ITC 4/52
Information Technology and Control
Vol. 52 / No. 4 / 2023
pp. 867-877
DOI 10.5755/j01.itc.52.4.34079

MPCM: Multi-modal User Portrait Classification Model Based on Collaborative Learning

Received 2023/05/10 | Accepted after revision 2023/07/17

HOW TO CITE: Liu, J., Li, L. (2023). MPCM: Multi-modal User Portrait Classification Model Based on Collaborative Learning. *Information Technology and Control*, 52(4), 867-877. https://doi.org/10.5755/j01.itc.52.4.34079

MPCM: Multi-modal User Portrait Classification Model Based on Collaborative Learning

Jinhang Liu
School of Computer Science, Hubei University of Technology, Wuhan, 430068, China

Lin Li
School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan 430070, China

Corresponding author: liujinhang831209@126.com

A social-media user portrait is an important means of improving the quality of an Internet information service. Current user profiling methods do not discriminate the emotional differences of users of different genders and ages on social media against a background of multi-modality and a lack of domain sentiment labels. This paper adopts the sentiment analysis of images and text to improve label classification, incorporating gender and age differences in the sentiment analysis of multi-modal social-media user profiles. In the absence of domain sentiment labels, instance transfer learning technology is used to express the learning method with the sentiment of text and images; the semantic association learning of multi-modal data of graphics and text is realized; and a multi-modal attention mechanism is introduced to establish the hidden image and text. Alignment relationships are used to address the semantic and modal gaps between modalities. A multi-modal user portrait label classification model (MPCM) is constructed. In an analysis of the sentiment data of User users on Facebook, Twitter, and News, the MPCM method is compared with the naive Bayes, Latent Dirichlet distribution, Tweet-LDA and LUBD-CM(3) methods in terms of accuracy, precision, recall and the FL-score. At a 95% confidence, the performance is improved by 5.6% to 8.9% by using the MPCM method.

KEYWORDS: Data mining, user portraits, sentiment analysis, multi-modal data, attribute label classification.

1. Introduction

With the rapid development of artificial intelligence and big data technology, there has been enormous growth in the volume of rich data, such as content data, behavioral data, and social data, generated in the form of multi-modal text, images, audio, and video [14]. Research on multi-modal social-media user portraits integrated with sentiment analysis is important to precision marketing, intelligent recommendation, information retrieval, and other services and it is an inevitable requirement for improving the quality and level of consumers. There are three types of sentiment analysis, namely text sentiment analysis, image sentiment analysis, and multi-modal sentiment analysis. Text sentiment analysis removes invalid text data through methods such as word segmentation, and stop word removal; uses bag of words, term frequency–inverse document frequency, and word2vec methods to represent the text in a language that can be processed by a computer; and constructs a sentiment classification model for the text [8]. The sentiment classification indicates the judgment of emotional polarity. Image sentiment analysis classifies sentiment polarity through image per-processing, sentiment feature extraction, and sentiment classification modelling. First, image normalization is per-processed through image cutting and enhancement, Then allowing a more comprehensive model to be extracted to obtain image emotional information, and then more effective analysis of emotional features. The reason why different modalities are divided is firstly because the information that can be accessed in different scenarios is different, and secondly, the information provided by different modalities is often different, and the most important thing is the processing and modeling that need to be adopted for the information of different modalities The way is also different. In simple cases, we can get a judgment on emotional attitude through only a single modality, such as a piece of evaluation text, a recording of a conversation, a piece of commentary video, etc. Naturally, we can also combine data from multiple modalities and model them uniformly.

At present, the user portrait derived from emotion analysis has the following main shortcomings: The description of user attributes is not in-depth. At present, user portrait technology mainly mines the user's basic information, such as age, gender, living area, and education level. However, such information can be obtained when the user registers, and the user at-tributes obtained through the analysis of such information are too simple to determine the user's interests and behaviours; User attributes are not updated with time. A user's interests and hobbies change over time, and previously mined attributes may no longer be applicable. Because the interests of users evolve over time in the widespread use of similar network terms, The semantic similarity between users' profile and words cannot be effectively measured; The user attribute description is not comprehensive. Social media to the diversity of content such as pictures, audio, video only use this information to the user essay analyses portrait will not be able to fully reflect the characteristics of the user [9].

Sentiment analysis is an important branch of natural language processing conducted to process social media and mine its emotional colour. In recent years, research on sentiment analysis [6] has attracted wide attention. By mining the views and tendencies of users of social media, we can judge fashion trends and hot spots, which helps enterprises analyse consumers' purchasing tendencies, carry out precise marketing, and improve the quality and level of the users' consumption. The use of a single-mode user profiling such as sentiment analysis means that only one mode is used to analyse and predict user attributes. Different data sources have different analysis methods. Faradic et al. [5] studied the gender, age, and personality of users using information such as text, images, and relationships of multi-modal user portraits. Faradic et al. [4] perform a comparative analysis of state-of-the-art computational personality recognition methods on a varied set of social media ground truth data from Facebook, Twitter and YouTube, and achieved the decay in accuracy when porting models trained in one social media environment to another Takahashi. Zhu et al. [15] present a Knowledge-Driven Location Privacy Preserving (KD-LPP) scheme, in order to mine user preferences and provide customized location privacy protection for users. Firstly, the UBPG algorithm is proposed to mine the basic portrait. User familiarity and user curiosity are modelled to generate psychological portrait. Then, the location transfer matrix based on the user portrait is built to transfer the real location to an anonymous lo-

cation. In order to achieve customized privacy protection, the amount of privacy is modelled to quantize the demand of privacy protection of target user. Finally, experimental evaluation on two real datasets illustrates that our KD-LPP scheme can not only protect user privacy, but also achieve better accuracy of privacy protection. Nguyen et al. [12] studied the relationship between stock prices and sentiment. Adopting Latent Dirichlet Allocation (LDA), they proposed the Topic Sentiment Latent Dirichlet Allocation model, which captures topic information and emotional information at the same time.

