challenger competition 2018; it contains a training set of 15,000 labelled reviews and a validation set of 1,500 labelled reviews. This dataset is for fine-grained sentiment analysis and contains five fine-grained sentiment tendencies disgust, enjoyment, happiness, sadness, and others. The batch_size of the model training is 16, the epoch is 50, the learning rate is 0.001, and the dropout is set to 0.5 in order to prevent overfitting, and the parameters are updated by back propagation, and the training is optimised by using the Adam optimiser.

Figure 5 User clustering results of CNN-BiGRU-MAT (see online version for colours)

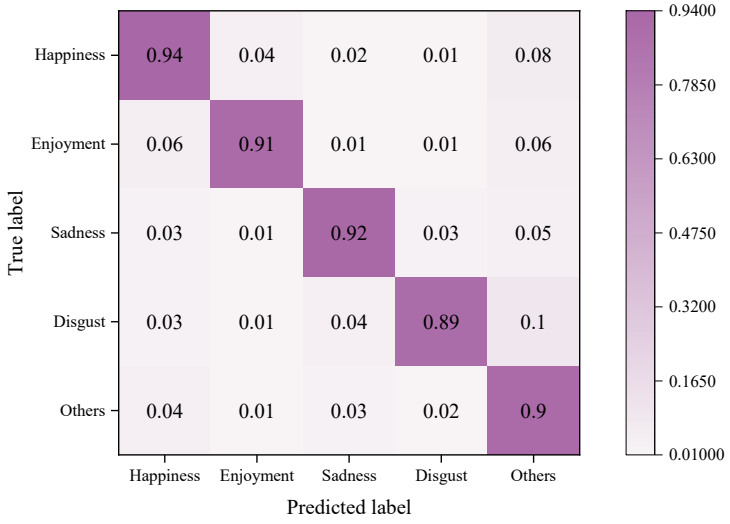


In this paper, the experiments are conducted under Windows 10 64-bit operating system, running on a desktop computer with 16 G of RAM, Intel (R) Core (TM) i7-4770K CPU @ 3.50 GHz, and NVIDIA GeForce GTX 780 Ti graphics card. In the process of experiment, the front end of the system uses pure JavaScript diagram class library Echarts.js and D3.js to design and realise the visualisation effect and interactive function. Visual analysis results of multi-dimensional emotion classification of the proposed model CNN-BiGRU-MAT are shown in Figure 5. Figure 5 demonstrates the user clustering results of CNN-BiGRU-MAT in 3-dimensional space based on multidimensional sentiment categories. The clustering results show that users who show the same emotion towards digital media content are assigned to the same linear space, indicating that our model can well support the application of user clustering, and that different users have different emotional tendencies towards digital media content. The analysed results can identify users in multi-dimensional emotional tendency categories and provide users with better services and experiences, so it has high practical value and application prospects.

At the same time, 20% of the data in the data set are randomly chosen as the test set, and the data results are presented in the form of confusion matrix, as shown in Figure 6.

From the meaning of the confusion matrix, we can see that the data on the diagonal is the numerator value of the accuracy rate, from which we calculate the classification accuracy rate in the confusion matrix of CNN-BiGRU-MAT. The classification recognition accuracies of CNN-BiGRU-MAT for disgust, enjoyment, happiness, sadness, and others emotions are 0.94, 0.91, 0.92, 0.89 and 0.9, respectively, with an average classification accuracy of 91.2%. This indicates that CNN-BiGRU-MAT not only uses CNN-BiGRU to capture local characteristics and global semantic characteristics of the comment text, but also extracts multidimensional sentiment topic words as auxiliary features by the improved TF-IDF, which greatly enriches the multidimensional sentiment feature representation and improves the classification accuracy.

Figure 6 Visual analysis results of CNN-BiGRU-MAT model (see online version for colours)



To verify the classification performance of the proposed models, this paper uses accuracy, recall, and Marco-F1 values to compare the Sent-SVM model in literature (Urologin, 2018), WDLDA model in literature (Park et al., 2017), BERT-CNN model in literature (Yu et al., 2016), LSTM-AM model in literature (Tembhurne and Diwan, 2021), MSTAM in literature (Hou et al., 2023) model and the suggested CNN-BiGRU-MAT model for comparison experiments. Comparison of classification performance metrics for different models is shown in Table 1. CNN-BiGRU-MAT has a large improvement in all the metrics compared to the other models, and its accuracy and recall are 0.9324 and 0.9015, respectively, which are improved by 5.39%–28.64% compared to the other five models. Comparing Marco-F1 again, CNN-BiGRU-MAT shows excellent sentiment classification by 17.51%, 15.19%, 10.66%, 9.11%, and 5.09% compared to Sent-SVM, WDLDA, BERT-CNN, LSTM-AM, and MSTAM, respectively.

