International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

AFFINITY PROPAGATION BASED STRATIFIED SAMPLING FOR EMOTION DETECTION

SANJAI H(1903049)
Department Of Computer Science And Engineering
PSN College Of Engineering And Technology
Tirunelveli.

Dr.S.Radhakrishnan,
Associate Professor,
Department Of Computer Science And Engineering,
PSN College Of Engineering And Technology,
TIrunelveli.

Abstract— **Emotions were considered an important component. Everyday lives need emotions on a regular basis. Electroencephalogram (EEG)-based emotion identification was gaining popularity quickly. Signals from electroencephalograms (EEGs) are one of the key resources. The biggest benefit of using EEG signals is that they accurately represent real feelings and it could be prepared by computer systems. EEG is a physiological marker that may be captured from the brain activity within the context of scalp-transmitted brain waves which was used to collect brain signals. With this test, waves representing the brain's activity are recorded. The SEED-IV dataset was utilized to categorize emotions as happy, sad, fear, and neutral. Sampling techniques and classification techniques were used to raise the level of performance of the algorithm. One of the most important advantages of using EEG signals is that is accurately prepared by computer systems and portrayed the real experience. The performance of the emotion recognition systems by brain signals depended on the efficiency of the algorithms used. The instability of the brain's impulses is one of the reasons why this process is seen as challenging. So, pre-process was done to remove noisy data. The proposed method focuses on implementing an algorithm that accurately classified emotions into the said four categories.**

Keywords— **EEG, Seed Dataset, Stratified sampling, Emotion, Affinity Propagation, Classification**

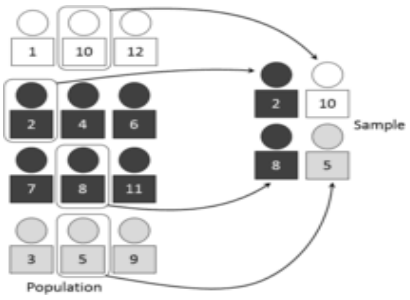
I. INTRODUCTION

Emotions are physical states that are connected to all of the nerve structures and are influenced by neurophysiological changes that are somehow connected to ideas, emotions, behavioral reactions, and some degree of joy or disappointment. However, systems haven't developed to the point where they can recognize emotions. The last couple of decades have seen a remarkable amount of research on programmed emotion recognition using BCI frameworks.

There are three kinds of emotions namely neutral, positive, and negative. Consistency, progress, and development are needed for fundamentally good emotions of gladness and pleasure. The most common unpleasant emotions, such as sympathy, irritation, nausea, and terror, often pass quickly. The field has been researching brain-computer interfaces over the last several years. BCI frameworks look at electroencephalogram (EEG) data coming from the brain. Over a decade ago, the idea behind BCI was to assist individuals with their physical and physiological problems.

An electrical capture of the brain's activity called an electroencephalogram (EEG) was made from the scalp. Waveforms that were captured showed the electrical brain action. Microvolts, the unit of measurement for EEG, indicated extremely little activity (mV). A signal's frequency Humans' EEG waves mostly have the following frequencies: Delta: The most amplitude and slowest motion are typically seen in it. It was considered common for it to be predominant in newborns up to one and through stages 3 and 4 of sleep. Theta: A "slow" activity, theta has a frequency range of 3.5 to 7.5 Hz. Adults being awake was weird when children under the age of 13 are sleeping, which was entirely natural. Fig-1: Sampling

Alpha: 7.5 to 13 Hz. The backs of the heads on both sides



had felt alpha waves more frequently, with the dominant side having a bigger amplitude. It emerged when one closes their

eyes and relaxes; when awakened by any eventually, it goes away. It was the primary rhythm seen in generally satisfied. The bulk of one's life is spent with it, especially after the age of thirteen. Beta: The term "beta" denotes "rapid" action. It oscillates at or above 14 Hz. It was frequently distributed symmetrically on both sides and it was the most obvious from the front. It may be absent or reduced in areas where the cortical layer has been damaged. It was frequently considered to have a characteristic rhythm.

Sampling, A technique that involved picking a portion of a bigger group to study to estimate the characteristics of the whole group. Getting information from a huge data collection took some time; getting information from sample data can be faster and it provided findings that are comparable. A whole population was drawn, to create the individuals that make up a sample. The process of selecting a group for your study was known as sampling. The terms "observations," "sampling units," and "sample points" were used to describe these elements. A successful research strategy was used to produce a sample. The researcher can thus use the knowledge obtained from examining the sample to comprehend the entire population. A probability sampling approach known as stratified sampling reflected a miniature replica of the population. Here, the researcher splits the total population into several subgroups or strata before utilizing straightforward probability sampling to randomly choose the final participants proportionately from the various strata. Affinity Propagation, an algorithm that located examples among data points and groups similar data points together. Until a useful collection of exemplars and clusters develops, it functions by concurrently examining each data point as a possible exemplar and exchanging signals across data points.