In existing text-based unimodal user portraits, most of the extracted features are lexical, syntactic, and word embedding features [18], [2], [7]. There are differences in the emotional expression of users of different genders on social networks, but few researchers combine user portraits with sentiment analysis. In the field of social-media user portraits, most existing multi-modal data fusion methods adopt techniques such as feature splicing and taking the weighted average of the prediction results of multiple models. In this paper, we extract an emotional feature representation to improve the effect of user portrait label classification. We plan to design a feature fusion method based on the alignment of multi-modal data, whereby the relationship between different modal data are learned and the feature representation extracted from text representation and image representation is enriched, to improve the effect of user portrait label classification.

This paper designs a collaborative representation and bridge network to integrate text and image modalities. The main contributions of the paper are as follows. A method of extracting key information based on the self-attention mechanism is proposed to bridge the semantic gap. The paper studies a multi-modal deep learning classification model based on co-learning, evaluates the performance of the model with age, gender, and region as the main classification labels in the existing data set, and discusses the general performance of the model extended to other labels.

In recent years, the application and research of deep learning theory in the field of speech emotion recognition is a very hot research direction. Deng et al. used sparse auto encoder model to carry out feature transfer learning of speech signals, and experimentally verified it on several speech emotion databases, achieving good results [3]. Zheng-Wei et al. used a model composed of sparse auto-encoder and sparsely restricted Boltzmann phase to conduct speech emotion recognition, and verified on interface data set that larger speech Windows and more hidden nodes can obtain better recognition performance [11]. Albornoz et al. used restricted Boltzmann machine and deep belief network to establish a speech emotion recognition model. Aiming at the speaker-independent experimental mode, the recognition accuracy rate was 8.67% above the baseline [1]. Mao et al. proposed a feature extraction method of salient discriminate feature analysis, combined with convolutional neural network model for speech emotion recognition, and achieved satisfactory results [16]. Huang Chen et al. proposed to use deep belief network to automatically extract speech emotion features, and use support vector machine classifier for recognition and classification, and achieved good results [10]. Zheng et al. proposed to use deep convolutional neural network model to achieve automatic extraction for emotional feature extraction of speech signals, which has achieved better classification effect compared with hand-crafted features in emotion recognition [13]. George Trigeorgis et al. [17] proposed an end-to-end speech emotion recognition method, which integrates convolutional neural network and Long Short Term memory network, can make use of contextual emotion information, and can directly use the original speech as input to automatically extract features.

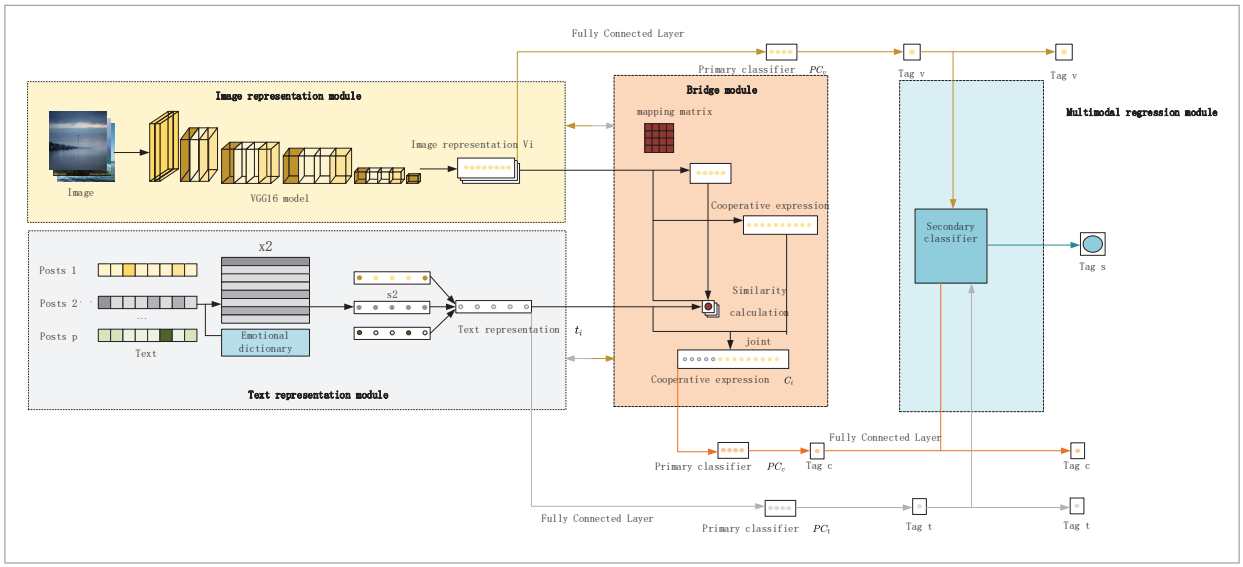
This paper describes how to leverage sentiment analysis in user profiling. The main technologies include image and text sentiment representation learning, multi-modal data machine learning, and deep learning to predict the attributes of social media users.