Although Sent-SVM considers multidimensional sentiment words, the SVM algorithm needs to solve the support vectors with the help of quadratic programming during training, which involves a large number of matrix computations and reduces the classification efficiency. WDLDA only considers the features of the subject words but not the global semantic features of the text, which leads to poor classification accuracy. BERT-CNN is used for text embedding and feature extraction by BERT and CNN

respectively without considering multidimensional sentiment features. Although LSTM-AM considers the global semantic characteristics of the text, there is no multi-dimensional expansion for the emotion words in the text, so the classification performance is not as good as MSTAM. Therefore, CNN-BiGRU-MAT, which incorporates multidimensional sentiment words and text global features, achieves the best classification results.

Table 1 Comparison of classification performance metrics of different models

Model	*Accuracy*	*Recall*	*Marco-F1*
Sent-SVM	0.7248	0.7569	0.7366
WDLDA	0.7625	0.7581	0.7598
BERT-CNN	0.8218	0.8169	0.8051
LSTM-AM	0.8463	0.8271	0.8206
MSTAM	0.8847	0.8529	0.8608
CNN-BiGRU-MAT	0.9324	0.9015	0.9117

7 Conclusions

According to the above analysis of the current state of research, it is known that the single-dimension digital media emotion classification model cannot comprehensively and accurately mine the emotional tendency of users, resulting in unsatisfactory classification results. Therefore, to address the above issues, this paper proposes a multi-dimensional digital media sentiment classification method based on machine learning, and carries out a multi-dimensional visualisation analysis. Firstly, based on the improved Word2vec, the multidimensional emotion words of the basic emotion dictionary are expanded, and the multidimensional emotion dictionary of digital media is constructed, and the text of the digital media commentaries is inputted into the improved Word2vec model to generate the multidimensional text emotion word vectors. The improved TF-IDF method is utilised in order to create dense clustering to obtain the representation of multidimensional text sentiment topic words. Subsequently, CNN is utilised to achieve the initial extraction of local semantic features of digital media review texts, and GRU enhances the attention to the contextual semantic information to obtain the global semantic fusion representation through the splicing operation. The extracted multidimensional topic word representations are used as auxiliary features, the global features and auxiliary features are fully semantically interacted using the multidimensional attention mechanism, and finally the fused features are projected to the multidimensional sentiment polarity space using the full connectivity layer to obtain the multidimensional digital media sentiment classification results. Visual analysis outcome implies that the proposed model significantly improves the classification accuracy and provides a new solution idea for digital media sentiment classification.

There is still some room for improvement in this study, the number of neural network layers of the offered model needs to be further increased, and there is still room for optimisation in the accuracy of the classification results due to the limited training data of the model.

Declarations

All authors declare that they have no conflicts of interest.

References

Cahyani, D.E. and Patasik, I. (2021) 'Performance comparison of TF-IDF and word2vec models for emotion text classification', *Bulletin of Electrical Engineering and Informatics*, Vol. 10, No. 5, pp.2780–2788.

Chang, Y-C., Ku, C-H. and Chen, C-H. (2020) 'Using deep learning and visual analytics to explore hotel reviews and responses', *Tourism Management*, Vol. 80, p.104129.

Chu, K.E., Keikhosrokiani, P. and Asl, M.P. (2022) 'A topic modeling and sentiment analysis model for detection and visualization of themes in literary texts', *Pertanika Journal of Science & Technology*, Vol. 30, No. 4, pp.2535–2561.

Dhaoui, C., Webster, C.M. and Tan, L.P. (2017) 'Social media sentiment analysis: lexicon versus machine learning', *Journal of Consumer Marketing*, Vol. 34, No. 6, pp.480–488.

Duric, A. and Song, F. (2012) 'Feature selection for sentiment analysis based on content and syntax models', *Decision Support Systems*, Vol. 53, No. 4, pp.704–711.

