Emotion Recognition From Multi-Channel EEG Signals by Exploiting the Deep Belief-Conditional Random Field Framework

HAO CHAO AND YONGLI LIU
School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454003, China
Corresponding author: Hao Chao (chaohao1981@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61502150, in part by the Fundamental Research Funds for the Universities of Henan Province under Grant NSFRF1616, in part by the Foundation for Scientific and Technological Project of Henan Province under Grant 172102210279, and in part by the Key Scientific Research Projects of Universities in Henan under Grant 19A520004.

ABSTRACT Recently, much attention has been attracted to automatic emotion recognition based on multi-channel electroencephalogram (EEG) signals, with the rapid development of machine learning methods. However, traditional methods ignore the correlation information between different channels, and cannot fully capture the long-term dependencies and contextual information of EEG signals. To address the problems, this paper proposes a deep belief-conditional random field (DBN-CRF) framework which integrates the improved deep belief networks with glia chains (DBN-GC) and conditional random field. In the framework, the raw feature vector sequence is firstly extracted from the multi-channel EEG signals by a sliding window. Then, parallel DBN-GC models are utilized to obtain the high-level feature sequence of the multi-channel EEG signals. And the conditional random field (CRF) model generates the predicted emotion label sequence according to the high-level feature sequence. Finally, the decision merge layer based on K-nearest neighbor algorithm is employed to estimate the emotion state. According to our best knowledge, this is the first attempt that applies the conditional random field methodology to deep belief networks for emotion recognition. Experiments are conducted on three publicly available emotional datasets which include AMIGOS, SEED and DEAP. The results demonstrate that the proposed framework can mine inter correlation information of multiple-channel by the glia chains and catch inter channel correlation information and contextual information of EEG signals for emotion recognition. In addition, the classification accuracy of the proposed method is compared with several classical techniques. The results indicate that the proposed method outperforms most of the other deep classifiers. Thus, potential of the proposed framework is demonstrated.

INDEX TERMS Human-computer interaction, emotion recognition, multi-channel EEG signal, DBN-GC, conditional random field.

I. INTRODUCTION Emotion is a psychological and physiological state accompanied by cognitive and conscious processes, which plays a very important role in human communication. In human-computer interaction (HCI) system, the machine should have the functionality of understanding human affective behaviors. This functionality helps to understand human intentions accurately in order to trigger appropriate feedback. Therefore, affective computing has gradually become a focus field of emotional research in recent years. Specifically, affective computing is performed by computational model, which is established by constructing emotional states. It analyses behavior and physiological signals, and establishes emotional interaction between human and computer based on measured emotional states. As an indispensable part of affective computing, emotion recognition, which predicts and estimates the corresponding emotional state through the user’s behavior and physiological response, becomes increasingly attractive [1].

Emotion recognition can be implemented by external information such as voice intonation [2], [3], facial expression [4] and body posture [5], as well as by changes in the nervous system such as EEG signals. Because of various reasons, external information is easy to be artificially concealed, so the
model of the label sequence when giving the observation sequence. CRF provides two attractive advantages over other discriminative models [22]. Firstly, unlike Hidden Markov Models (HMMs), CRF does not require the strict independence assumptions on the observations. Therefore, it can effectively utilize the overlapping and long-range dependent information of observations. Secondly, unlike SVM and neural networks (NN), CRF determines an optimal result according to the whole sequence. The advantages make CRF more suitable for processing sequential data. Inspired by it, the paper investigates the possibility of integrating the CRF into deep learning framework to achieve superior recognition performance. In reference [23], Banda and Engelbrecht proposed the deep continuous conditional recurrent neural field framework to model dimensional emotion from multiple input features.

In this paper, we proposed a novel deep learning framework termed deep belief-conditional random field. In the framework, deep belief network with glia chains instead of DBN is employed to learn high-level abstract features. Compared with DBN, DBN-GC can not only utilize time-domain or frequency-domain information of EEG signals, but also mine inter correlation information of multiple-channel by the glia chains. Utilization of deep models reduces the difficulty of extracted task-related features and helps to bypass the feature selection stage. Considering that emotion state is a response to external events and evolves with the change of stimulus, EEG signals as an expression of emotional state contain rich contextual and semantic information. Therefore, CRF is adopted to handle the evolution and long term dependencies of the EEG signals for emotion recognition. Therefore, the contribution of this paper lies in two aspects. Firstly, in addition to time-domain and frequency-domain features, this paper also attempts to inter channel correlation information and contextual information of EEG signals into emotion recognition. Secondly, the deep belief-conditional random field framework is proposed to catch inter channel correlation information and contextual information of EEG signals.

The remaining of this paper is organized as follows. The basic background knowledge of DBN and CRF is elaborated in Section II. Section III provides the detailed description of the proposed DBN-CRF framework. The experimental settings and evaluation results are described in Section IV. Section V briefly concludes this work.

II. THEORETICAL BACKGROUND

In order to better understand the following DBN-CRF framework, in this section, we briefly introduce the theoretical background of relevant models, namely restricted Boltzmann machine (RBM), DBN, and CRF.