2. Multi-modal Social Media User Profiling Enhanced by Sentiment Analysis

A multi-modal deep-learning classification model based on co-learning is used to evaluate the performance of the model with age, gender, and region as the main classification labels in the existing data-set.

The framework of the multi-modal user portrait classification model based on collaborative learning proposed in this paper is shown in Figure 1.

Figure 1
Multi-modal user portrait classification model based on collaborative learning



The input of the text representation module is the textual data of the user, and the output is the textual representation of the user. The text classification model comprises word-level encoding and sentence-level encoding. This paper extracts the intermediate layers in sentence-level encoders as text representations. The text representation of this module is followed by a fully connected layer to obtain a text uni-modal gender classification model, which can still classify attribute labels when there is no image modality during testing.

When the above model is trained, the text representation and image representation are mapped to a common semantic space, and aligned and fused representations of the two modalities of text and image are obtained. The aligned and fused representations propagate the loss to the text representation and image representation, respectively, during back-propagation. The rich information contained in the text is transferred to the image representation through the bridge of fusion representation, and at the same time, the information contained in the image is transferred to the text representation through the bridge. In this manner, the data information of different modalities complement each other, and the feature representation ability of the text representation and image representation are improved. While improving the effect of multi-modal gender classification, the effect of single-modal classification will be improved.

2.1. Multi-modal Alignment and Fusion Based on Collaborative Representation Learning

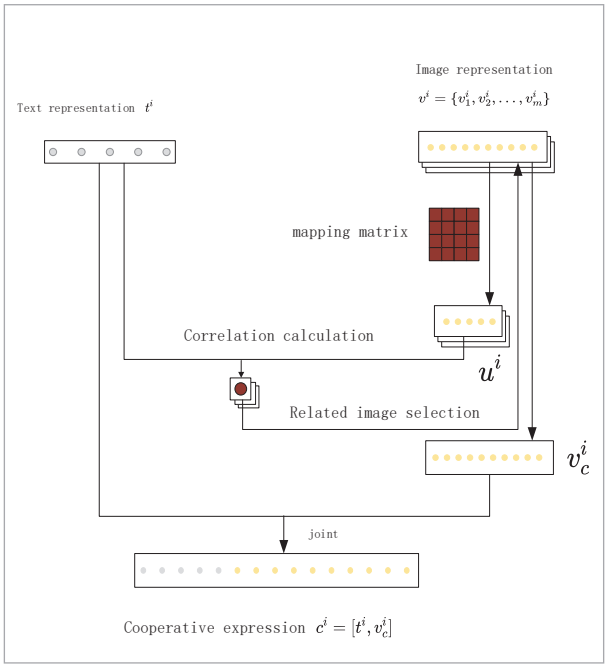
In text and image data representation (including sentiment representation), the multi-modal classification module uses the collaborative representation method to fuse the data of different modalities. Assuming that the output of the text representation module is and the number of pictures provided by each user is, the output of the image representation module is. Similarly, each picture has a corresponding image representation. To fuse text modalities and image modalities, in this paper, we chose a mapping matrix to train, as shown in Figure 2.

In Figure 2, each image representation is linearly transformed by multiplication with a mapping matrix, and the similarity is then computed with the text representation. The obtained similarity is fed back to the original image representation, and the original image representation is weighted and averaged. The weighted average image representation is stitched with the image representation into a synergistic representation. The specific formulas of the process are presented as Equations (3)-(6).

$$u_i = \tanh(W_m f_i + b_m) \tag{3}$$

$$\alpha_i = \frac{\exp(t^{\cdot} u_i)}{\sum_n \exp(t^{\cdot} u_i)} \tag{4}$$

Figure 2
Multi-modal alignment and fusion based on collaborative representation



$$f = \sum_n \alpha_i f_i \tag{5}$$

$$c = [t, f] \tag{6}$$

In the above formula, W_m is the mapping matrix and b_m is the offset, which f_i is obtained after linear transformation of u_i. Each linearly transformed image representation f_i will calculate the similarity with the image representation, and apply the softmax function for normalization, and the obtained normalized similarity is expressed as $\alpha_i, i = 1, 2, ..., n$. The similarity is taken as the importance of each image and returned to the weighted average of the original image representation f_i, which can be spliced with the text representation to obtain a collaborative representation. This paper informs on mapping matrices to align image representations into the text representation space, and in other cases, text representations into the image representation space, or both into a common space.

2.2. Multi-modal Label Classification Based on Bridge Networks

While studying multi-modal fusion, we hope to transfer knowledge from one modality to another, extract richer image representations, and improve multi-modal fusion gender classification models.