II. LITERATURE SURVEY

This paper [1] explored a variety of sample tactics and methodologies. The methods to carry out the sampling were described in this paper. The sampling process began with a precise definition of the target demographic. Next, a sample frame was picked. Before examining the various sampling methods, it was important to define sampling and take into account the justifications for sample selection in research. When the intended audience, sampling strategy, response rate, and sampling procedure had been decided, the next process would be data collecting. In some cases, a big sample size, the right sampling technique, and a precisely defined sample might help to reduce the likelihood of sample bias.

In addition to probability sampling approaches, this paper [2] provided knowledge on how non-probability sampling techniques and deliberate sampling are employed in research. One of the most crucial elements affecting accuracy was sampling. Statistics offered a wide range of sampling methods, so one would have got the relevant data from the population.

In this paper [3], The first provided a proper definition of sampling. Additionally, numerous sample techniques and procedures were also discussed. The benefits and drawbacks of various approaches were also discussed. Before choosing a sample strategy, it was critical to comprehend the "Pros" and "Cons." Knowing the "Pros" and "Cons" would help the researcher decide which sampling technique would be ideal for the study.

In this paper [4], a look at the theoretical underpinnings of stratified was discussed. The components of the population were sampled. The bulk ideas covered in this article might be applied to these more complex designs, even though stratified multistage cluster sampling was frequently utilized in "real-life" surveys.

This paper [5] presented a method for grouping massive amounts of data based on stratified sampling. The basic stages were outlined. Initially, a number of representative samples from different strata were obtained using a stratified sampling approach and the location-sensitive hashing technique. The collected samples are then divided into different groups using the fuzzy c-means clustering technique. to put the out-of-sample items in the clusters that were closest, using the data labeling approach. Stratified Sampling Extension FCM was used to boost computing efficiency. The C-means clustering method, which makes use of some hashing techniques, was used because it performs better in terms of efficacy and efficiency. Furthermore, this method does not lower the clustering quality.

This paper's [6] major focus was on the challenges of categorizing material from a deep web or hidden source. Accurate categorization of the input properties from a deep web data source revealed the relationships between the input and output features. The output attribute space of a deep web data source was subdivided into smaller regions. The research has offered a good estimate of the statistics, including proportions and centers, within the sub-spaces of the output characteristics.

In this publication,[7] K-means clustering was carried out using cluster exemplars made using AP. By lowering the total squared error, this clustering technique has the potential to outperform AP and K-means clustering. In addition, K-means with different initializations cannot yield reliable clustering results, unlike CAK-means. The data similarity matrix has been generated if it hasn't already been known as a prior in order to achieve high results using CAK-means. Therefore, CAK-means effectively addressed the clustering issue, when the data size is not too big.

The combination of selective clustering and fusion might improve the accuracy of clustering analysis, as this paper [8] illustrates. Selective polymerization was the aim of that investigation. A look at fusion-related techniques such as selection criteria, fusion function design, and data dimensionality reduction was there. By making use of the selective clustering fusion approach, several clustering problems were examined. The process of fusing components of a quantitative clustering algorithm involved applying a large method, the same method with methods to obtain the final clustering results. Clustering fusion had started to draw more attention in recent years.

In this paper [9], a unique strategy was suggested in order to reduce the runtime by examining fewer data points. We ran simulation research to compare our sampling-based means algorithm's speed and accuracy to that of the traditional method. According to the results, the resulted method was more efficient than the present algorithm while it retained a comparable degree of accuracy. Both the k-means clustering and the k-means sampler provided extremely accurate results, even though when the k-means sampler converged faster. The Accuracy for 2-D was 98.76.

This paper [10] introduced an idea on novel AP-based strategy for clustering mixed numerical and categorical data. The experimental findings in this research demonstrated the effectiveness of the suggested approach on several real-world mixed-type datasets. A number of user-defined parameters were provided however, it was not always apparent to check whether the value was optimum for these parameters, which looked similar to many other algorithms with parameter tuning problems. Future work concentrated on the enhancement of the AP algorithm and expanding its applicability across many areas. This paper [11] stated that after the usage of adaptive preference scanning to explore the space of the number of clusters, adaptive AP used the cluster validation approach to discover the best clustering solution suited for a data set. The adaptive escape in adaptive AP was developed to eliminate oscillations. when the adaptive damping approach fails and it was meant to do it automatically rather than manually. With these adaptive methods, adaptive AP might perform better than or on par with AP algorithm in terms of oscillation removal and clustering quality.

