Emotion Annotation Using Hierarchical Aligned Cluster Analysis

Wei-Ye Zhao[1], Sheng Fang[1], Ting Ji[1], Qian Ji[1], Wei-Long Zheng[1], and Bao-Liang Lu[1,2,3(✉)]

[1] Department of Computer Science and Engineering,
Center for Brain-like Computing and Machine Intelligence, Shanghai, China
{andylaw,weilong}@sjtu.edu.cn
[2] Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Shanghai, China
[3] Brain Science and Technology Research Center,
Shanghai Jiao Tong University, Shanghai, China
bllu@sjtu.edu.cn

Abstract. The correctness of annotation is quite important in supervised learning, especially in electroencephalography(EEG)-based emotion recognition. The conventional EEG annotations for emotion recognition are based on the feedback like questionnaires about emotion elicitation from subjects. However, these methods are subjective and divorced from experiment data, which lead to inaccurate annotations. In this paper, we pose the problem of annotation optimization as temporal clustering one. We mainly explore two types of clustering algorithms: aligned clustering analysis (ACA) and hierarchical aligned clustering analysis (HACA). We compare the performance of questionnaire-based, ACA-based, HACA-based annotation on a public EEG dataset called SEED. The experimental results demonstrate that our proposed ACA-based and HACA-based annotation achieve an accuracy improvement of 2.59% and 4.53% in average, respectively, which shows their effectiveness for emotion recognition.

Keywords: Neural data analysis · Time series analysis · EEG annotations · Emotion recognition

1 Introduction

Emotion is a subjective, conscious experience when people are faced with internal or external stimuli, and it is crucial for natural communication, decision making and human-machine interface. Emotion recognition can be performed through facial movement, voice, speech, text, and physiological signals [5,9]. Among these approaches, emotion recognition from electroencephalography (EEG) has attracted increasing interest [12,14,15].

In many supervised learning tasks, accurate training data annotations are the key factors to obtain ideal learning performance [4]. It was shown that the

D. Liu et al. (Eds.): ICONIP 2017, Part IV, LNCS 10637, pp. 572–580, 2017.
https://doi.org/10.1007/978-3-319-70093-9_60

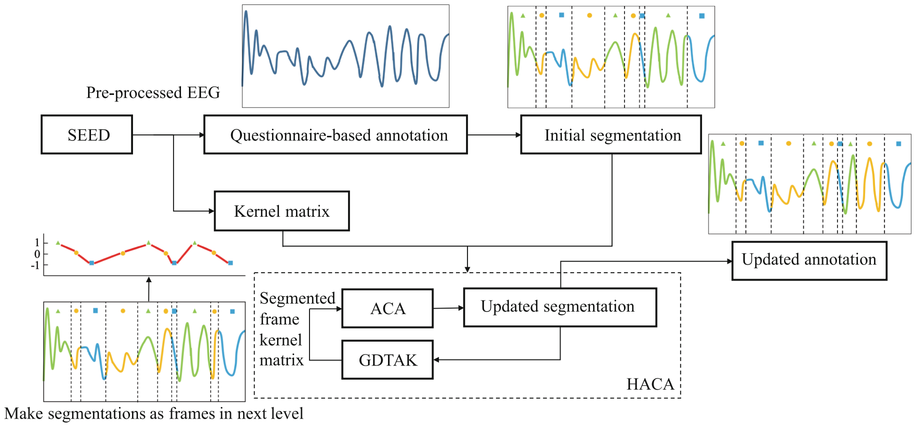

Fig. 1. The main processing steps to optimize questionnaire-based annotations with ACA and HACA.

presence of mislabeled training data, even on a very small scale, can deteriorate the performance of classifiers in a broad range of classification problems [3,8] and inaccurate annotations are even more harmful than noisy training features [18]. However, obtaining annotations is error-prone since the process is inherently subjective [1]. For example, EEG signals are often annotated according to testers' evaluation of their emotion states in affective brain-computer interfaces [7].

Two categories of methods to eliminate mislabeled training examples have been widely studied recently: designing noise robust supervised learning algorithm and filtering mislabeled data [4]. Some classification methods in the presence of label noise were proposed [2]. As for filtering mislabeled data, a simple approach is to remove low-quality data in preprocessing stage, whereas such strategy might remove useful instances. Guan and colleagues proposed a nearest neighbor editing method to remove noisy samples from the training dataset [3].