Adopting collaborative representation learning, a bridge of multi-modal fusion representation is built between the text representation module and the image representation module. By training the mapping matrix, the text representation and the image representation are mapped into a common semantic space, and the connection between different modalities is established. During back-propagation, the loss of the common semantic space is fed back to the text representation module and image representation module, so as to migrate the information contained in the text representation into the image representation. At the same time, unique information in the image representation can be transferred to the text representation. This method improves the feature representation capability of the text representation module and image representation module, extracts richer representations, and improves the effectiveness of the multi-modal gender classification model.

In this paper, there are three outputs of the first-level classifier of the multi-modal user portrait classification model based on collaborative learning, allowing multi-task joint learning. The outputs of the classification model are the output of the text representation part, the output of the image representation part, and the output of the multi-modal fusion part. During the training process, the loss functions generated by these three outputs are fed back to the entire model, and the text representation, image representation, and multi-modal fusion representation affect each other, relate to each other, and promote each other. During testing, the three outputs have their own duties without affecting each other. This not only improves the ability of the image part to extract information but also improves the ability of the output of the text representation part. At the same time, it meets the needs of not only multi-modal gender prediction but also single-modal gender prediction.

2.3. Algorithm Description

Owing to the inconsistent distribution of data in different modalities, the initialization of the parameters of the text representation part of the model comes from the parameters of the text single modality. Therefore, before training the model, it is necessary to train the text single-modality model for the initialization of the text representation part of the model.

The text uni-modal model training algorithm is presented as Algorithm 1.

Algorithm 1. Text Single-modal Model Training Algorithm

Input: text information $T=\{T^1,T^2,...,T^N\}$,
user portrait label $\hat{y}=\{y^1,y^2,...,y^N\}$,
text sentiment representation D;
Output: predict label $y_t=\{y_t^1,y_t^2,...,y_t^N\}$,
Repeat:
For $i \leftarrow 1$ *to* N
1. Connect the user's Weibo to a virtual document as text information, $T^i=\{w_{11}^i,w_{12}^i,...,w_{1T_1}^i,w_{21}^i,w_{22}^i,...,w_{2T_2}^i,...,w_{L1}^i,w_{L2}^i,...,w_{LT_L}^i\}$;
2. Perform word-level encoding, $T_w^i=W_w(T^i,D)=\{s_1^i,s_2^i,...,s_L^i\}$;
3. Perform sentence-level encoding to get the textual
representation of user i, $t^i=W_s(T_w^i)$;
4. Predict user attribute labels, $y_t^i=Dense_t(t)$, through the fully connected layer;
5. Calculate the loss function, $Loss=\sum_{i=1}^{N}e(y_t^i,\hat{y}^i)$;
6. Back-propagate the update parameters
Until: $Loss$ minimum

After training the text bimodal model, the parameters of the text uni-modal model are saved and used for the initialization of the text representation part of the model. The multi-modal label classification model is divided into a primary classifier and secondary classifier. The model training algorithm is presented as Algorithm 2.

Algorithm 2. Multi-modal Model Training Algorithm

Input: text information $T=\{T^1,T^2,...,T^N\}$,
image information $F=\{F^1,F^2,...,F^N\}$,
user portrait labels $\hat{y}=\{y^1,y^2,...,y^N\}$,
text and image sentiment D
Output: text representing part of the predicted label $y_t=\{y_t^1,y_t^2,...,y_t^N\}$,
images representing partially predicted labels $y_f=\{y_f^1,y_f^2,...,y_f^N\}$,
integration of partially predicted labels $y_s=\{y_s^1,y_s^2,...,y_s^N\}$
Initialization: Word-level encoding W_w and sentence-level encoding W_s of text parts;
Repeat:
For $i \leftarrow 1$ *to* N
1. multi-modal model text section:
1.1. Connect micro-blogs L of user i to a virtual document as text information, $T^i=\{w_{11}^i,w_{12}^i,...,w_{1T_1}^i,w_{21}^i,w_{22}^i,...,w_{2T_2}^i,...,w_{L1}^i,w_{L2}^i,...,w_{LT_L}^i\}$;
1.2. Perform word-level encoding, $T_w^i=W_w(T,D)=\{s_1^i,s_2^i,...,s_L^i\}$;
1.3. Perform sentence-level encoding to get the textual representation of user i $t^i=W_s(T_w^i)$;
1.4. Partially predict user attribute labels, $y_t^i=Dense_t(t)$, from the text representation;
2. multi-modal model image section:
2.1. Get user i image information, n $F^i=\{F_1^i,F_2^i,...,F_n^i\}$;
2.2. Get the image representation of each image, $f^i=VGG16(F^i)=\{f_1^i,f_2^i,...,f_n^i\}$;
2.3. Calculate the mean of the image representation $f_a^i=AVG(f^i)$;
2.4. Partially predict user attribute labels, $y_f^i=Dense_f(f_a^i)$, from the image representation;
3. multi-modal alignment and fusion part of the multi-modal model:
3.1. Get computational collaborative representations, $c^i=Col(t^i,f^i)$,
3.2. multi-modal Fusion Partially Predicting User Attribute Labels $y_c=Dense_c(c^i)$
4. Calculation of the loss function, $Loss=\sum_{i=1}^{N}[e(y_t^i,\hat{y}^i)+e(y_f^i,y^i)+e(y_c^i,y^i)]$
5. Update of parameters through back-propagation
Until: $Loss$ minimum
Repeat:
For $i \leftarrow 1$ *to* N
6. Calculation of the final predictions using secondary classifiers, $y_s^i=SC(y_t^i,y_f^i,y_c^i)$
7. Calculation of the error rate, $E=1/N\sum_{i=1}^{N}e(y_s^i,\hat{y}^i)$
8. Update of secondary classifier parameters
Until: E minimum

The multi-modal model is suitable for both multi-modal classification and single-modality classification. Taking a single-modal image as an example, it is only necessary to obtain the prediction result of the image representation part, which has nothing to do with the input of text information. The test algorithm for the image representation part of the multi-modal model is presented as Algorithm 3.