A. RBM AND DBN

As a kind of Boltzmann machine with special topological structure, RBM is a generative stochastic artificial neural network that can learn a probability distribution by input data set [24]. RBM consists of two parts: the visible layer...
Deep belief network was proposed by Hinton et al. [25] in 2006. As a deep learning model, it has attracted wide attention and has been successfully applied in object recognition, speech recognition and other fields. Deep belief network is a multi-hidden layers network composed of several restricted Boltzmann machines and a supervised output layer, as shown in Figure 2. $v$ represents the visual layer, and $h_i$ represents the hidden layer. DBN training includes two steps: pre-training and fine-tuning.

In the pre-training stage, unsupervised layerwise training is employed to train the RBMs in each layer. The output of the hidden layer of the lower RBM is used as the input of the visible layer of the upper RBM. This training method can learn some non-linear complex functions by mapping data directly from input to output, which makes DBN have strong feature extraction ability. In fine-tuning stage, supervised learning is used to train the network in the last layer, and the errors between actual output and expected output are propagated backwards layer by layer, and the weights of the whole DBN network are fine-tuned.

### B. CONDITIONAL RANDOM FIELD

CRF is a discriminative probabilistic model which is often used to annotate or analyze sequence data, such as natural language [26], [27]. CRF constructs conditional probability distribution $P(Y \mid X)$, which represents Markov random fields of a group of output random variables $Y$ given a set of input random variables $X$. Similar to Markov random fields, conditional random fields are undirected graph models, where vertices represent random variables and links between vertices represent dependencies among random variables. In CRF, the distribution of random variables $Y$ is conditional probability, and the given observation value is random variables $X$. In principle, the layout of graph model of conditional random fields can be given arbitrarily. However, the common layout is the chain architecture because there are efficient algorithms in training, inference and decoding for chain architecture. Among the linear chain, conditional random field is the most commonly used one.

### III. DBN-CRF FRAMEWORK

To employ the temporal dependencies of EEG signal and utilize the correlation information between different channels, we present the DBN-CRF framework for EEG based emotion recognition task, and offers the details of the whole framework.

#### A. OVERVIEW OF THE FRAMEWORK

We integrate the DBN-GC methodology and conditional random field together to establish a novel deep framework. The architecture consists of three key parts: parallel DBN-GCs, CRF based sequence learning model and decision merge layer, as shown in Figure 3.

The multi-channel EEG signals of current sample are split into several segments by a specific time window, and a raw feature vector can be extracted from the multi-channel EEG.
signals of the current segment. Thus, a raw feature vector sequence can be achieved. Each raw feature vector is fed into the corresponding DBN-GC to extract the high level representations. Thus, a high-level feature sequence can be achieved. Due to the existence of glia chain, DBN-GC can mine inter-channel correlation information of EEG signal. Then, the unit of sequence learning, which adopts a CRF structure, models the contextual information for the high-level feature sequence and generates the predicted emotion label sequence $Y = \{y_1, y_2, \ldots, y_l\}$. Finally, the decision merge layer base on k-nearest neighbor algorithm is utilized to determine the emotion state of the observation sequence. The proposed framework combines the strong ability of the DBN-GC model in processing intra-channel and cross-channel information and the ability of CRF in processing sequential data. Therefore, it is quite suitable for processing sequential signal such as EEG or speech.

### B. INDEPENDENT DBN-GC LEARNING

DBN-GC is proposed based on deep belief network which is composed of many restricted Boltzmann machines (RBM) in the stacking way. In DBN-GC, each RBM has two node layers, as well as a group of glia cells represented by stars and linked into a chain structure, as shown in Figure 4. Each glia cell is also connected to a unit in the hidden layer of RBM, as shown in Figure 5. There are connection weights between glia cell and corresponding node unit. In the training process, the role of all glia cells can be directly applied to the hidden unit, so that the output of the hidden layer nodes can be adjusted accordingly. By connecting glia cells, each glia cell can also transmit activated signals to other glia cells and regulate the glia effect of other glia cells. In order to simplify the calculation, all signals produced by glia cells propagate along specific directions of the glia chain. It means that the signal is transmitted from the first glia cell to the last one in the chain.

For RBM with glia chain, the output rule of hidden units is modified as follows:

$$h_j = \sigma(h_j^a + \alpha \cdot g_j)$$

where $h_j^a$ is the input value of the $j$-th node of the hidden layer. $g_j$ is the glia effect value of the corresponding glia cell. $\alpha$ is the weight coefficient of glia effect value. $\sigma$ is the sigmoid function. The weight coefficient $\alpha$ is preset, which can control the effect of glia effect value on hidden units. The calculation formula of $h_j^a$ is as follows:

$$h_j^a = \sum_i W_{ij}v_i + c_j$$

where $W_{ij}$ is the connection weight of the visual layer node $i$ and the hidden layer node $j$. $v_i$ is the state value of node $i$ in the visual layer. $c_j$ is the offset of node $j$. The glia effect value $g_j$ of node $j$ is defined as:

$$g_j(t) = \begin{cases} 1, & (h_j^a > \theta \cup g_{j-1}(t-1) = 1) \cap (d_j > T) \\ \beta g_{j-1}(t-1), & \text{other} \end{cases}$$

where $\theta$ is a predetermined threshold. $T$ is an unresponsive time threshold after activation, and $\beta$ represents the attenuation factor. Signals generated by activated glia are transmitted to the next glia cell. The activation of glia cells will depend on whether the input of the corresponding hidden unit reaches the predetermined threshold theta or whether the previous glia cells transmit signals to it. At the same time, the difference between the last activation time and the current time $d_j$ must be greater than the threshold $T$. If the glia cell is activated, it will send a signal to the next glia cell, otherwise it will not produce a signal, and its glia effect will gradually attenuate. $t$ represents the number of steps in which signals produced by activated glia are transmitted towards a specific direction, each step leading to the next adjacent glia cell.

