Research Article

Personalized Clothing Recommendation Based on User Emotional Analysis

Xueping Su [iD],[1] Meng Gao,[1] Jie Ren,[1] Yunhong Li,[1] and Matthias Rätsch[2]

[1]*School of Electronics and Information, Xi'an Polytechnic University, Xi'an710048, China*
[2]*Interactive and Mobile Robotics and Artificial Intelligence, Department of Engineering, Reutlingen University, Reutlingen, Germany*

Correspondence should be addressed to Xueping Su; yifeichongtian1201@163.com

With the continuous development of economy, consumers pay more attention to the demand for personalization clothing. However, the recommendation quality of the existing clothing recommendation system is not enough to meet the user's needs. When browsing online clothing, facial expression is the salient information to understand the user's preference. In this paper, we propose a novel method to automatically personalize clothing recommendation based on user emotional analysis. Firstly, the facial expression is classified by multiclass SVM. Next, the user's multi-interest value is calculated using expression intensity that is obtained by hybrid RCNN. Finally, the multi-interest value is fused to carry out personalized recommendation. The experimental results show that the proposed method achieves a significant improvement over other algorithms.

1. Introduction

With the rapid development of e-commerce, online shopping has become one of the main ways people spend shopping. On the one hand, there is too much information about online clothing, which may drown users in the mass clothing information; how to quickly choose the clothing they need and improve the shopping efficiency are crucial for the merchant. On the other hand, users have their own preferences and focus on individual needs of clothing. Therefore, research on the clothing personalized recommendation method is very important to improve the user's shopping efficiency and meet the user's personalized needs. However, in the traditional personalized recommendation, due to the lack of user information, the recommendation quality of the clothing recommendation system is not high enough to meet the user's expectations. In addition, personalized recommendation function is limited to recommending products related to the user or favoured by the user.

Furthermore, with the development of emotional computing and intelligent human-computer interface, the computers are required to perceive and understand human's expression. There are a variety of facial expression-recognition methods, but existing expression analysis frameworks are less likely to measure the intensity of expressions. In the field of human-computer interaction, the measurement of facial expression intensity is also very necessary; it can help computers understand people's emotions. For example, when a user browses clothing information, the intensity of his expression can reflect the degree of affection of clothing. The computer can use it to recommend the clothing of interest to the user. Therefore, we propose an efficient method for the personalized recommendation of clothing based on user emotional analysis. The novel contributions of the proposed work are as follows:

(1) The scheme of the hybrid recurrent convolutional neural network (RCNN) is proposed to compute the expression intensity. These implementations improve the precision of personalized clothing recommendations.

(2) To the best of our knowledge, this is the first time to describe the user's multi-interest value by combining

expression intensity and the expression duration, which well captures the user's preferences and improves the recall of personalized clothing recommendation.

The remainder of this paper is organized as follows. We first describe some related research for facial expression recognition and personalized clothing recommendation. In section 3, we introduce our method, focusing on facial expression recognition by multi-class support vector machine (SVM), computing the expression intensity that is obtained by the hybrid RCNN, fusing the user's multi-interest value and personalized recommendation. In section 4, we present detailed experimental results and compare the performance of our proposed method with other current approaches. Finally, we conclude with discussions.

2. Related Work

Deep neural networks have been successfully applied in computer vision, especially in face recognition, where the use of convolutional neural networks (CNNs) outperforms all the previously proposed methods, and the obtained results surpass the human performances [1–5]. Subsequently, Zhou et al. [6] proposed the recurrent convolutional neural network (RCNN) for object recognition by applying recurrent connections with the same layer. With fewer parameters, the RCNN achieved better results than the state-of-the-art CNNs by testing object recognition datasets [7]. The end-to-end RCNN framework can predict the pain intensity of each frame by considering sufficiently large historical frames while limiting the scale of the parameters within the model [6]. Besides that, the RCNN outputs continuous scores rather than discrete labels as in the problem of classification.