Algorithm 3. Multi-modal Model Image Representation Partial Test Algorithm

1 Input: image information $F = \{F^1, F^2, ..., F^N\}$

2 Output: image representation partial predicted label $y_f = \{y_f^1, y_f^2, ..., y_f^N\}$

3 For $i \leftarrow 1$ *to* N:

4 Get image information n $F^i = \{F_1^i, F_2^i, ..., F_n^i\}$ for user i,

5 Get the image representation of each image, $f^i = VGG16(F^i) = \{f_1^i, f_2^i, ..., f_n^i\}$,

6 Calculate the mean of the image representation, $f_a^i = AVG(f^i)$,

7 Partially predict user attribute labels, $y_f^i = Dense_f(f_a^i)$, through image representation

The same is true for the case of uni-modal text in that only the prediction result f_t^i of the text part needs to be obtained. In the case of multi-modality, only the prediction results f_s^i of the secondary classifiers need to be obtained.

3. Analysis of Results

3.1. Experiment Environment

From Table 1, the model, algorithm, and optimization technology proposed in this paper are implemented using high-performance central processing unit (CPU) and graphic processing unit (GPU) heterogeneous architecture server clusters. The experiment is performed on typical software and hardware platforms, such as the Linux operating system, using C++, Java, Python and other programming languages, with the help of Tensor flow and Pytorch. An open-source deep-learning framework is established for the experiment in Table 2.

Table 1
Cluster hardware relationship

Type	Model
CPU	2 quad-/hex-/octo-core CPUs ,operating frequency2.5GHz
GPU	GEFORCE RTX 3090,10496cores,1.70GHZ
RAM	512G
Network	Gigabit Ethernet

Table 2
Experimental framework

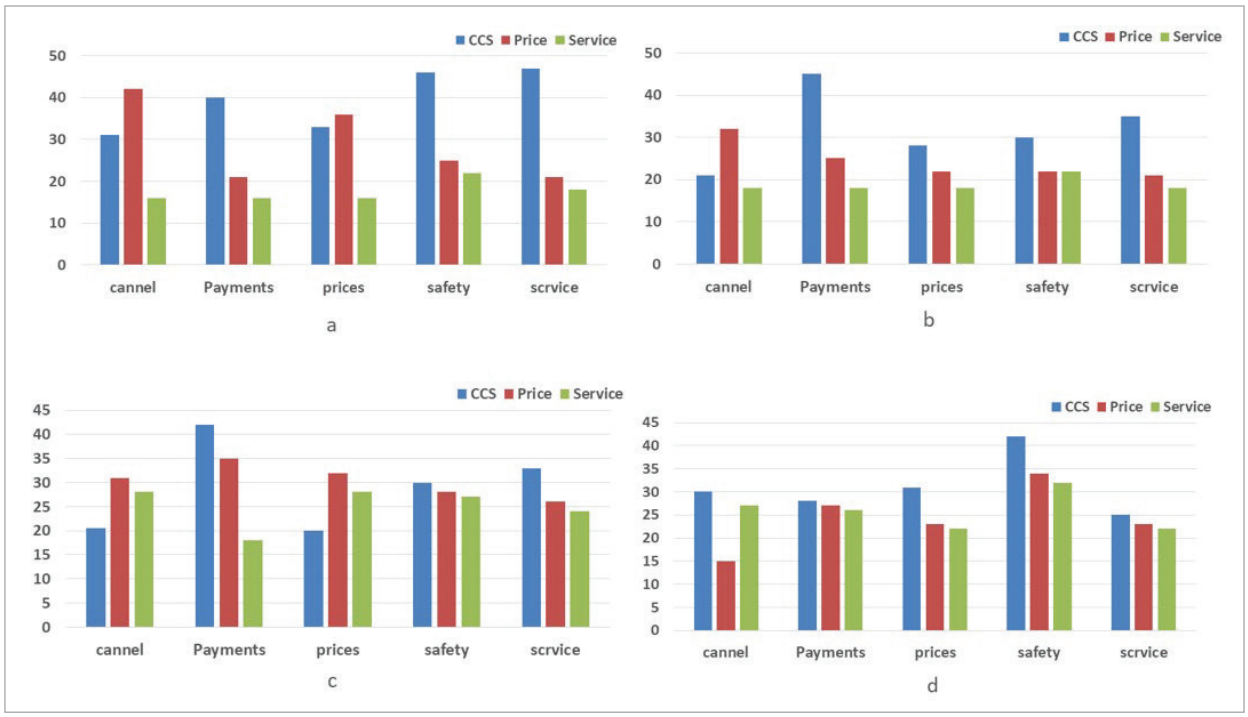
Name	Description
Tensorflow	Multi-level structure, deployed on cluster servers, PC terminals and web pages and supports GPU and TPU high-performance numerical computing
Pytorch	Provides powerful GPU-accelerated tensor computing (such as NumPy) and deep neural networks for automatic derivation systems.