Based on our previous work [28], this paper improves the way by which glia signals are transferred between glia cells. According to (3), the glia effect values of the glia cells of the nodes in the hidden layer of RBM are updated as follows Table 1. In the process, the newly activated node as the starting node can only transmit a signal once. Thus, the newly activated node that transmits the signal in the current cycle becomes an old activation node which doesn’t transmit signals in the next cycle.

Similar to DBN, the training process of DNB-GC includes two steps: pre-training and fine-tuning. The glia cell mechanism only acts on pre-training. The specific training process can be found in reference [29].

### C. SEQUENCE LEARNING AND DECISION MARGE LAYER

After extracting the high-level feature sequence by parallel DBN-GCs, CRF is employed to perform sequential learning and generate the predicted emotion label sequence.
TABLE 1. Pseudo codes for training the RBM with glia chain.

| Input: connection Weight Matrix \( W \), state value \( v \) in the visual layer, offset \( c_0 \), threshold \( \theta \), threshold \( T \), attenuation factor \( \beta \) |
| Output: glia effect value \( g \) of each node |

**Start training**

| Initialize \( t=0 \) |
| Calculate the input value \( h^* \) of each node in the hidden layer of RBM according to formula (2) |
| If \( h^* > \theta \) and \( d < T \) (for each node) |
| \( m=1 \)  \( (m \) is Cyclic termination judgment symbol. \( m=1 \) denotes the presence of newly activated nodes) |
| The corresponding node is marked as the activated node |
| The activation time point \( a=t \) |
| \( d=0 \)  \( (d \) is the difference between the activation time \( a \) and the current time \( t \) |
| End if |

| While \( (m=1) \) |
| \( m=0 \) |
| \( t=t+1 \) (The signals produced by glia cells of the newly activated nodes are transmitted to the next adjacent nodes, and these adjacent nodes are regarded as candidate activation nodes) |
| \( d=d+1 \) (For all nodes) |
| If \( d>T \) (for each unactivated node in these candidate activation nodes) |
| \( m=1 \) |
| The corresponding node is marked as the activated node |
| The activation time point \( a=t \) |
| \( d=0 \) |
| End if |

| End while |

| \( g=1 \) (for each activated nodes, \( g \) is the glia effect value) |
| \( g=\beta g(0) \) (for each unactivated node) |
| \( g(0)=\beta^0 \alpha \) (for each activated nodes, \( g(0) \) is the initial glia effect value for next update of glia) |
| \( g(0)=\beta g(0) \) (for each unactivated node) |

**End training**

CRF adopts the form of undirected graphical model, which combines the characteristics of maximum entropy model and hidden Markov model. Given the observation sequence \( X = \langle X_1, X_2, \ldots, X_n \rangle \) which corresponds to the high-level feature sequence derived from EEG signals in this emotion recognition task, a conditional probability distribution over label sequence \( Y = \{y_1, y_2, \ldots, y_{2^n}, \ldots, y_n\} \) which corresponds to emotion label sequence is defined as

\[
p(Y|X) = \frac{\exp\left(\sum_l \sum_k u_l f_k(y_{l-1}, y_l, X, l)\right)}{Z(X)}
\]

where \( f_k(y_{l-1}, y_l, X, l) \) is either a state feature function or a transition feature function. \( u_l \) is the weight coefficient. \( Z(X) \) is the normalization factor, which is defined as

\[
Z(X) = \sum_{Y \in Y^*} \exp\left(\sum_l \sum_k u_l f_k(y_{l-1}, y_l, X, l)\right)
\]

where \( Y^* \) is the set of all possible label sequences.

In the decision merge layer, k-nearest neighbor algorithm (KNN) is adopted to determine the emotion state of the observation sequence \( X \). Given the predicted emotion label sequence \( Y = \langle y_1, y_2, \ldots, y_n \rangle \), the emotion state \( Y^* \) of the observation sequence \( X \) can be defined as

\[
Y^* = \max_v s_v, \quad v = 1, 2, \ldots, m
\]

where \( s_v \) is the number of elements belonging to the \( v \)-th emotion state in \( Y \), and \( m \) is the total number of emotion states.

IV. EXPERIMENT AND DISCUSSION

A. DATABASE AND DATA PREPROCESSING

Datasets which are often used in EEG emotion experiment include AMIGOS [30], SEED [31], HR-EEG4EMO [32] and DEAP [33], etc. In this paper, DEAP, AMIGOS and SEED were adopted to validate the effectiveness of the proposed DBN-CRF framework.
The DEAP dataset recorded various physiological signals of 32 subjects (16 males and 16 females) when they watched 40 Music Video excerpts with different emotional markers. Among them, EEG signal was collected by 32-lead electrode cap of “10-20” international lead standard, and EEG signal of subjects was collected at 512 Hz sampling frequency. After watching the video, all subjects were asked to mark the valence, arousal, liking and dominance of the video they watched on a scale of 1–9.