To our best knowledge, there are limited studies reported in the literature dealing for emotion annotation optimization using temporal clustering algorithms. In our paper, we explore aligned clustering analysis (ACA) [16] and hierarchical aligned clustering analysis (HACA) [17]. HACA is a hierarchical extension of ACA using generalized dynamic time alignment kernel (GDTAK). The performance are evaluated on a public EEG dataset named SEED, which gives EEG annotations according to test subjects' evaluation on their emotional states (questionnaire-based annotations). The process for optimizing questionnaire-based EEG annotations is shown in Fig. 1.

2 Methods

2.1 Dynamic Time Alignment Kernel

Temporal clustering method utilizes a distance metric that is capable of matching time series points even for series of different length. Shimodaira *et al.* [11] proposed Dynamic Time Alignment Kernel (DTAK) as an efficient metric between time sequences. Given two sequences $\boldsymbol{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_{n_x}]$ and $\boldsymbol{Y} = [\boldsymbol{y}_1, \dots, \boldsymbol{y}_{n_y}]$, and their distance kernel matrix $\boldsymbol{K} \in \mathbb{R}^{n_x \times n_y}$ $\left(\kappa_{i,j} = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2}{2\sigma^2}\right)\right)$, DTAK is defined by recursively computing the similarity between the two time sequences.

$$\tau(\boldsymbol{X}, \boldsymbol{Y}) = \frac{\omega_{n_x,n_y}}{n_x + n_y}, \omega_{i,j} = \max \begin{cases} \omega_{i-1,j} + \kappa_{i,j} \\ \omega_{i-1,j-1} + 2\kappa_{i,j} \\ \omega_{i,j-1} + \kappa_{i,j} \end{cases} \tag{1}$$

2.2 Aligned Clustering Analysis

Aligned Clustering Analysis [16] extends kernel k-means for temporal clustering. Given an EEG sequence $\boldsymbol{X} \in \mathbb{R}^{dim \times n_x}$, instead of minimizing the sum of distances, ACA minimizes the energy function:

$$\boldsymbol{J}_{aca}(\boldsymbol{G}, \boldsymbol{s}) = \sum_{c=1}^{k} \sum_{i=1}^{m} g_{ci} \underbrace{\|\psi(\boldsymbol{X}_{[s_i, s_{i+1}]}) - \boldsymbol{z}_c\|^2}_{dist_\psi^2(\boldsymbol{Y}_i, \boldsymbol{z}_c)} = \|[\psi(\boldsymbol{Y}_1), \dots, \psi(\boldsymbol{Y}_m)] - ZG\|_F^2$$
$$\text{s.t.} \qquad \boldsymbol{G}^T \boldsymbol{1}_k = \boldsymbol{1}_m, \tag{2}$$

where $\psi(\cdot)$ denotes a mapping of the sequence into a feature space, $\boldsymbol{G} \in \{0,1\}^{k \times m}$ is the EEG segment indicator matrix with k classes and m segments ($g_{ci} = 1$ only when segment i belongs to class c), $\boldsymbol{s} \in \mathbb{R}^{m+1}$ is a vector contains start and end position of each EEG segment ($s_{i+1} - s_i \in [1, n_{max}]$), $\boldsymbol{Y}_i = \boldsymbol{X}_{[s_i, s_{i+1}]}$ denotes an EEG segment, and $\boldsymbol{z}_c$ is the geometric centroid for class c.

With the ability to handle variable length features and DTAK to calculate the distance between the segment and the class centroid, ACA is suitable for EEG sequence analysis. An effective algorithm called Dynamic Programming Search (DPSearch) [17] is proposed to minimize ACA energy function. Process of using ACA to optimize questionnaire-based annotations is described in Algorithm 1.

2.3 Hierarchical Aligned Clustering Analysis

In this section, we introduce HACA optimization (see Algorithm 2), which extends ACA with hierarchical implementation [17] to optimize questionnaire-based annotations. The core idea is using ACA to find an optimal segmentation in a smaller segment length constraint n_{max1} in the first level, then propagating the solution to the second level by treating segments in the first level as temporal series frames. After calculating the "frame" kernel matrix in the second level, ACA is applied in the second level to gain the segmentation of longer temporal scale in the constraint of n_{max2}. The generalized dynamic time alignment kernel (GDTAK) is used for calculating kernel matrix based on segmented frames.