3.2. Sentiment Analysis

Uber as a ride-sharing service serves 500 cities around the world and has received large volumes of feedback, suggestions, and complaints. This study collected 34,176 user comments made on Facebook, 21,605 tweets published on Twitter, and 4245 press releases.

Sentiment analysis was performed for sentiments published by CSS, Multiple Signal Classification (MUIC), and Uber. Figure 3 presents a comparative analysis of the five categories of positive reviews for cancellations, payments, prices, safety, and service in Uber's datasets on Facebook, Twitter, and The News using the method of CSS, our method, MUIC, and the results of the Uber study. In the first evaluation, the method of CSS is higher than the method of MUIC and Uber by 100% and 107%, except for cannel and price, which are lower than the results announced by MUIC and Uber. We consider that Facebook, Twitter, and News comments have certain random content and news. Irrelevant content includes shared content, marketing, and promotions, and we generated new results by deleting this content. Except that the index of Cannel's positive comment of CSS was smaller than the results published by MUIC and Uber, the results of the method proposed in this paper were close to the results published by Uber. Additionally, it is found that the sentiment of Cannel, Payment, Price, Safety and Service changed appreciably.

Figure 3
Uber sentiment analysis



Uber can get product and business feedback from positive reviews by analysing tweets, and in this paper, we re-analyses Payment and Safety from the latest analysis. After filtering out spam, marketing, news, and random irrelevant information, the third graphs and the fourth graphs are obtained. It is seen that the results of the method proposed in this paper, MUIC, are close to the results of Uber, compared to CSS.

3.3. User Portrait

The datasets used in the experiment are SemEVal and Twitter public datasets. The SemEval datasets are related to Latop and Facebook. Table 3 provides an overview of the datasets.

To better analyse the experimental results, detailed statistics on the text length of the dataset are carried out; the statistical results are given in Table 4. The length statistics of the data set are the maximum length, minimum length, average length, and proportion of lengths greater than 15.

The statistical results show that the text lengths in the Laptop data-set and Restaurant data-set are similar. The average text length is 21.5 and 20 words, respectively, and the proportions of the lengths greater

Table 3
Experiments on half semi-precision training data

	Positive		Negative		Neutral	
	Train	Test	Train	Test	Train	Test
Facebook	2154	718	815	176	603	190
Twitter	937	340	886	108	440	160
News	1561	172	1960	163	3226	356

Table 4
Statistics of data-set lengths

	Facebook		Twitter		News	
	Train	Test	Train	Test	Train	Test
Maximum length	80	70	78	69	40	30
Minimum length	2	2	3	5	2	3
Average length	21.5		20		16.3	
Length >15 Proportion	80.6%		70.1%		59.2%	

than 15 words are 80.6% and 70.1%, respectively. The text lengths in the Twitter data-set are shorter than those in the Latop and Restaurant datasets, with the average length being 16.3 words and the maximum text length being 41 words.

Through the above data analysis, the proposed multi-modal user portrait label classification model (MPCM) method is compared with the naive Bayes classifier, LDA, Tweet-LDA, and LUBD-CM(3) methods on evaluation criteria of the average accuracy rate, precision rate, recall rate, and FL-score; Naive Bayes model (NBM): A competitive classification algorithm in text classification, it is still the standard algorithm for spam filtering. LDA can easily set hyper-parameters and initialize parameters. The distribution Dirichlet distribution adopted by LUBD-CM is the same as that used in LDA; Tweet-LDA is a topic model suitable for short-text micro-blogs and for extracting topics related to microblogs, and is a core part of LUBD-CM; LUBD-CM(3) is a variant of LUBD-CM. Unlike the MPCM, which considers five behavior types, LUBD-CM(3) considers only the first three behavior types in the model; for other three behavior combinations.

A performance comparison was conducted between the naive Bayes model, LDA, Tweet-LDA, LUBD-CM(3), and MPCM in terms of the average precision, precision, recall, and F1-score. Figures 4-5 clearly show that the MPCM consistently outperforms the four benchmark algorithms of the naive Bayes model, LDA, Tweet-LDA, and LUBD-CM(3). We also performed h-tests for different measures, finding that, with 95% confidence .there was a significant difference between the MPCM and benchmark algorithms. This result verifies the effectiveness of considering multiple behavioral data in creating accurate user portraits, which is higher than usual practice. Although the performance improvement of the MPCM relative to the benchmark algorithms is only between 1% and 4%, in real life, these performance are more accurate for users' sentiment analysis portraits.