The AMIGOS dataset recorded various physiological signals of 40 volunteers. In the first experiment, 40 volunteers watched a set of 16 short affective video extracts from movies. In the second experiment, 37 of the participants of the previous experiment watched a set of 4 long affective video extracts from movies. Each participant rated each video in valence, arousal, dominance, familiarity and liking, on a scale of 1–9 and selected basic emotions. Among them, EEG data was recorded using the EMOTIV Epoc, with 14 channel and a sampling frequency of 128 Hz. EEG channels according to the 10-20 system are: AF3, F7, F3, FC5, T7,P7, O1, O2, P8, T8, FC6, F4, F8, AF4. For the DEAP dataset and the AMIGOS dataset, only arousal, valence and dominance scales are employed for evaluation. All samples are divided into two classes according to arousal, valence and dominance scales. The thresholds of arousal, valence and dominance are both 5.

The SEED dataset contains EEG signals and eye movement signals from 15 subjects (7 males and 8 females) during watching emotional movie clips. The dataset contains 15 movie clips and each clip lasts about 4 minutes long. Each subject performed the experiment three times with an interval of about one week. The EEG signals are of 62 channels at a sampling rate of 1000Hz. Then, the data is down sampled to 200Hz. In this paper, only the 12 channels selected by reference [31] are employed. There are three emotional labels (−1 for negative, 0 for neutral and +1 for positive).

Among the physiological signals, only EEG signals are used to perform emotion recognition task. A sliding window is employed to divide the raw EEG signals into several segments, and there is no overlapping region between adjacent windows. As an independent sample, each segment inherits the label of the corresponding original sample. The window length of the three datasets is 12800, 50320 and 14895. In the experiments, 10-fold cross validation technique is adopted. For each volunteer, the corresponding samples are divided into 10 subsets, of which 9 subsets are assigned to the training set and the remaining one to the test set. Above process is repeated 10 times.

**B. FEATURE EXTRACTION**

In order to verify the effectiveness of the proposed model, three different types of features are employed in this paper. These features describe the salient information related to emotion states in time domain, frequency domain and time-frequency domain of EEG signals respectively [33]–[37], and Table 2 shows the detailed description.

The statistical measures from time domain include mean, variance, zero-crossing rate, and approximate entropy of all EEG channels. Power features from frequency domain include average PSD in theta (4–8 Hz), slow-alpha (8–10 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–45 Hz) bands for all EEG channels and difference of average PSD in theta, alpha, beta, and gamma bands for all EEG channel pairs between right and left scalp. HHS features from time-frequency domain include average values of squared amplitude and instantaneous frequency of HHS-based time–frequency representation in delta(1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–45 Hz) bands for all EEG channels.

**C. PERFORMANCE OF THE INDEPENDENT DBN-GC**

To analyze the performance of the improved DBN-GC, statistical measures, power features and HHS features are all used to train independent DBNs and independent DBN-GCs. Considering the similarity between DEAP and AMIGOS, these two datasets are used in the experiment. Each type of features is employed to train a DBN and a DBN-GC, respectively. This means that a DBN and its corresponding DBN-GC are trained by the same features and they have the same hyper-parameters. Furthermore, the parameters of a DBN and its corresponding DBN-GC, such as learning rate, are all set to the same value. Specifically, DBN1 and DBN-GC1 are trained by statistical measures of the DEAP dataset. DBN2 and DBN-GC2 are trained by power features of the DEAP dataset, and DBN3 and DBN-GC3 are trained by HHS features of the DEAP dataset. Similarly, DBN4 and
DBN-GC4 are trained by statistical measures of the AMIGOS dataset. DBN5 and DBN-GC5 are trained by power features of the AMIGOS dataset, and DBN6 and DBN-GC6 are trained by HHS features of the AMIGOS dataset. The above models implement the same emotion recognition task, and accuracy and F1-score are adopted to evaluate the recognition performance of these models.

In order to obtain the appropriate combination of hyper-parameters, we set different combinations of parameters to perform the experiment, and the parameter combination with the minimal recognition error is adopted. The hyper-parameters of DBNs and DBN-GCs are shown in Table 3 and Table 4. The training epoch is set to 200 and the batch size is 40.

The recognition results are shown in Figure 6 and Figure 7 with the form of mean value of recognition result. Two phenomena can be observed from Figure 6 and Figure 7. Firstly, regardless of which feature is used, the median values of accuracy and F1-score of DBN-GC model are obviously better than those of the corresponding DBN. Secondly, among the DBN-GC models trained by different features, DBN-GCs constructed by the power features acquires the highest accuracy (0.7293 on arousal and 0.7375 on valence for DEAP, 0.7568 on arousal and 0.7397 on valence for AMIGOS) and F1-score (0.6705 on arousal and 0.6732 on valence for DEAP, 0.7005 on arousal and 0.7089 on valence for AMIGOS).

The above results show that the glia chain can effectively improve the deep learning performance of DBN model. Specifically, the same hidden layer nodes can transmit information to each other through the glia chain. Thus, the correlation information between the adjacent hidden nodes, which is beneficial to pattern recognition task, can be caught by the DBN-GC model. When the DBN-GC model is applied to multi-channel EEG based emotion recognition, glia chain can mine and utilize inter-channel information, which is related to emotion states and helps to judge emotional state. In addition, the results also validate that the frequency domain feature of the three types of features contains most salient information regarding emotion states.