Facial expression is salient information to understand certain target's emotional situation. Most of the human emotional expressions are able to be observed on their face than any other signs. At the same time, the CNN is used for the facial expression recognition task with Tang [8], Bergstra [9], and Jeon et al. [10] and achieved the best performance on Kaggle facial expression recognition challenge. Tang used the CNN with linear-SVM instead of the SoftMax layer in the classification phase. His model performed the best accuracy of 69.77% on the challenge. Bergstra's model is concentrated to hyperparameter optimization. Jeon constructed a real-time facial expression recognizer using a deep neural network which is invariant to the subject. Soon after, many deep learning methods are used for facial expression recognition and have achieved good performance [11–14]. In summary, there are a variety of facial expression recognition methods, but the method of expression intensity based on deep learning is less [15].

Meanwhile, with the continuous development of economy, consumers pay more attention to the demand for personalization of clothing. Personalized clothing recommendation not only meets the personalized needs of consumers but also greatly saves time for consumers to choose clothes. Therefore, the personalized clothing recommendation

has attracted the attention of domestic and foreign clothing experts, and the method of personalized clothing recommendation has emerged [16–18]. The clothing personalized recommendation system mainly obtains user preferences based on the user's purchase record, browsing history, and neighbouring user information analysis. There are problems such as cold start and low degree of personalization, which cannot satisfactorily satisfy the personalized recommendation effect.

In summary, we propose a novel method to automatically personalize clothing recommendation based on user multi-interest value, which is calculated using expression intensity that is obtained by the hybrid RCNN.

3. The Proposed Framework

3.1. Initialization Recommendation. Clothes are divided into multiple classes by the affinity propagation cluster. For each class of clothes $(\overline{c}_i, \sigma_i)$, $\overline{c}_i$ represents the mean of the class and σ_i represents the variance of the class. For the user u, we calculate the similarity between each class of clothes and the user, and recommend the suitable class of clothes for the user according to the ranking of similarity. The formula for calculating similarity between the class of clothes and user is as follows:

$$d_{cu}(\overline{c}_i, u) = \sqrt{\sum_{k=1}^{K} \left(\frac{(\overline{c}_{ik} - u_k)}{\sigma_{ik}} \right)^2},$$

$$S_{cu}(\overline{c}_i, u) = -\exp\left(d_{cu}(\overline{c}_i, u)\right), \tag{1}$$

$$i = 1, 2, \ldots, N,$$

where $d_{cu}(\overline{c}_i, u)$ is the Euclidean distance of the class of clothes $\overline{c}_i$ and the user u, $\overline{c}_{ik}$ and u_k are the k-th features of the class of clothes $\overline{c}_i$ and the user u, N is the total number of the class of clothes, and K is the feature dimension number. $S_{cu}(\overline{c}_i, u)$ represents the similarity between the class of clothes $\overline{c}_i$ and the user u.

Similarly, the other users are also divided into multiple clusters by the affinity propagation cluster. For each class of users $(\overline{u}_j, \sigma_j)$, $\overline{u}_j$ represents the mean of the class, σ_j represents the variance of the class. We calculate the similarity between each class of users and the user and recommend a suitable other class of users' clothes for the user according to the ranking of similarity. The formula for calculating similarity between the class of users and user is as follows:

$$d_{uu}(\overline{u}_j, u) = \sqrt{\sum_{k=1}^{K} \left(\frac{(\overline{u}_{jk} - u_k)}{\sigma_{jk}} \right)^2},$$

$$S_{uu}(\overline{u}_j, u) = -\exp\left(d_{uu}(\overline{u}_j, u)\right). \tag{2}$$

$$j = 1, 2, \ldots, M$$

where $d_{uu}(\overline{u}_j, u)$ is the Euclidean distance of the class of users $\overline{u}_j$ and the user u, $\overline{u}_{jk}$ and u_k are the k-th features of

other user $\overline{u}_j$ and the user u, M is the total number of the class of other users, and K is the feature dimension number. $S_{uu}(\overline{u}_j, u)$ represents the similarity between the class of users $\overline{u}_j$ and the user u.

3.2. Calculation of Expression Intensity.

The high-definition camera was used to obtain the user's expression feedback of the initialization recommendation clothing video and recognize the user's expression in the video. We adopt the multiclass SVM method for dynamic expression recognition and use the method of the recurrent convolution neural network (RCNN) to evaluate the expression (happy, angry, etc.) intensity.