Figure 4
Performance ratio of the MPCM and benchmark algorithm (accuracy, precision)

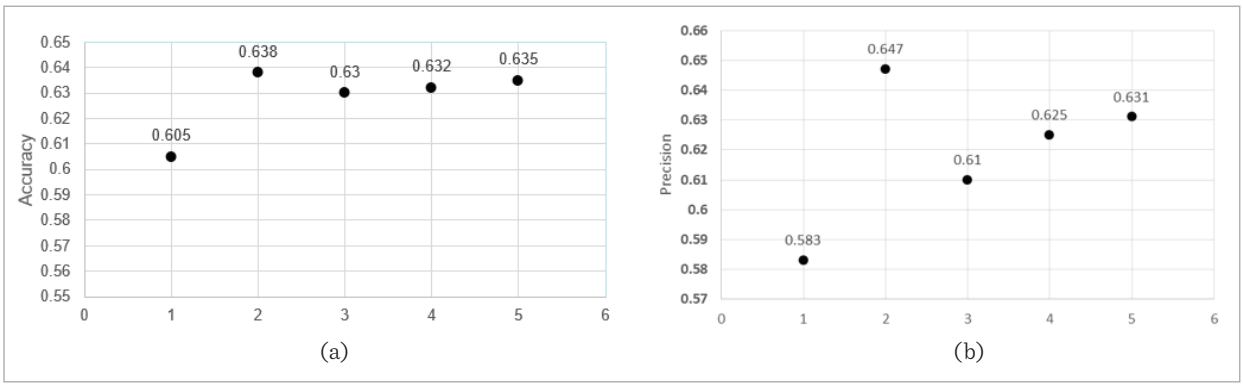


Figure 5
Performance comparison of the MPCM and benchmark algorithm (recall, F1-score)

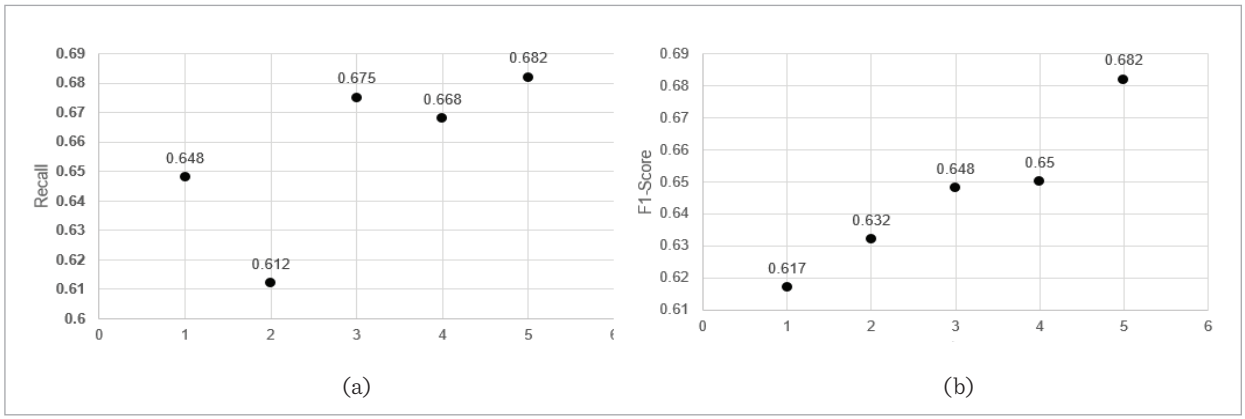
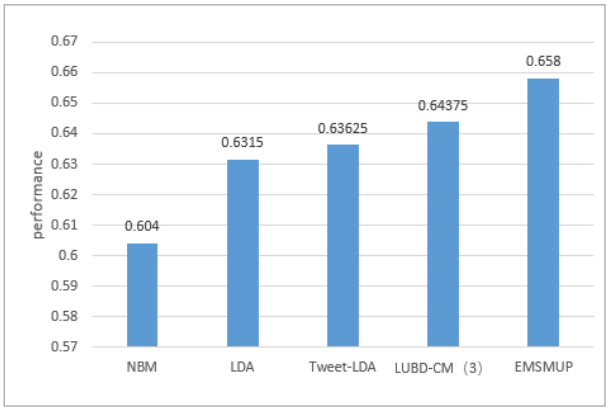


Figure.6
Performance comparison



4. Conclusion

Current user profiling methods do not well discriminate emotional differences among users of different genders and ages on social media under the conditions of multi-modality and a lack of domain sentiment labels. This paper conducted the sentiment analysis of images and text as a multi-modal social media user portrait to improve tag classification for sentiment analysis incorporating gender and age differences. Through emotional analysis computing, multi-modal data analysis, and creating a model for attribute label prediction, a method for learning the emotional representation of text and images was proposed; a multi-modal attention mechanism was introduced to realize the semantic association learning of multi-modal data of graphics and text; and a multi-modal for classification of user portrait tags. Through the analysis of User sentiment data on Facebook, Twitter, and News, the MPCM method was compared with the naive Bayes, LDA, Tweet-LDA and LUBD-CM(3) methods in terms of the accuracy, precision, recall, and FL-score of positive reviews. At a 95% confidence level, the performance improved by 5.6%–8.9% when using the MPCM method.

In the follow-up research work, we can improve the algorithm so that it can parse logs online, and when calculating the similarity between log messages and log templates, it is necessary to traverse all log templates, and the time performance is still insufficient. In the future, we will Keywords and log templates can be stored in tree form to improve retrieval speed.

Acknowledgments

The authors are grateful for the support of the National Science Foundations of China No.62106070.