### TABLE 3. Hyper-parameters of DBNs and DBN-GCs.

|                  | Time domain | Frequency domain | Time-frequency domain |
|------------------|-------------|-----------------|----------------------|
|                  | DBN1 | DBN-GC1 | DBN2 | DBN-GC2 | DBN3 | DBN-GC3 |
| \( v_1 \)        | 128  | 128    | 216  | 216    | 320  | 320    |
| \( h_1 \)        | 280  | 280    | 365  | 365    | 515  | 515    |
| \( h_2 \)        | 195  | 195    | 240  | 240    | 320  | 320    |
| \( h_3 \)        | 65   | 65     | 105  | 105    | 140  | 140    |
| \( \alpha \)     | 2    | 2      | 2    | 2      | 2    | 2      |

Note: \( v_1 \) represents the input layer, \( \alpha \) represents the output layer and \( h_i \) represents the \( i \)-th hidden layer.

### TABLE 4. Hyper-parameters of DBNs and DBN-GCs.

|                  | Time domain | Frequency domain | Time-frequency domain |
|------------------|-------------|-----------------|----------------------|
|                  | DBN4 | DBN-GC4 | DBN5 | DBN-GC5 | DBN6 | DBN-GC6 |
| \( v_1 \)        | 56   | 56     | 126  | 126    | 140  | 140    |
| \( h_1 \)        | 160  | 160    | 310  | 310    | 345  | 345    |
| \( h_2 \)        | 125  | 125    | 175  | 175    | 220  | 220    |
| \( h_3 \)        | 50   | 50     | 80   | 80     | 105  | 105    |
| \( \alpha \)     | 2    | 2      | 2    | 2      | 2    | 2      |

Note: \( v_1 \) represents the input layer, \( \alpha \) represents the output layer and \( h_i \) represents the \( i \)-th hidden layer.

![FIGURE 6. Accuracies and F1 Score of the six models in two dimensions on the DEAP dataset.](image-url)
D. RESULTS OF THE DBN-CRF FRAMEWORK

To verify the effectiveness of the integrated framework in Figure 3, two kinds of methods are utilized for emotion recognition. One uses the DBN-GC-CRF model just described in Figure 3. The other makes use of the DBN-CRF model which uses several parallel DBNs instead of DBN-GCs of the framework.

In the experiments, only the frequency domain features are employed to construct the DBN-CRF framework and the DBN-GC-CRF framework. The hyper-parameters of DBN and DBN-GC in the two frameworks constructed by the DEAP dataset are the same as those of DBN2 and DBN-GC2 in Table 3, respectively. Similarly, DBN5 and DBN-GC5 in Table 4 are applied to emotion recognition task based on the AMIGOS dataset. When constructing DBN-CRF framework and DBN-GC-CRF framework with the SEED dataset, the node number of the three hidden layers in DBN and DBN-GC is 260, 180 and 45 respectively.

User-dependent model is built to perform leave-one-out cross validation, and user-independent model is built to perform leave-one-subject-out cross validation. In leave-one-out cross validation, for each volunteer, the corresponding samples are divided into 10 subsets, of which 9 subsets are assigned to the training set and the remaining one to the test set. Above process is repeated 10 times. In leave-one-subject-out cross validation, one subject set is assigned to the test set, and the remaining subject sets are assigned to the training set. Above process is repeated until each subject set is used as test set.

Table 5-7 show the results of leave-one-out cross validation on all datasets, and Table 8-10 show the results of leave-one-subject-out cross validation on all datasets. As can be seen from these tables, no matter which dataset is used, the DBN-GC-CRF framework outperforms the DBN-CRF framework in all dimensions. In addition, the recognition results of the DBN-GC-CRF framework are better than those of DBN-GC.

| TABLE 5. The mean accuracies in arousal dimension on all datasets for leave-one-out cross validation (%) |
|---------------------------------|
| leave-one-out cross validation (%) | DBN | DBN-CRF | DBN-GC | DBN-GC-CRF |
|---------------------------------|
| DEAP                           | 69.03 | 74.28 | 72.93 | 76.13 |
| AMIGOS                         | 70.35 | 74.89 | 75.68 | 77.96 |
| SEED                           | 78.54 | 81.21 | 80.63 | 83.46 |

| TABLE 6. The mean accuracies in valence dimension on all datasets for leave-one-out cross validation (%) |
|--------------------------------------------------|
| leave-one-out cross validation (%) | DBN | DBN-CRF | DBN-GC | DBN-GC-CRF |
|-----------------------------------|
| DEAP                              | 69.56 | 74.79 | 73.75 | 77.02 |
| AMIGOS                            | 71.56 | 75.22 | 73.97 | 78.13 |
| SEED                              | -     | -     | -     | -     |

| TABLE 7. The mean accuracies in dominance dimension on all datasets for leave-one-out cross validation (%) |
|-----------------------------------------------|
| leave-one-out cross validation (%) | DBN | DBN-CRF | DBN-GC | DBN-GC-CRF |
|-------------------------------------|
| DEAP                                | 68.88 | 74.25 | 73.16 | 76.45 |
| AMIGOS                              | 70.97 | 74.75 | 74.32 | 76.28 |
| SEED                                | -     | -     | -     | -     |