3.2.1. Facial Expression Recognition.

We convert facial expression recognition into a classification problem, and recognition of the expression (happy, sad, etc.) in the video. The specific expression recognition framework is shown in Figure 1:

The specific steps are as follows:

Firstly, the video is transformed into a series of frames of image sequences, and the active shape model method is used for face detection, and the video volume is created [7]. Secondly, the Local Gabor Binary Pattern Histogram Sequence (LGBPHS) features [19] of the three planes XY, XT, and YT are extracted, and all the features of the video volume are combined as the features of the final image sequence (the specific steps are shown in Figure 2).

Multiclass SVM aims to assign labels to instances by using SVM, where the labels are drawn from a finite set of several elements (Sad, Happy, Angry, Disgust, Fear, Surprise, and Neutral). The dominant approach for doing this is to reduce the single multiclass problem into multiple binary classification problems. Common methods for such reduction include one versus all and one versus one. Both methods have been found to produce approximately similar results when dealing with face recognition. Compared with one versus one, one versus all constructed a much less number of decision planes. When the number of classes is large, the prediction speed is faster. In our paper, we use one versus all to train multiclass SVM.

We train multiclass SVM for expression recognition. The procedure is as follows. Firstly, for the samples of happy, we assign 1 as the class label (Sad/Angry/Disgust/Fear/Surprise/Neutral set to 2/3/4/5/6/7). Secondly, the samples of all emotions are trained for multiclass SVM.

3.2.2. Calculation of Expression Intensity

(1) Calculation of Expression Duration. We use the expression duration as one of the expression intensities, and calculate the expression duration based on the recognized frame expressions. The calculation formula is as follows:

$$I_i = \frac{TE_i}{T_i}, \quad i = 1, 2, 3, \ldots, M, \tag{3}$$

where I_i is the time value of the i-th interest measure and TE_i is the duration of the interest in the i-th interest measure (expression: happy, angry, etc.). T_i is the viewing total time. Then, the time values of all interest measures are sorted in descending order. The smaller the serial number, the more interested in trying on the clothing. Set a threshold for I_i, and recommend it to the user if I_i is greater than the threshold.

(2) Evaluate of Expression Intensity. From the field of face recognition, the face model strained on several specific facial parts can significantly improve the recognition accuracy [20, 21]. Compared with the full-face model, the specific part model is able to extract more detailed information. For the sake of exploring the effectiveness of different face parts, we divide the entire face into several parts. In addition, based on the promising results obtained by the RCNN, we trained a hybrid RCNN using different face parts. The main idea of our method can be concluded as (1) train a hybrid RCNN based on the face region, eye region, and mouth region. (2) Concatenate the last fully connected layer of the hybrid RCNN to constitute the features.

To save computation and reduce the time consumption, we simplified the architecture of the RCNN [6]. The RCNN used in our paper contains one convolutional layer, three Recurrent Convolutional Layers (RCLs), three max pooling layers, and one fully connected layer. The first layer is the standard feed-forward convolutional layer without recurrent connections, followed by max pooling. Three RCLs are used with a max pooling layer in the middle. Between neighbouring RCLs, there are only feed-forward connections. The output of the third RCL follows a global max pooling layer, which outputs the maximum over every feature map, yielding a feature vector representing the image. The main architecture of the hybrid RCNN is shown in Figure 3.

The framework of the hybrid RCNN to evaluate expression intensity is shown in Figure 4. The specific steps are as follows.

For each frame of the face image in the video sequence, firstly, the active shape model method is used to detect face feature points [7]. Secondly, in the process of face aligning and warping, we warped every facial image in R, G, and B channels separately, then combined all channels back to get the final RGB warped faces. Third, the input samples of our hybrid RCNN structure should be no more than two dimensions, but to reserve the temporal information among frames and the spatial pixel information of warped facial images at the same time, each frame is converted into a one-dimensional vector by flattening. After flattening, we concatenated all 1D flattened warped facial images in frame order to achieve frame vector sequences.