Hahsler, M., Piekenbrock, M. and Doran, D. (2019) 'dbscan: fast density-based clustering with R', *Journal of Statistical Software*, Vol. 91, pp.1–30.

Healy, J. and McInnes, L. (2024) 'Uniform manifold approximation and projection', *Nature Reviews Methods Primers*, Vol. 4, No. 1, p.82.

Hou, S., Tuerhong, G. and Wushouer, M. (2023) 'Visdanet: visual distillation and attention network for multimodal sentiment classification', *Sensors*, Vol. 23, No. 2, p.661.

Jain, R., Kumar, A., Nayyar, A., Dewan, K., Garg, R., Raman, S. and Ganguly, S. (2023) 'Explaining sentiment analysis results on social media texts through visualization', *Multimedia Tools and Applications*, Vol. 82, No. 15, pp.22613–22629.

Jindal, K. and Aron, R. (2021) 'A novel visual-textual sentiment analysis framework for social media data', *Cognitive Computation*, Vol. 13, pp.1433–1450.

Khodabandelou, G., Jung, P-G., Amirat, Y. and Mohammed, S. (2020) 'Attention-based gated recurrent unit for gesture recognition', *IEEE Transactions on Automation Science and Engineering*, Vol. 18, No. 2, pp.495–507.

Kucher, K., Martins, R.M., Paradis, C. and Kerren, A. (2020) 'StanceVis Prime: visual analysis of sentiment and stance in social media texts', *Journal of Visualization*, Vol. 23, No. 6, pp.1015–1034.

Kuo, C-C.J. (2016) 'Understanding convolutional neural networks with a mathematical model', *Journal of Visual Communication and Image Representation*, Vol. 41, pp.406–413.

Li, H. and Huang, Q. (2023) 'MAF-Net: multidimensional attention fusion network for multichannel speech separation', *Multimedia Systems*, Vol. 29, No. 6, pp.3703–3720.

Li, Z., Li, R. and Jin, G. (2020) 'Sentiment analysis of danmaku videos based on naïve Bayes and sentiment dictionary', *IEEE Access*, Vol. 8, pp.75073–75084.

Liang, M. and Niu, T. (2022) 'Research on text classification techniques based on improved TF-IDF algorithm and LSTM inputs', *Procedia Computer Science*, Vol. 208, pp.460–470.

Liu, M., Shi, J., Li, Z., Li, C., Zhu, J. and Liu, S. (2016) 'Towards better analysis of deep convolutional neural networks', *IEEE Transactions on Visualization and Computer Graphics*, Vol. 23, No. 1, pp.91–100.

Park, D., Kim, S., Lee, J., Choo, J., Diakopoulos, N. and Elmqvist, N. (2017) 'ConceptVector: text visual analytics via interactive lexicon building using word embedding', *IEEE Transactions on Visualization and Computer Graphics*, Vol. 24, No. 1, pp.361–370.

Pellert, M., Metzler, H., Matzenberger, M. and Garcia, D. (2022) 'Validating daily social media macroscopes of emotions', *Scientific Reports*, Vol. 12, No. 1, p.11236.

Rodríguez-Ibánez, M., Casánez-Ventura, A., Castejón-Mateos, F. and Cuenca-Jiménez, P-M. (2023) 'A review on sentiment analysis from social media platforms', *Expert Systems with Applications*, Vol. 223, p.119862.

Tembhurne, J.V. and Diwan, T. (2021) 'Sentiment analysis in textual, visual and multimodal inputs using recurrent neural networks', *Multimedia Tools and Applications*, Vol. 80, No. 5, pp.6871–6910.

Urologin, S. (2018) 'Sentiment analysis, visualization and classification of summarized news articles: a novel approach', *International Journal of Advanced Computer Science and Applications*, Vol. 9, No. 8.

Yu, Y., Lin, H., Meng, J. and Zhao, Z. (2016) 'Visual and textual sentiment analysis of a microblog using deep convolutional neural networks', *Algorithms*, Vol. 9, No. 2, p.41.

Zhou, Y., Wang, X., Zhang, M., Zhu, J., Zheng, R. and Wu, Q. (2019) 'MPCE: a maximum probability based cross entropy loss function for neural network classification', *IEEE Access*, Vol. 7, pp.146331–146341.