Table-2: Review Of Previous works

DATASET	ALGORITHM	ACCURACY
SEED and SEED-IV	Regularized Graph Neural Networks	85.30%
SEED	Knn Classifier	94.06%
SEED	SVM and LibSVM	79.3852%
SEED	Kmeans clustering	92%

In order to choose a limited but useful collection of genes, this paper [12] suggested a hybrid FS approach (mAP-KL), which combined affinity propagation (AP)-clustering algorithm, multiple hypothesis testing, and the Krzanowski & Lai cluster quality score. The mAP-effectiveness KL's against 13 other feature selection methods were evaluated, using both actual and simulated microarray data. mAP-KL displayed competitive classification results across a range of disorders and samples, It's total AUC score was 0.91, particularly in neuromuscular disorders.

This paper [13] highlighted the key underlying data patterns that were crucial to remote sensing tasks, such as picture composition and spatial arrangements. To do this, it established a successful information extraction strategy. The affinity algorithm was a brand-new, very effective method that handled odd data which included both category and numerical qualities. The choice of the starting preference value, the occurrence of oscillations, and the processing of enormous data sets were considered as some of the constraints of AP. In this study, the AP algorithm's clustering performance was assessed while taking the preferences parameter and damping factor into consideration.

This paper [14] suggested adaptive AP, which made use of adaptive priority checking to explore the space of the clusters in order to determine the ideal grouping option for a specific data collection. Furthermore, the adaptable escape feature in adaptive AP was designed to automatically eliminate oscillations rather than manually doing so when the adaptive damping strategy failed. Using these adaptive techniques, adaptive AP might be on par with or superior to AP algorithm in terms of oscillation reduction and clustering quality.

In order to cluster random two-dimensional data points quickly, this paper [15] examined four affinity propagation expansion strategies. The outcome of the theoretical analysis was consistent with the outcome of the running test, from a theoretical standpoint. Based on the theoretical investigation, it was discovered that LAP (Landmark Affinity Propagation), when compared to other Affinity Propagation expansion techniques, had the lowest computing cost.

This paper [16], makes the suggestion of estimating emotion using EEG. A hierarchical RNN and CNN were used in the suggested model's dual approach to take into account the spatial link between EEG channels. On three datasets, it displays encouraging results. It could be interesting to take into account new ML models with various representations and easier-to-understand methodologies with other feature extraction techniques. For three of the suggested datasets, the acquired results beat findings from other artistic techniques with a smaller standard deviation that suggests improved stability.

This paper [17], attempts to address the EEG sentiment classification problem in this study were prompted by the discovery that not all training samples contribute equally to emotion classification, which also holds true for the relative value of various brain areas in this sample. TANN's capacity to learn both good and bad information at the sample-level and brain-regional levels can enhance the ability to spot emotions in EEG signals. The suggested framework was simple to use, and comprehensive testing on three open EEG emotion datasets showed that the suggested TANN technique achieves the various art performance measures. The lobe and visual cortex are shown to contribute more to emotion expression when

TANN was used to explore the transferability of various brain areas in EEG emotion identification.

In this paper [18], the 5 sub bands are obtained when the DEAP dataset is decomposed using DWT. In order to convert the decomposed signals into high dimensions. ToC is used as it preserves both the original signal information and the harmonic information. For the signal decomposition Cascaded technique is used. Five layered model is proposed using deep learning. It contains a SoftMax layer, classification layer, layers of LSTM etc., Then it is divided into mini batches and the learning rate is set to speed up the process. 90% accuracy is provided by this model.

In order to effectively classify the emotions, This paper [19] make use of SVM classification algorithm for both DEAP and SEED dataset. For DEAP dataset events are separated before the data is pre-processed. Then statistical, frequency and other features are extracted. Then the obtained data is given to the classifier. Both the datasets provided best results.

In this paper [20], subjects are allowed to watch music videos(emotional). CNN is used to classify emotions. Both publicly available EEG datasets SEED and DEAP is used. Instead of providing EEG signal(raw) directly to the network, it is provided to localisation algorithm as the data should contain both spatial and temporal information. A new method is introduced which uses a toolbox from MATLAB to calculate active brain regions. The results are evaluated for cross subject scenarios. The performance of the algorithm is high when localisation of the source is combined with the proposed model.