3.2.3. Calculation of Interest Value.

We combine the expression intensity and the expression duration as the user interest value. The calculation formula of interest value is as follows:

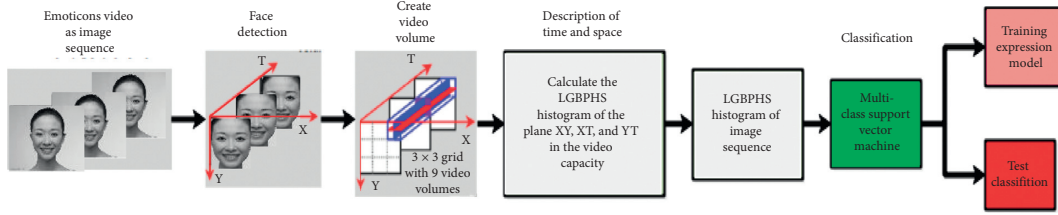$$I_e = \alpha I_i + \beta I_t, \tag{4}$$

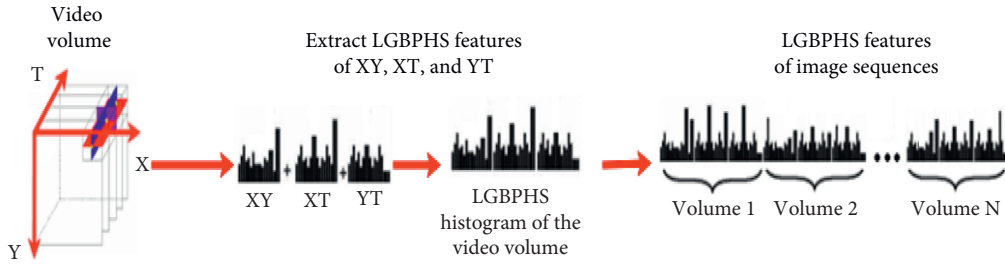FIGURE 1: Expression recognition framework.

FIGURE 2: LGBPHS feature flow chart for extracting image sequences.
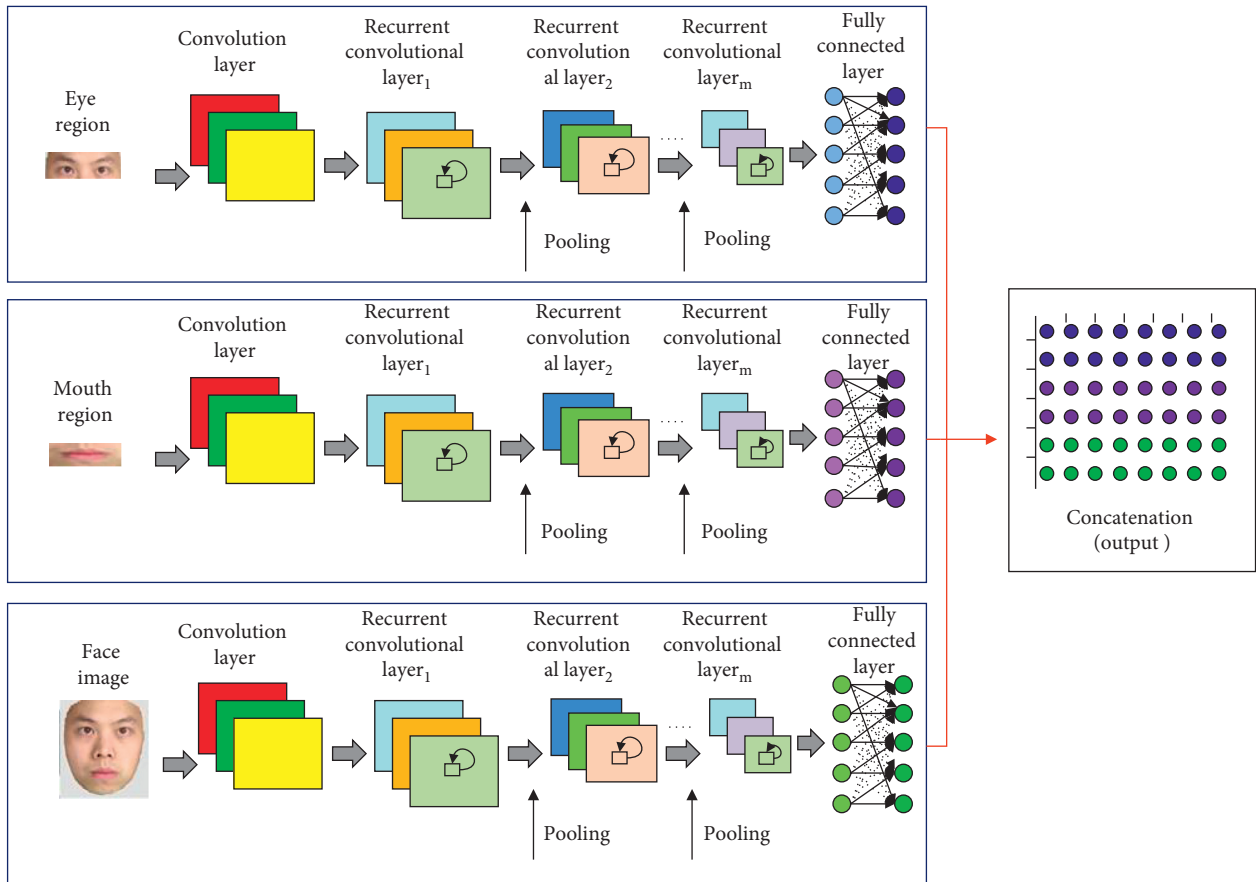