Note that both ACA and HACA share the same energy function minimizing algorithm, ACA minimizes distance among sampling points, whereas HACA

Algorithm 1. ACA Optimization

ACA Optimization ($\boldsymbol{X}, \boldsymbol{label}$);
Parameter: EEG segment length constraint n_{max} and number of emotion classes k
Input : EEG sequence $\boldsymbol{X}$ and questionnaire-based EEG annotations $\boldsymbol{label}$
Output : Optimized EEG annotations $\boldsymbol{label}_{new}$
Construct indicator matrix $\boldsymbol{G}$ and segment vector $\boldsymbol{s}$;
Construct kernel matrix $\boldsymbol{K}$ from distance matrix of $\boldsymbol{X}$;
$\boldsymbol{s}_{new} \leftarrow \boldsymbol{s}$;
do
  $\boldsymbol{s} \leftarrow \boldsymbol{s}_{new}$;
  Use $DPSearch(\boldsymbol{G}, \boldsymbol{s}, \boldsymbol{K})$ to obtain $\boldsymbol{s}_{new}$ and $\boldsymbol{G}_{new}$;
while $\boldsymbol{s}_{new} \neq \boldsymbol{s}$;
$m + 1 \leftarrow length(\boldsymbol{s}_{new})$;
for $i = 1$ *to* m **do**
  for $j = \boldsymbol{s}_{new}(i)$ *to* $\boldsymbol{s}_{new}(i+1)$ **do**
    Create $\boldsymbol{label}_{new}(j) \leftarrow \boldsymbol{G}(i)$;
  end
end

Algorithm 2. HACA Optimization

HACA Optimization ($\boldsymbol{X}, \boldsymbol{label}$);
Parameter: EEG segment length constraint in 1^{st} level n_{max1}, EEG segment length constraint in 2^{nd} level n_{max2} and number of emotion classes k
Input : EEG sequence $\boldsymbol{X}$ and questionnaire-based EEG annotations $\boldsymbol{label}$
Output : Optimized EEG annotations $\boldsymbol{label}_{new}$
Construct indicator matrix $\boldsymbol{G}$ and segment vector $\boldsymbol{s}$;
Construct kernel matrix $\boldsymbol{K}$ from distance matrix of $\boldsymbol{X}$;
Use ACA to optimize segmentation in first level:$(\boldsymbol{G}, \boldsymbol{s}) \leftarrow ACA(\boldsymbol{G}, \boldsymbol{s}, \boldsymbol{K})$;
Construct segmented frame kernel matrix $\boldsymbol{T} \leftarrow GDTAK(\boldsymbol{K}, \boldsymbol{s})$;
for $i = 1$ *to* $length(\boldsymbol{s}) - 1$ **do**
  Create $\boldsymbol{s}_{hierarchy}(i) \leftarrow i$, $\boldsymbol{G}_{hierarchy} \leftarrow \boldsymbol{G}_j$;
end
Create $\boldsymbol{s}_{hierarchy}(length(\boldsymbol{s})) \leftarrow length(\boldsymbol{s})$;
Use ACA to optimize segmentation in second level:
$(\boldsymbol{G}_{hierarchy}, \boldsymbol{s}_{hierarchy}) \leftarrow ACA(\boldsymbol{G}_{hierarchy}, \boldsymbol{s}_{hierarchy}, \boldsymbol{T})$;
for $j = 1$ *to* $length(\boldsymbol{s}_{hierarchy}) - 1$ **do**
  Create $\boldsymbol{s}_{new}(j) \leftarrow \boldsymbol{s}(\boldsymbol{s}_{hierarchy}(j))$, $\boldsymbol{G}_{new}(j) \leftarrow \boldsymbol{G}_{hierarchy}(j)$;
end
Create $\boldsymbol{s}_{new}(length(\boldsymbol{s}_{hierarchy})) \leftarrow \boldsymbol{s}(length(\boldsymbol{s}))$;
Create $\boldsymbol{label}_{new}$ based on $\boldsymbol{G}_{new}$ and $\boldsymbol{s}_{new}$;

minimizes distance among ACA optimized segments. Therefore, HACA replace DTAK with GDTAK. As for more details about ACA and HACA algorithms, please refer to the previous study [17].