| TABLE 8. The mean accuracies in arousal dimension on all datasets for leave-one-subject-out cross validation (%) |
|--------------------------------------------------------|
| leave-one-subject-out cross validation (%) | DBN | DBN-CRF | DBN-GC | DBN-GC-CRF |
|---------------------------------------------|
| DEAP                                        | 68.21 | 73.55 | 71.48 | 74.59 |
| AMIGOS                                      | 68.57 | 73.11 | 74.31 | 75.75 |
| SEED                                        | 77.86 | 80.55 | 79.22 | 83.02 |

| TABLE 9. The mean accuracies in valence dimension on all datasets for leave-one-subject-out cross validation (%) |
|---------------------------------------------------------|
| leave-one-subject-out cross validation (%) | DBN | DBN-CRF | DBN-GC | DBN-GC-CRF |
|--------------------------------------------|
| DEAP                                       | 68.72 | 72.53 | 72.45 | 75.88 |
| AMIGOS                                     | 70.67 | 73.97 | 73.05 | 76.29 |
| SEED                                       | -     | -     | -     | -     |

Table 5-7 show the results of leave-one-out cross validation on all datasets, and Table 8-10 show the results of leave-one-subject-out cross validation on all datasets. As can be seen from these tables, no matter which dataset is used, the DBN-GC-CRF framework outperforms the DBN-CRF framework in all dimensions. In addition, the recognition results of the DBN-GC-CRF framework are better than those of DBN-GC.
FIGURE 8. Results of DBN-CRF and DBN-GC-CRF with different section number.

FIGURE 9. Comparison of DBN-CRF and DBN-GC-CRF.

TABLE 10. The mean accuracies in dominance dimension on all datasets for leave-one-subject-out cross validation (%).

| datasets for leave-one-subject-out cross validation (%) | DBN | DBN-CRF | DBN-GC | DBN-GC-CRF |
|--------------------------------------------------------|-----|---------|--------|------------|
| DEAP                                                   | 68.33 | 73.04 | 71.88 | 75.10 |
| AMIGOS                                                 | 69.48 | 73.55 | 73.67 | 75.24 |
| SEED                                                   | -     | -      | -     | -         |

E. PARAMETER ANALYSIS

As mentioned in Section 3.1, each sample is split into several sections. Then, a raw feature vector can be extracted from each section and is fed into the corresponding DBN or DBN-GC. In order to investigate the effect of section number on recognition performance and achieve the appropriate section number, the experiments with different section numbers are carried out. The experiments are performed on the DEAP dataset. Section number ranges from 1 to 20. When the section number is set to 1, the DBN-CRF framework is simplified to the DBN model and the DBN-GC-CRF framework is simplified to the DBN-GC model. Figure 8 shows the mean recognition accuracies (MRA) of the experiments.

As can be seen from Figure 8, when the section number increases from 1 and 11, the MRA on arousal as well as the MRA on valence also increase rapidly whether using DBN or DBN-GC. The results indicate that the contextual information of EEG signal contains salient information related to emotion states, which is beneficial to improvement of emotion recognition. Meanwhile, the proposed framework can effectively combine the feature extraction ability of DBN with the sequence data processing ability of CRF. Once the section number exceeds 11, the two MRAs increase very slowly and remain basically stable. With the increase of the section number, the computational complexity of the proposed framework will increase dramatically. Therefore, it is appropriate to set the section number to 11. In this case, the recognition rates of DBN-CRF and DBN-GC-CRF in arousal dimension are 0.7428 and 0.7613, respectively. And the recognition rates of DBN-CRF and DBN-GC-CRF in valence dimension are 0.7479 and 0.7702, respectively. In addition, regardless of the number of sections, the DBN-GC-CRF framework
TABLE 11. Details of several reported studies.

| Study                  | Feature                  | Classifier    | Arousal accuracy | F1-score | Valence accuracy | F1-score |
|------------------------|--------------------------|---------------|------------------|----------|------------------|----------|
| Tripathi et al.[20]    | Statistical parameters   | DNN           | 0.7313           | —        | 0.7578           | —        |
| Tripathi et al.[20]    | Statistical parameters   | CNN           | 0.7336           | —        | 0.8141           | —        |
| Li et al.[38]          | Grid-like frames         | CNN/RNN       | 0.7412           | —        | 0.7206           | —        |
| Li et al.[38]          | Grid-like frames         | DBN-SVM       | 0.6420           | —        | 0.5840           | —        |
| Wang and Shang [14]    | DBN features             | DBN           | 0.5120           | —        | 0.6090           | —        |
| Koelstra et al. [33]   | Power spectral features  | Gaussian naive Bayes | 0.6200 | 0.5830 | 0.5760 | 0.5630 |
| The proposed method    | Power features           | DNB-GC-CRF    | 0.7613           | 0.7016   | 0.7702           | 0.7089   |

outperforms the DBN-CRF framework in both arousal and valence dimensions. This indicates that the glia chain of DBN-GC is also effective in the integrated framework.