FIGURE 3: Primary architecture of the hybrid RCNN.

where I_e is the interest value, I_i and I_t are the duration value and intensity value of the expression, respectively, and α β are the weights value, and it meets the formula $\alpha + \beta = 1$.

3.3. *Fusion User's Multi-Interest Value and Personalized Recommendation.* We intend to calculate the user's multi-interest (colour, style, texture, price, etc.) value. The method
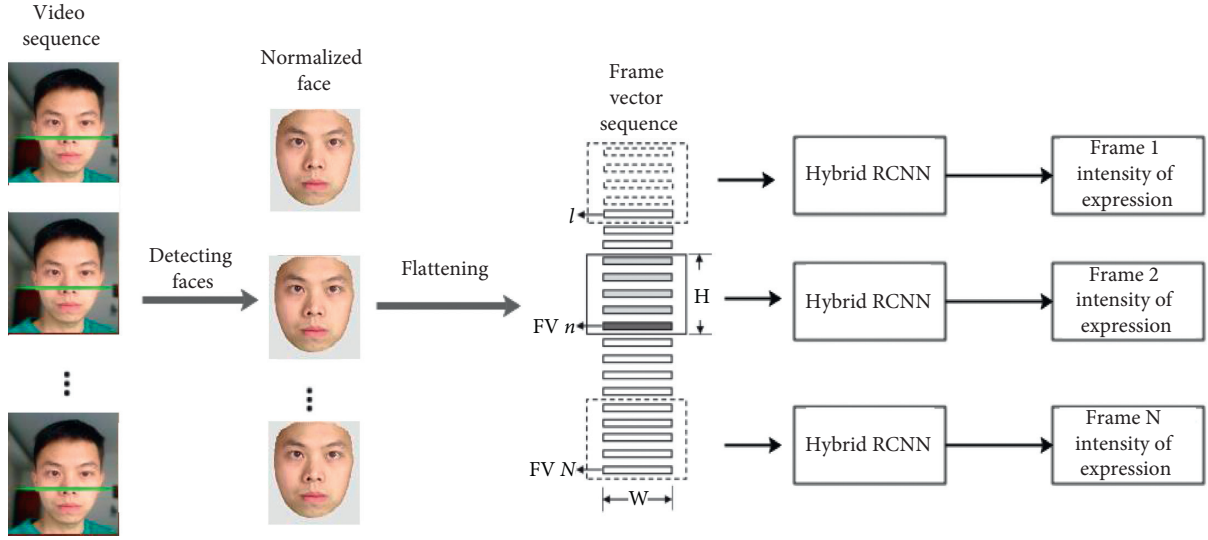
FIGURE 4: Frame diagram of the expression intensity evaluation method.

of multi-interest value fusion is adopted to carry out personalized recommendation.