Conflicts of Interest Statement

No potential conflict of interest was reported by the author.

Data Availability Statement

The datasets generated or analyzed during this study are available from Facebook and Twitter on reasonable request.

References

1. Albornoz, E. M., Milone, D. H., Rufiner, H. L. Spoken Emotion Recognition Using Hierarchical Classifiers. Computer Speech & Language, 2011, 25(3), 556-570. https://doi.org/10.1016/j.csl.2010.10.001
2. Bertani, R. M., Bianchi, R. A., Costa, A. H. R. Combining Novelty and Popularity on Personalised Recommendations Via User Profile Learning. Expert Systems with Applications, 2020, 146, 113149. https://doi.org/10.1016/j.eswa.2019.113149
3. Chen, L., Su, W., Wu, M., Pedrycz, W., Hirota, K. A Fuzzy Deep Neural Network with Sparse Autoencoder for Emotional Intention Understanding in Human-Robot Interaction. IEEE Transactions on Fuzzy systems, 2020, 28(7), 1252-1264. https://doi.org/10.1109/TFUZZ.2020.2966167
4. Farnadi, G., Sitaraman, G., Sushmita, S., Celli, F., Kosinski, M., Stillwell, D., Davalos, S., Moens, M.-F., De Cock, M. Computational Personality Recognition in Social Media. User Modeling and User-adapted Interaction, 2016, 26, 109-142. https://doi.org/10.1007/s11257-016-9171-0
5. Farnadi, G., Tang, J., De Cock, M., Moens, M. F. User Profiling Through Deep Multimodal Fusion. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, 2018, 171-179. https://doi.org/10.1145/3159652.3159691
6. Gong, L., Wang, H. When Sentiment Analysis Meets Social Network: A Holistic User Behavior Modeling in Opinionated Data. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, 1455-1464. https://doi.org/10.1145/3219819.3220120
7. Ji, Y., Yin, M., Fang, Y., Yang, H., Wang, X., Jia, T., Shi, C. Temporal Heterogeneous Interaction Graph Embedding for Next-Item Recommendation. In Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14-18, 2020, Proceedings, Part II, 2021, 314-329. https://doi.org/10.1007/978-3-030-67664-3_19
8. Li, J., Lu, K., Huang, Z., Shen, H. T. On Both Cold-Start and Long-Tail Recommendation with Social Data. IEEE Transactions on Knowledge and Data Engineering, 2019, 33(1), 194-208. https://doi.org/10.1109/TKDE.2019.2924656
9. Liang, S., Zhang, X., Ren, Z., Kanoulas, E. Dynamic Embeddings for User Profiling in Twitter. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, 1764-1773. https://doi.org/10.1145/3219819.3220043
10. Liu, D., Chen, L., Wang, Z., Diao, G. Speech Expression Multimodal Emotion Recognition Based on Deep Belief Network. Journal of Grid Computing, 2021, 19(2), 22. https://doi.org/10.1007/s10723-021-09564-0
11. Liu, N., Zhang, B., Liu, B., Shi, J., Yang, L., Li, Z., Zhu, J. Transfer Subspace Learning for Unsupervised Cross-corpus Speech Emotion Recognition. IEEE Access, 2021, 9, 95925-95937. https://doi.org/10.1109/ACCESS.2021.3094355
12. Nguyen, T. H., Shirai, K. Topic Modeling Based Sentiment Analysis on Social Media for Stock Market Prediction. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2015, 1354-1364. https://doi.org/10.3115/v1/P15-1131
13. Tang, D., Kuppens, P., Geurts, L., van Waterschoot, T. End-to-End Speech Emotion Recognition Using a Novel Context-Stacking Dilated Convolution Neural Network. EURASIP Journal on Audio, Speech, and Music Processing, 2021, (1), 18. https://doi.org/10.1186/s13636-021-00208-5
14. Yang, Y., Wang, H., Zhu, J., Wu, Y., Jiang, K., Guo, W., Shi, W. Dataless Short Text Classification Based on Biterm Topic Model and Word Embeddings. In Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence, 2021, 3969-3975. https://doi.org/10.24963/ijcai.2020/549
15. Zhang, J. Knowledge-Driven Location Privacy Preserving Scheme for Location-Based Social Networks. Electronics, 2022, 12. https://doi.org/10.3390/electronics12010070
16. Zhang, S., Tao, X., Chuang, Y., Zhao, X. Learning Deep Multimodal Affective Features for Spontaneous Speech Emotion Recognition. Speech Communication, 2021, 127, 73-81. https://doi.org/10.1016/j.specom.2020.12.009
17. Zheng, W. Q., Yu, J. S., Zou, Y. X. An Experimental Study of Speech Emotion Recognition Based on Deep Convolutional Neural Networks. In 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), 2015, 827-831. https://doi.org/10.1109/ACII.2015.7344669
18. Zheng, Y., Li, L., Zhang, J., Xie, Q., Zhong, L. Using Sentiment Representation Learning to Enhance Gender Classification for User Profiling. In Web and Big Data: Third International Joint Conference, APWeb-WAIM 2019, Chengdu, China, August 1-3, 2019, Proceedings, Part II 3, 3-11. https://doi.org/10.1007/978-3-030-26075-0_1