III. PROPOSED SOLUTION

An electrical capture of the brain's activity called an electroencephalogram (EEG) was made from the scalp. The SEED-IV dataset was used in this case to obtain the EEG dataset. There were four types of emotions: joyous, depressing, frightened, and neutral. The drama recordings were specifically selected to elicit diverse emotional states. Preprocessing, the procedure for converting unprocessed data into a form that was more suitable for further analysis and understandable by the user. Pre- processing wants to be able to tell the difference between the significant neural activity at random and brain signals during EEG recording. In order to do this, a low pass filter and an artefact removal system employing signal processing techniques were created. Although blinking and eye instant were often thought of as sources of noise, it can also show significant patterns. In preprocessing, the bad channel was simply deleted. Continue with feature extraction techniques and look for the most important data in any signal that is being analyzed. Both statistical and nonstatistical data might be included. A number of feature extraction methods have been developed and reported the literature. So, finalized a DWT.

A discrete wavelet transform separates an input signal into many sets, each set comprising a time series of coefficients that illustrate the temporal evolution of the signal in the corresponding frequency range. It is possible to condense an algorithm to a more manageable set of characteristics when its input data was too large to analyse and was deemed redundant. It had another named called features vector. This method was referred to as feature extraction. Wavelet transform was used for EEG feature extraction.

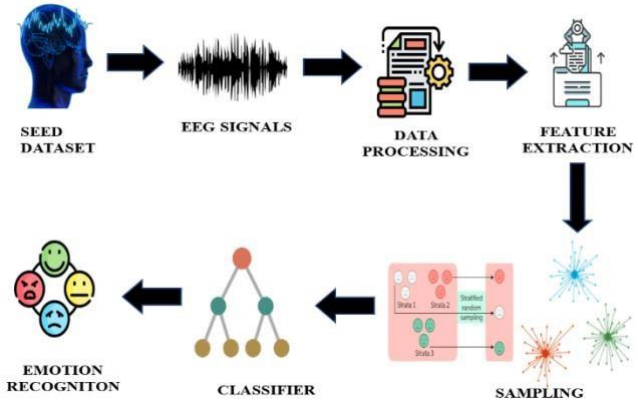


Fig-2:Workflow

The input signal is treated as a wavelet by the DWT method since it spans the frequency and time domains. These domains were used to extract 10 characteristics, including alpha, beta, gamma, theta, and delta. Sampling, a technique that involved picking a portion of a bigger group to study in order to estimate the characteristics of the whole group. Getting information from a huge data collection took some time; getting information from sample data can be faster and it provided findings that are comparable. In many types of sampling analyzing in literature survey, chosen as stratified sampling. Stratified Sampling achieved by affinity propagation algorithm. Stratified Sampling splits the total population into several subgroups or strata before utilizing straightforward probability sampling to randomly choose the final participants proportionately from the various strata. Stratified sampling offered a more accurate representation of the broader population.

Affinity propagation concurrently considers all data points as prospective exemplars and utilises similarity scores between pairs of data points as input. Real values are used to communicate amongst data points until a high-quality set of exemplars and linked clusters finally emerge. nor its cluster size nor the number of clusters must be provided as input. It may be used each time there was a technique to calculate or quantify a number that represents how similar each pair of data points. Low numbers signify little similarity, whereas high values signify great similarity.

Next, apply classification algorithms to the output of sampling algorithm, it acts as input to the classifier algorithm. Analyzing literature, SVM, KNN, DNN picked. SVM, By identifying the best binary classifier or line that can split the n-dimensional space into classes, it may be possible to categorise new data points in the future.KNN, it was frequently used as a classification approach since it was predicated on the notion

that similar points could be found nearby. DNN, the Keras deep learning tool enabled the speedy construction and evaluation of models of neural networks for multi-label classification problems. Emotions were classified and analyzed as performance metrics. By averaging the accuracy across all individuals throughout two data sessions, we assess the model's performance. . Recall was related to that of True Positive rate and Values, where it might be characterized as the degree of success in compartmentalizing the sensations. Precision was the accuracy in the center of facts, figures of specific data, and anticipating the specifics.

IV. CONCLUSION

This paper gives an overview of the research that has already been done on EEG and recognising emotions. Initially, an in-depth explanation of the EEG's mechanism, as well as its emotion trigger mode and categorization model, is provided. Then, we expanded the previously developed EEG emotion identification algorithms by looking at them from three different perspectives: the extraction of features, the selection of features, and the classifier. Lastly, a study of the literature is done in order to analyse and compare the outcomes of various emotion classification systems. There are still numerous issues with EEG emotion recognition that need to be resolved before it can be used in practise, such as the dearth of adequate emotion datasets and classifications, which will also be the focus of further research in this field.