Considering the similarity between DBN-CRF and DBN-GC-CRF, hold-out validation is utilized to evaluate the two models for further comparison. The section number is set to 11. All samples are randomly divided into training, validation, and test sets, with respective proportions of 8:1:1 for each label. Figure 9 describes the results in the validation set. By using the same data, the validation accuracy of DBN-GC-CRF increases faster both in arousal and valence dimensions, and the accuracy of DBN-GC-CRF is higher than that of DBN-CRF with a similar network structure using the same epoch.

F. RESULTS COMPARISON

Finally, the proposed method is compared with other deep learning methods which also use the DEAP dataset, and the detail results are described in Table 11. As shown in Table 11, the highest recognition accuracies both in arousal and valence dimensions can be achieved by the DBN-GC-CRF model, with the exception of the results of reference [20]. The mean recognition accuracy of 0.7702 on valence provided by the DBN-GC-CRF model is lower than that of 0.8141 reported in reference [20]. The reason may be that the reference [20] trains a subject-related model, which is only utilized to recognize the samples belonging to the same subject. In addition, the F1-scores of the DBN-GC-CRF model are superior to 0.5830 and 0.5630 of reference [33] with the same kind of features.

V. CONCLUSION

In this paper, we have proposed an ensemble deep learning framework, which integrates parallel DBN-GCs and CRF, for emotion classification based on multi-channel EEG signals. Specifically, a raw feature vector sequence is firstly obtained from the multi-channel EEG signals of the current sample by dividing the EEG signals into several segments. Then, each raw feature vector is fed into the corresponding DBN-GC model to generate the high level representations. Thus, a high-level feature sequence is achieved. According to the high-level feature sequence, the CRF model generates the predicted emotion label sequence. Finally, the KNN based decision merge layer is adopted to determine the emotion state. The proposed method is evaluated on datasets AMIGOS, SEED and DEAP. The results indicate that the proposed method outperforms most of the previous studies with the same EEG data set. The main contributions of this work lie in two aspects. Firstly, the DBN-GC can mine inter-channel or cross-modal correlation information of EEG signals through the glia chain, which includes salient information related to emotion states. Furthermore, the CRF based model structure can catch the long-term dependencies and contextual information of EEG sequences and is beneficial to the improvement of recognition performance. Especially, this is the first attempt to apply the conditional random field methodology to deep belief networks.

REFERENCES

[1] P. Marrero-Fernández, A. Montoya-Padrón, A. Jaume-i-Capó, and J. M. B. Rubio, "Evaluating the research in automatic emotion recognition," IEEE Tech. Rev., vol. 31, no. 3, pp. 220–232, May 2014.
[2] M. Chen, X. He, J. Yang, and H. Zhang, "3-D convolutional recurrent neural networks with attention model for speech emotion recognition," IEEE Signal Process. Lett., vol. 25, no. 10, pp. 1440–1444, Oct. 2018.
[3] Z.-T. Liu, X. Xie, M. Wu, W.-H. Cao, Y. Mei, and J.-W. Mao, "Speech emotion recognition based on an improved brain emotion learning model," Neurocomputing, vol. 309, pp. 145–156, Oct. 2018.
[4] Y. Huang, J. Yang, P. Liao, and J. Pan, "Fusion of facial expressions and EEG for multimodal emotion recognition," Comput. Intell. Neurosci., vol. 2017, no. 2, pp. 1–8, 2017.
[5] P. D. Ross, L. Polson, and M.-H. Grosbras, "Developmental changes in emotion recognition from full-light and point-light displays of body movement," PLoS ONE, vol. 7, no. 9, Sep. 2012, Art. no. e44815.
[6] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from EEG data using machine learning approach," Neurocomputing, vol. 129, pp. 94–106, Apr. 2014.
[7] A. Al-Nafjan, M. Hosny, Al-Ohali, and A. Al-Wabil, “Review and classification of emotion recognition based on EEG brain-computer interface system research: A systematic review,” Appl. Sci., vol. 7, no. 12, p. 1259, Nov. 2017.
[8] R. M. Mehmood and H. J. Lee, “Towards emotion recognition of EEG brain signals using Hjorth parameters and SVM,” Adv. Sci. Technol. Lett., Biosci. Med. Res., vol. 91, pp. 24–27, Apr. 2015. [Online]. Available: http://sersc.org/proceedings/vol91_2015.php
[9] R. M. Mehmood and H. J. Lee, “EEG based emotion recognition from human brain using Hjorth parameters and SVM,” Int. J. Bio-Sci. Bio-Tech., vol. 7, no. 3, pp. 23–32, Jun. 2015.
[10] H. J. Yoon and S. Y. Chung, “EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm,” *Comput. Biol. Med.*, vol. 43, no. 12, pp. 2230–2237, Dec. 2013.

[11] N. Jadhav, R. Manthalkar, and Y. Joshi, “Electroencephalography-based emotion recognition using gray-level co-occurrence matrix features,” in *Proc. Int. Conf. Comput. Vis. Image Process.*, Dec. 2016, pp. 335–343.

[12] A. Samara, M. L. R. Menezes, and L. Galway, “Feature extraction for emotion recognition and modelling using neurophysiological data,” in *Proc. 15th Int. Conf. Ubiquitous Comput. Commun. Int. Symp. Cyberspace Secur. (IUCC-CSS)*, Dec. 2016, pp. 138–144.