Rank aggregation is the fusion of decision-making results which is expressed by the order list. Because the order list expresses the decision-making result, it is necessary to facilitate direct comparison of the different results, and contains a wealth of information for decision-making result. Therefore, we exploit the weighted Borda count method [22] to fuse multi-interest value. Ballots in the Borad count are counted by assigning a point value to each place in the hierarchy, and the choice with the largest number of points is selected. The Borda method scores each sequence of the list of interests (colour, style, texture, and price) linearly, and the score of objects i in the sequence τ_j of interest measure is estimated as follows:

$$B^j(i) = -\tau_j(i), \quad i = 1, 2, 3, \ldots, M, j = 1, 2, 3, \ldots, N; \tag{5}$$

where $\tau_j(i)$ is the ordinal function of the sequence τ_j of interest value, and it indicates the order of the object i in the interest value list τ_j, M is the total number of objects, and N is the total number of interest lists. The symbol in the formula is to place the value of the object in front of the order of bits higher. When J sequence of interest value is fused, the weighted Borda method is adopted according to the performance difference of sequences of different interest measure, which is:

$$B(i) = \sum_{j=1}^{J} w_j B^j(i), \quad i = 1, 2, 3, \ldots, M, j = 1, 2, 3, \ldots, N; \tag{6}$$

where $B^j(i)$ is the score of the object i in the j-th sequence of interest value, M is the total number of objects, N is the total number of interest value lists, and w_j is the weight of the interest measure sequence j. The calculation formula is as follows:

$$w_j = \text{AP}_j, \quad j = 1, 2, 3, \ldots, N;$$
$$\sum_{j=1}^{J} w_j = 1, \quad 0 < w_j < 1, \tag{7}$$

where AP_j (average precision) is the average accuracy of the j-th sequence of interest value.

We sort the multi-interest (colour, style, texture, price, etc.) value and then use the weighted Borda method to perform rank aggregation. The final rank aggregation result is the sorting of the weighted scores from high to low, which gives a personalized recommendation based on the results.

4. Experiments

4.1. Dataset. Kaggle [23] facial expression recognition challenge database is used for training and testing performance. This dataset has 7 facial expression categories (angry, disgust, fear, happy, sad, surprise, and neutral), 28,709 training images, 3,589 validation images, and 3,589 test images. This dataset contains the human frontal face with various illumination, poses, and domains, and even cartoon characters are included. Moreover, in Kaggle facial recognition challenge training dataset, 7215 images are in the happy category and 436 images are in the disgust category.

Forty female university students from Soochow University conducted a verification experiment on the personalized recommendation system. Forty subjects had a rating of points for the system's recommended clothing, and it contains 132 sample pictures of women's winter wool coats [23]. Among them, 32 clothing images (See Figure 5) are used to train and 100 clothing images are used to recommend. Women's winter woollen coat has 7 attributes, of which the colour attribute has 8 attribute values, and the other 6 attributes have 4 attribute values.

FIGURE 5: Training images.

4.2. Experimental Setup. For the clothing sample, we let each image in the sample to automatically play at intervals of 3 seconds. During the playback, the user is asked to evaluate the clothing according to their own preferences, so as to obtain the user's evaluation form for the sample. *E* is used to indicate the user's evaluation of the sample clothing. The values of *E* are 1, 2, 3, 4, and 5, respectively, indicating that the consumer's evaluation of each garment in the sample is "very dislike, "dislike," "general," "like," and "very like." Similarity, the expression intensity values are graded into 5 intensity levels: [0,0.2] for level 1, the weakest level, [0.2, 0.4] is level 2, and so on, [0.8,1] is level 5.

The hybrid RCNN was implemented and run on two GeForce GTX TITAN Black GPUs. Initially, the weight of feed-forward/recurrent is set to 0.02, and the bias is set to 1. In addition, the parameter β in equation (4) are analysed in Figure 6. In order to better trade off average precision, average recall and average *E*, α is set to 0.2 and β is set to 0.8 for calculating interest value.

4.3. Results and Analysis. We compare the proposed method with Wang's method [16], Hu's method [17], and Melo's method [18]. Wang's method is tested on the same clothing dataset which we use in our experiments. Hu's method used the user interest degree to express the user preference model; the idea is similar to Melo's method and our paper. To evaluate the effectiveness of the proposed method, we select 40 people to evaluate the recommend clothing, and evaluate the performance by calculating the average precision, average recall, and average *E*. The experiment result is shown in Table 1.