3 Experiment and Result

3.1 EEG Dataset

We evaluate the performance of these approaches on a public dataset, SEED dataset[1] [13], which consists of stimuli and EEG data. There are 15 emotional film clips which elicit three emotions: positive, neutral and negative. For each session, a 5-s hint for starting is given before clip and a 45-s self-assessment and a 15-s rest after. There are totally 45 subjects participating in the experiments, who are required to elicit their own corresponding emotions while watching the clips. EEG data are recorded with a 62-electrode cap according to the international 10–20 system using ESI Neuroscan system.

3.2 Data Preprocessing and Feature Extraction

For data preprocessing, we apply bandpass filter between 1 Hz and 75 Hz to process raw EEG data. 62-channel EEG signals are further down-sampled to 200 Hz to reduce the computational complexity. Then EEG features are calculated using short-term Fourier transform from pre-processed EEG segments with non-overlapping 1-s time window.

For feature extraction, differential entropy (DE) features as EEG features are used for emotion recognition [13], which shows superior performance when compared to the conventional power spectral density features. For a fixed length EEG segment, DE is equivalent to the logarithm energy spectrum in a certain frequency band [10]. Therefore, DE features can be calculated in five frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (14–31 Hz), and gamma (31–50 Hz). The total dimension of a 62-channel EEG segment is 310.

3.3 Questionnaire-Based Annotations Optimization using HACA

In this section, we carry out experiments to evaluate the effectiveness of ACA (HACA) for optimizing questionnaire-based annotations. There is problem that if the annotations are optimized by temporal clustering methods, relabeled classes have higher similarity within each class, classification on these EEG data could be easier, which will cause inaccurate results. In order to avoid this problem, we use first 9 sessions for training and rest 6 sessions for testing. Only training data are clustered by ACA and HACA, and testing data annotations remain questionnaire-based to compare the performance of different annotation methods. Unlike typical clustering problem which determines the intrinsic grouping in a collection of unlabeled data, we make use of questionnaire-based annotations initialization to rectify the mislabeled annotations. Figure 2 shows that HACA annotation reaches the peak accuracy when $n_{max1} = 15$ and $n_{max2} = 4$.

Figure 3 shows the average accuracies of questionnaire-based (original), ACA-based, HACA-based annotation methods for total 45 experiments. The

[1] http://bcmi.sjtu.edu.cn/~seed/index.html.

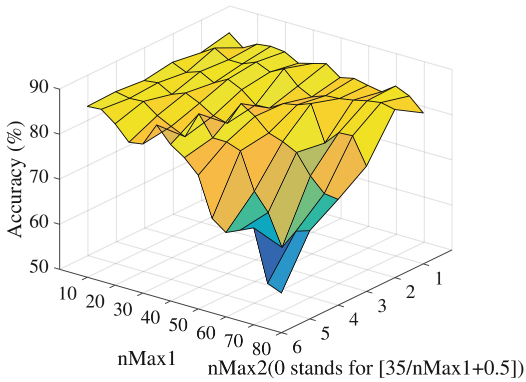

Fig. 2. Classification accuracy surface of HACA method using questionnaire-based initiation in 1^{st} level and 1^{st} level annotation optimization-based initiation in 2^{nd} level.

questionnaire-based annotation achieves the moderate performance with an average accuracy of only 85.13%. HACA outperforms the other two approaches with an average accuracy of 89.70%.

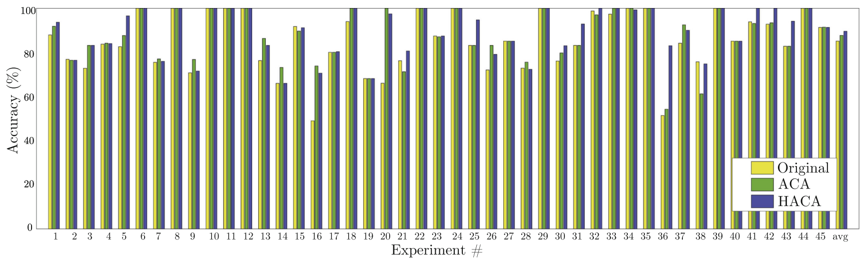

Fig. 3. Accuracies of questionnaire-based annotation (original), ACA-based annotation, HACA-based annotation on 45 experiments.