V. FUTURE WORK

In future, research on the following would be done to assist additional domain-invariant EEG representations learned by our model:

1. Using more complicated classifiers or more sophisticated methods to manage unbalanced data between training and test sets, training of a domain classifier could be done which was considered to be more discriminative.
2. To prevent over-smoothing on these tiny graphs, a more simplified version of our model and more sophisticated regularizations could be required. Additionally, it could be worthwhile to investigate data processing methods like spatial filtering that might enhance the spatial resolution of EEG signals.

REFERENCES

[1] Taherdoost, H. (2016). Sampling methods in research methodology; how to choose a sampling technique for research. How to choose a sampling technique for research (April 10, 2016).

[2] Bhardwaj, P. (2019). Types of sampling in research. Journal of the Practice of Cardiovascular Sciences, 5(3), 157.

[3] Sharma, G. (2017). Pros and cons of different sampling techniques. International journal of applied research, 3(7), 749-752.

[4] Parsons, V. L. (2014). Stratified sampling. Wiley StatsRef: Statistics Reference Online, 1-11.

[5] Zhao, X., Liang, J., & Dang, C. (2019). A stratified sampling-based clustering algorithm for large-scale data. Knowledge-Based Systems, 163, 416-428.

[6] Liu, T., & Agrawal, G. (2012, August). Stratified k-means clustering over a deep web data source. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1113-1121).

[7] Zhu, Y., Yu, J., & Jia, C. (2009, August). Initializing k-means clustering using affinity propagation. In 2009 Ninth International Conference on Hybrid Intelligent Systems (Vol. 1, pp. 338-343). IEEE.

[8] Xiao, Y., & Wang, P. (2018, September). Application of Clustering Algorithm in Audit Stratified Sampling. In 4th Workshop on Advanced Research and Technology in Industry (WARTIA 2018) (pp. 401-406). Atlantis Press.

[9] Bejarano, J., Bose, K., Brannan, T., Thomas, A., Adragni, K., Neerchal, N. K., & Ostrouchov, G. (2011). Sampling within k-means algorithm to cluster large datasets. UMBC Student Collection.

[10] Zhang, K., & Gu, X. (2014). An affinity propagation clustering algorithm for mixed numeric and categorical datasets. Mathematical Problems in Engineering, 2014.

[11] Wang, K., Zhang, J., Li, D., Zhang, X., & Guo, T. (2008). Adaptive affinity propagation clustering. arXiv preprint arXiv:0805.1096.

[12] Sakellariou, A., Sanoudou, D., & Spyrou, G. (2012). Combining multiple hypothesis testing and affinity propagation clustering leads to accurate, robust and sample size independent classification on gene expression data. BMC bioinformatics, 13, 1-19.

[13] Moiane, A. F., & Machado, Á. M. L. (2018). Evaluation of the clustering performance of affinity propagation algorithm considering the influence of preference parameter and damping factor. Boletim de Ciências Geodésicas, 24, 426-441.

[14]Xia, D. Y., Wu, F., Zhang, X. Q., & Zhuang, Y. T. (2008). Local and global approaches of affinity propagation clustering for large scale data. Journal of Zhejiang University-Science A, 9(10), 1373-1381.

[15] Refianti, R., Mutiara, A. B., & Gunawan, S. (2017). Time complexity comparison between affinity propagation algorithms. Journal of Theoretical & Applied Information Technology, 95(7).

[16] Delvigne, V., Facchini, A., Wannous, H., Dutoit, T., Ris, L., & Vandeborre, J. P. (2022, July). A Saliency based Feature Fusion Model for EEG Emotion Estimation. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 3170-3174). IEEE.

[17] Li, Y., Fu, B., Li, F., Shi, G., & Zheng, W. (2021). A novel transferability attention neural network model for EEG emotion recognition. Neurocomputing, 447, 92-101.

[18] Sharma, R., Pachori, R. B., & Sircar, P. (2020). Automated emotion recognition based on higher order statistics

and deep learning algorithm. Biomedical Signal Processing and Control, 58, 101867.

[19] Kumar, D. K., & Nataraj, J. L. (2019). Analysis of EEG based emotion detection of DEAP and SEED-IV databases using SVM.

[20]Asadzadeh, S., Yousefi Rezaii, T., Beheshti, S., & Meshgini, S. (2022). Accurate emotion recognition using Bayesian model based EEG sources as dynamic graph convolutional neural network nodes. Scientific Reports, 12(1), 10282.