As shown in Table 1, our proposed method obtained the better result for personalized clothing recommendation.
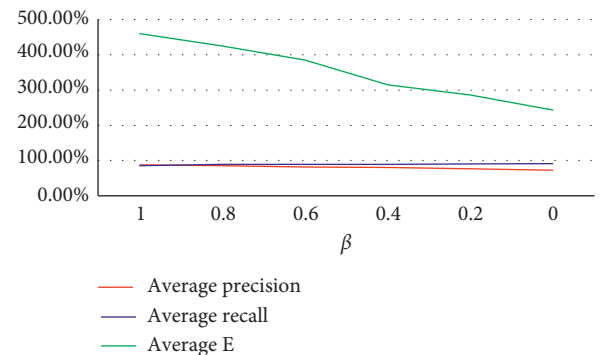
FIGURE 6: Compared results with varying β.

Because the measurement of facial expression intensity can help computers understand people's emotions, when a user browses clothing information, the intensity of his expression can reflect the degree of affection of the clothing. The computer can use it to recommend the clothing of interest to the user, which can satisfactorily satisfy the personalized recommendation effect.

In addition, we compare the classification accuracy of the facial expression recognition method with Tang's method [8] and Jeon's method [10]. These two methods are tested on the same facial expression dataset that we use in our experiments. The classification accuracy of facial expression is shown in Table 2.

As shown in Table 2, the average accuracy for all categories was 72.36% in our method. Accuracy for the happy and surprise category was higher than the others, but accuracy for the fear category was poor.

Meanwhile, we also compare the precision of classification of intensity levels for happy expressions with SVM

TABLE 1: Experiment result of different methods.

Method	Average precision (%)	Average recall (%)	Average E
Wang's method [16]	72.5	82.6	3.625
Hu's method [17]	72.3	80.5	3.86
Melo's method [18]	80.6	83.4	3.98
Our method	85.4	88.9	4.25

TABLE 2: Classification accuracy of facial expression in different methods.

Method	Angry (%)	Disgust (%)	Fear (%)	Happy (%)	Sad (%)	Surprise (%)	Neutral (%)
Tang's method [8]	48.2	55.4	38.6	69.7	50.8	70.9	59.2
Jeon's method [10]	61.1	67.3	49.5	88.7	65.2	83.2	69.2
Our method	65.4	69.9	55.8	90.6	66.8	86.7	71.3

TABLE 3: Precision of classification of intensity levels for happy expressions.

Method	Precision (%)
SVM [15]	81.3
RCNN [6]	84.2
Our method	90.6

[15] and the RCNN [6]. Although Zhou proposed an automatic frame-by-frame pain (not facial expression) intensity estimation framework in a video based on the RCNN [6], the solution to intensity estimation is similar.

As shown in Table 3, our proposed method obtained the better result for facial expression intensity estimation. Compared with the full-face model, the specific part model is able to extract more detailed information, so the face model strained on several specific facial parts can significantly improve the precision of expression intensity estimation.

5. Conclusion

We have presented a method for personalized recommendation of clothing based on the user's emotional analysis. Particularly, the hybrid RCNN is used to compute the expression intensity, which improves the precision of personalized clothing recommendation. In addition, to capture the user's preferences, the user's multi-interest value is computed by combining expression intensity and expression duration, which improves the recall of personalized clothing recommendation. For the datasets used in the experiments, our proposed method is superior to other existing methods.

There are still some possible directions to improve the performance of our method. In this study, we only process fewer datasets, and multiclass SVM can get much better results than other algorithms on small sample training sets, so multiclass SVM is used for expression classification. However, for a large-scale dataset, a large amount of storage space is required, which multiclass SVM cannot handle. Most deep learning methods can get good results for expression classification. Besides that, we use rank aggregation for recommendation, and collaborative filtering and knowledge graph could be also used for recommendation. The proposed idea can also be applied to other problems

such as personalized news recommendations, personalized travel recommendations, and so on.

Data Availability

Kaggle facial expression recognition challenge database: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation by joint identification verification," *Advances in Neural Information Processing Systems*, pp. 1988–1996, 2014, http://arxiv.org/abs/1406.4773.

[2] X. Chen, E. Zhou, Y. Mo, J. Liu, and Z. Cao, "Delving deep into coarse-to-fine framework for facial landmark localization," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2088–2095, IEEE, Honolulu, HI, USA, July 2017.

[3] Y. Sun, X. Wang, and X. Tang, "Sparsifying neural network connections for face recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4856–4864, IEEE, Las Vegas, NV, USA, June 2016.

[4] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: a unified embedding for face recognition and clustering," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823, IEEE, Boston, MA, USA, June 2015.