In supervised learning, the classification accuracy of mislabeled testing data will be aberrantly low [6], based on this observation, for investigating the relationship between mislabeled samples and classification accuracy, we employ 5-fold cross validation, which uses EEG sequences from 3 consecutive emotional film clips as testing data and the rest from 12 film clips as training data to compute the average classification accuracy ($curve\boldsymbol{Red}$) of every sampling point over 45 experiments. The annotation change ratios ($curve\boldsymbol{Blue}$) of every sampling point over 45 experiments are calculated after HACA annotation optimization. As shown in Fig. 4, the sampling points with low classification accuracy have high annotation change ratio. This comparison result shows HACA annotation optimization can efficiently rectify inaccurate annotations of EEG data.

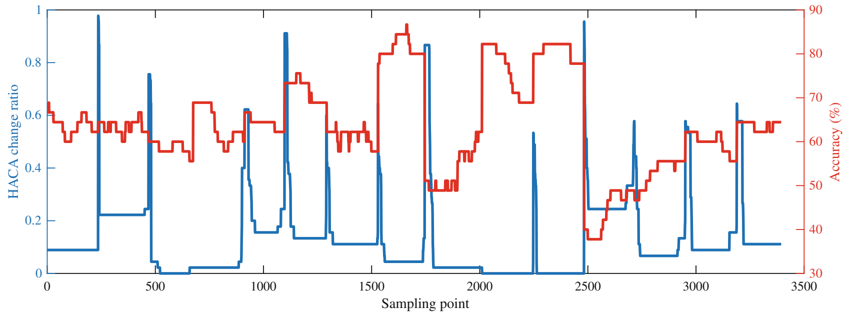

Fig. 4. Comparison between annotation change ratio and average classification accuracy. (Color figure online)

Figure 5 shows part of EEG annotations of 45 experiments using HACA, which indicates annotation variations across different experiments. To be specific, the classification accuracy of Experiment 16 and Experiment 20 improves from 48.99% to 70.59% and from 66.04% to 97.49%, respectively, which demonstrate the effectiveness of HACA annotation optimization compared with conventional questionnaire-based annotation.

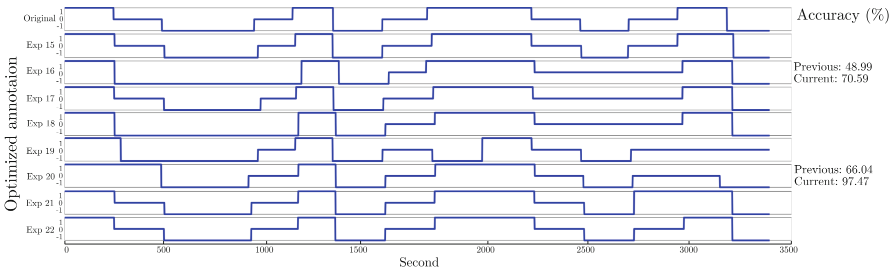

Fig. 5. Part of HACA EEG annotation optimization results of 45 experiments.

The average change ratios for three labels are shown in Table 1, in which negative emotions are often mislabeled as neutral (10%) and vice versa (10%).

Table 1. Average label change ratio of 45 experiments for HACA annotation

Rectified	Original		
	Positive	Neutral	Negative
Positive	0.90	0.03	0.04
Neutral	0.05	0.87	0.10
Negative	0.05	0.10	0.86

These results are consistent with previous finding that negative emotion is often confused with neutral emotion for EEG [7]. In summary, the experimental results demonstrate the efficiency of HACA approach for optimizing questionnaire-based EEG annotations.

4 Conclusion and Future Work

In this paper, we have adopted two temporal clustering methods, ACA and HACA to optimize the original questionnaire-based annotations. The experimental results have shown that ACA-based annotation achieves 2.59% average accuracy improvement and HACA-based annotation performs better with a 4.53% improvement. The comparison of HACA change ratio and performance improvement over original classification accuracy shows our annotation methods rectify inaccurate annotations.

Our future work will focus on testing performance of ACA and HACA annotation methods in larger real-world data sets, and comparing them with other existing works on mislead sampling filtering.