[13] N. Thammasan, K. Moriyama, K.-I. Fukui, and M. Numao, “Familiarity effects in EEG-based emotion recognition,” *Brain Inf.*, vol. 4, no. 1, pp. 39–50, Mar. 2017.

[14] D. Wang and Y. Zhang, “Modeling physiological data with deep belief networks,” *Int. J. Inf. Educ. Technol.*, vol. 3, no. 5, pp. 505–511, 2013.

[15] Y.-H. Kwon, S.-B. Shin, and S.-D. Kim, “Electroencephalography based fusion two-dimensional (2D)-convolution neural networks (CNN) model for emotion recognition system,” *Sensors*, vol. 18, no. 5, p. 1383, Apr. 2018.

[16] H. Chao, L. Dong, Y. Liu, and B. Lu, “Emotion recognition from multiband EEG signals using CapsNet,” *Sensors*, vol. 19, no. 9, p. 2212, May 2019.

[17] A.-R. Mohamed, G. E. Dahl, and G. Hinton, “Acoustic modeling using deep belief networks,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 20, no. 1, pp. 14–22, Jan. 2012.

[18] P. Tamilisivam and P. Wang, “Failure diagnosis using deep belief learning based health state classification,” *Rel. Eng. Syst. Saf.*, vol. 115, pp. 124–135, Jul. 2013.

[19] X. Li, D. Song, P. Zhang, Y. Hou, and B. Hu, “Deep fusion of multi-channel neurophysiological signal for emotion recognition and monitoring,” *Int. J. Data Mining Bioinformatic*, vol. 18, no. 1, p. 1, 2017.

[20] S. Tripathi, S. Acharya, and R.-D. Sharma, “Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset,” in *Proc. 29th AAAI Conf. Innov. Appl. Artif. Intell.*, Feb. 2017, pp. 4746–4752.

[21] C. Sutton, K. Rohaninanenash, and A. McCallum, “Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data,” in *Proc. 21st Int. Conf. Mach. Learn. (ICML)*, 2004, pp. 282–289.

[22] C. Sutton and A. McCallum, “An introduction to conditional random fields for relational learning,” in *Introduction to Statistical Relational Learning*, L. Getoor and B. Taskar, Eds. Cambridge, MA, USA: MIT Press, 2006.

[23] N. Banda and A. Engelbrecht, “Multimodal emotion recognition using tichannel EEG data and DBN-GC-based ensemble deep learning framework,” in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 408–413.

[24] G. E. Hinton, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[25] G. E. Hinton, S. Osindero, and Y.-W. Teh, “A fast learning algorithm for deep belief nets,” *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

[26] J. S. Zhou, X. Y. Dai, and C. Y. Yin, “Automatic recognition of Chinese organization name based on cascaded conditional random fields,” *Chin. J. Electron.*, vol. 34, no. 5, pp. 804–809, 2006.

[27] X. N. Mao, S. K. He, and S. C. Bao, “Chinese word segmentation and named entity recognition based on conditional random fields,” in *Proc. LICONLP*, 2008, pp. 90–93.

[28] H. Chao, H. Zhi, L. Dong, and Y. Liu, “Recognition of emotions using multichannel EEG data and DBN-GC-based ensemble deep learning framework,” *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–11, Dec. 2018.

[29] Z. Q. Geng and Y. K. Zhang, “An improved deep belief network inspired by glia chains,” *Acta Automatica Sinica*, vol. 42, no. 6, pp. 943–952, 2016.

[30] J. A. Miranda Correa, M. K. Abadi, N. Sebe, and I. Patras, “AMIGOS: A dataset for affect, personality and mood research on individuals and groups,” *IEEE Trans. Affective Comput.*, to be published.

[31] W.-L. Zheng and B.-L. Lu, “Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks,” *IEEE Trans. Austron. Mental Dev.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.

[32] H. Becker, J. Fleuret, P. Guilloite, F. Wendling, I. Merlet, and L. Albera, “Emotion recognition based on high-resolution EEG recordings and reconstructed brain sources,” *IEEE Trans. Affective Comput.*, to be published.

[33] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, “DEAP: A database for emotion analysis :Using physiological signals,” *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012.

[34] M. Khezri, M. Firoozabadi, and A. R. Sharafat, “Reliable emotion recognition system based on dynamic adaptive fusion of forehead biopotentials and physiological signals,” *Comput. Methods Programs Biomed.*, vol. 122, no. 2, pp. 149–164, Nov. 2015.

[35] J. Atkinson and D. Campos, “Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers,” *Expert Syst. Appl.*, vol. 47, pp. 35–41, Apr. 2016.

[36] C. Li, C. Xu, and Z. Feng, “Analysis of physiological for emotion recognition with the IRS model,” *Neurocomputing*, vol. 178, pp. 103–111, Feb. 2016.

[37] Z. Yin, M. Zhao, Y. Wang, J. Yang, and J. Zhang, “Recognition of emotions using multimodal physiological signals and an ensemble deep learning model,” *Comput. Methods Programs Biomed.*, vol. 140, pp. 93–110, Mar. 2017.

[38] X. Li, D. Song, P. Zhang, G. Yu, Y. Hou, and B. Hu, “Emotion recognition from multi-channel EEG data through Convolutional Recurrent Neural Network,” in *Proc. IEEE Int. Conf. Bioinformatics Biomed. (BIBM)*, Dec. 2016, pp. 352–359.