[5] Z. Zhu, G. Huang, W. Zou, D. Du, and C. Huang, "UCT: learning unified convolutional networks for real time visual tracking," in *Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 1973–1982, IEEE, Venice, Italy, October 2017.

[6] J. Zhou, X. Hong, F. Su, and G. Zhao, "Recurrent convolutional neural network regression for continuous pain intensity estimation in video," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 84–92, IEEE, Las Vegas, NV, USA, July 2016.

[7] X. Su and H. Zhou, "Automatic focus personage identification in multilingual news image," in *in Proceedings of the 2017 International Conference on the Frontiers and Advances in Data Science (FADS)*, pp. 1–6, IEEE, Xi'an, China, October 2017.

[8] Y. Tang, "Deep learning using linear support vector machines," 2013, https://arxiv.org/abs/1306.0239.

[9] J. Bergstra and D. D. Cox, "Hyperparameter optimization and boosting for classifying facial expressions: how good can a "null" model be?," 2013, https://arxiv.org/abs/1306.3476.

[10] J. Jeon, J.-C. Park, Y.J. Jo et al., "A real-time facial expression recognizer using deep neural network," in *Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication—IMCOM '16*, pp. 91–94, Association for Computing Machinery, New York, NY, USA, January 2016.

[11] H. Yang, U. Ciftci, and L. Yin, "Facial expression recognition by de-expression residue learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2168–2177, IEEE, Salt Lake City, UT, USA, June 2018.

[12] F. Zhang, T. Zhang, Q. Mao, and C. Xu, "Joint pose and expression modeling for facial expression recognition," in *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3359–3368, IEEE, Salt Lake City, UT, USA, June 2018.

[13] C.-M. Kuo, S.-H. Lai, and M. Sarkis, "A compact deep learning model for robust facial expression recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2121–2129, IEEE, Salt Lake City, UT, USA, June 2018.

[14] A. T. Lopes, E. de Aguiar, A. F. de Souza, and T. Oliveira-Santos, "Facial expression recognition with convolutional neural networks: coping with few data and the training sample order," *Pattern Recognition*, vol. 61, pp. 610–628, 2017.

[15] Li Yan, "Research on methods for facial expression intensity measurement based on video," pp. 1–57, Hunan University, Changsha, China, 2006, M.S. thesis.

[16] L.-L Wang, H.-Q Dai, and J. Wang, "The research on consumer's clothing preference based on clothing sample," *TBIS*, vol. 10, pp. 528–534, 2017.

[17] J. Hu, Z. Wang, and S. Han, "Research on personalized clothing recommendation mode based on user preference," *Journal of Zhejiang Science-Technology University (Social Sciences)*, vol. 40, no. 2, pp. 136–143, 2018.

[18] E. V. D. Melo, E. A. Nogueira, and D. Guliato, "Content-based filtering enhanced by human visual attention applied to clothing recommendation," in *2015 Proceedings of the IEEE 27th International Conference on Tools with Artificial Intelligence*, pp. 644–651, IEEE, Vietri sul Mare, Italy, November 2015.

[19] X. Su, J. Peng, X. Feng, J. Wu, J. Fan, and L. Cui, "Cross-modality based celebrity face naming for news image collections," *Multimedia Tools and Applications*, vol. 73, no. 3, pp. 1643–1661, 2013.

[20] J. Lu, V. E. Liong, G. Wang, and P. Moulin, "Joint feature learning for face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 7, pp. 1371–1383, 2016.

[21] L. Li, X. Feng, Z. Boulkenafet, Z. Xia, M. Li, and A. Hadid, "An original face anti-spoofing approach using partial convolutional neural network," in *Proceedings of the 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 1–6, IEEE, Oulu, Finland, December 2017.

[22] O. Melnik, Y. Vardi, and C.-H. Zhang, "A probability model for combining ranks," in *Proceedings of the Multiple Classifier Systems: 6th International Workshop*, pp. 74–85, Seaside, CA, USA, June 2005.

[23] I. J. Goodfellow, D. Erhan, P. L. Carrier et al., "Challenges in representation learning: a report on three machine learning contests," in *Proceedings of the International Conference on Neural Information Processing*, pp. 117–124, Springer, Berlin, Germany, 2013.