Acknowledgments. This work was supported in part by grants from the National Key Research and Development Program of China (Grant No. 2017YFB1002501), the National Natural Science Foundation of China (Grant No. 61673266), the Major Basic Research Program of Shanghai Science and Technology Committee (Grant No. 15JC1400103), ZBYY-MOE Joint Funding (Grant No. 6141A02022604), and the Technology Research and Development Program of China Railway Corporation (Grant No. 2016Z003-B).

References

1. Brodley, C.E., Friedl, M.A.: Identifying mislabeled training data. J. Artif. Intell. Res. **11**, 131–167 (1999)
2. Frénay, B., Verleysen, M.: Classification in the presence of label noise: a survey. IEEE Trans. Neural Netw. Learn. Syst. **25**(5), 845–869 (2014)
3. Guan, D., Yuan, W., Lee, Y.K., Lee, S.: Nearest neighbor editing aided by unlabeled data. Inf. Sci. **179**(13), 2273–2282 (2009)
4. Guan, D., Yuan, W., Ma, T., Khattak, A.M., Chow, F.: Cost-sensitive elimination of mislabeled training data. Inf. Sci. **402**, 170–181 (2017)
5. Kim, K.H., Bang, S.W., Kim, S.R.: Emotion recognition system using short-term monitoring of physiological signals. Med. Biol. Eng. Comput. **42**(3), 419–427 (2004)
6. Lam, C.P., Stork, D.G.: Evaluating classifiers by means of test data with noisy labels. In: IJCAI, pp. 513–518 (2003)
7. Lu, Y., Zheng, W.L., Li, B., Lu, B.L.: Combining eye movements and EEG to enhance emotion recognition. In: IJCAI, pp. 1170–1176 (2015)
8. Sáez, J.A., Galar, M., Luengo, J., Herrera, F.: A first study on decomposition strategies with data with class noise using decision trees. In: Corchado, E., Snášel, V., Abraham, A., Woźniak, M., Graña, M., Cho, S.-B. (eds.) HAIS 2012. LNCS, vol. 7209, pp. 25–35. Springer, Heidelberg (2012). doi:10.1007/978-3-642-28931-6_3

9. Schuller, B., Rigoll, G., Lang, M.: Hidden Markov model-based speech emotion recognition. In: Proceedings of the 2003 International Conference on Multimedia and Expo, vol. 1, p. I–401. IEEE (2003)
10. Shi, L.C., Jiao, Y.Y., Lu, B.L.: Differential entropy feature for EEG-based vigilance estimation. In: 35th Annual International Conference of the IEEE, Engineering in Medicine and Biology Society, pp. 6627–6630. IEEE (2013)
11. Shimodaira, H., Noma, K.I., Nakai, M., Sagayama, S., et al.: Dynamic time-alignment kernel in support vector machine. In: NIPS, vol. 2, pp. 921–928 (2001)
12. Wang, X.W., Nie, D., Lu, B.L.: Emotional state classification from EEG data using machine learning approach. Neurocomputing **129**, 94–106 (2014)
13. Zheng, W.L., Lu, B.L.: Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE Trans. Auton. Ment. Dev. **7**(3), 162–175 (2015)
14. Zheng, W.L., Zhu, J.Y., Lu, B.L.: Identifying stable patterns over time for emotion recognition from EEG. IEEE Trans. Affect. Comput. (2017). doi:10.1109/TAFFC.2017.2712143
15. Zheng, W.L., Zhu, J.Y., Peng, Y., Lu, B.L.: EEG-based emotion classification using deep belief networks. In: IEEE International Conference on Multimedia and Expo, pp. 1–6. IEEE (2014)
16. Zhou, F., De la Torre, F., Hodgins, J.K.: Aligned cluster analysis for temporal segmentation of human motion. In: 8th IEEE International Conference on Automatic Face & Gesture Recognition, pp. 1–7. IEEE (2008)
17. Zhou, F., De la Torre, F., Hodgins, J.K.: Hierarchical aligned cluster analysis for temporal clustering of human motion. IEEE Trans. Pattern Anal. Mach. Intell. **35**(3), 582–596 (2013)
18. Zhu, X., Wu, X.: Class noise vs. attribute noise: a quantitative study. Artif. Intell. Rev. **22**(3), 177–210 (2004)