Personalizing emotion recognition using incremental random forests

Jordan Gonzalez
Learning, Data and Robotics Lab
ESIEA
Paris, France
jordan.gonzalez@esiea.fr

Lionel Prevost
Learning, Data and Robotics Lab
ESIEA
Paris, France
lionel.prevost@esiea.fr

***Abstract*—Affective learning analytics is based on video exchanges within students and between students and teachers. Their aim is to manage both cognitive and affective loads. Affective state detection from videos is subject to identity bias, particularly morphological and behavioral. Although machine learning can fairly detect basic expressions (as defined by Paul Ekman), it hardly deals with these biases and perform better after personalization to a given user (or set of users). But this adaptation needs to train incrementally algorithms with new data. We propose here to use incremental random forests that show abilities to deal with covariate shift in data. We initially train our model on CFEE dataset. Then we train incrementally and evaluate on benchmark CK dataset. Experimental results show that, after progressive training, precision increases. These results suggest that the use of incremental training can improve the accuracy of emotion recognition by specializing the model on a given subject.**

***Index Terms*—Incremental Learning, Semi-supervised Learning, Emotion Recognition, Random Forest**

I. Introduction

Emotions are the heart of the didactic triangle linking teachers, students and knowledge. Emotional stimuli can cause either disturbance or motivation and affect the learning experience. It is difficult for a teacher to focus on each student and adapt his speech in order to maintain the student's concentration to avoid student dropouts. It is all the more difficult given the current context of the pandemic where teaching is mostly done at a distance and is similar to Massive Open Online Courses (MOOCs). Automatic recognition of emotions could help in this respect. Research of facial emotion has gained a lot of attention over the past decades within the so-called affective computing domain. The main reason is that facial expressions are one of the most informative channels in interpersonal communication. However, although major progress has been made in recent years, this field is still challenging [7]. Identity bias is one challenge and interests us in this context, as all subjects express emotions in their own way. Behavioral and morphological variability imply that some systems will not be able to recognize the emotions of several subjects. However, few studies focus specifically on identity bias. This is why we are interested in so-called incremental learning systems that could help to deal with this constraint, in particular, to customize a generic model on a specific subject (or set of subjects, a group of students, for example).

Incremental Learning is a field of study in machine learning whose aim is to copy the human ability to learn new tasks throughout their lives, while not forgetting how to perform old ones. One of the main characteristics of incremental techniques is the ability to update models using only recent data (i.e. without accessing old data). This is often the only practical solution when it comes to learning data "on the fly" as it would be impossible to keep in memory and relearn from scratch every time new information becomes available. Unfortunately, when neural networks are uniquely formed on new data, they are rapidly overloaded with a phenomenon known as catastrophic forgetting [10].

The (deep) convolutional neural network (CNN) is one of the most popular network model and has achieved state-of-the-art results in various fields, including facial expressions recognition [7] [8] [9]. However, those systems suffer from catastrophic forgetting when learning incrementally new data, we decided to focus on another family of algorithms. Random Forests (RF) [2] look interesting because of their multi-class nature and their generalization ability. They have other advantages: very quick learning, possible incrementation in either data and classes. Moreover, thay may exhibit better interpretability. Nearest class mean forests have demonstrated that they can outperform RF and allow an easy way to perform incrementation [1]. That is why they will be used in this study.

In this paper, we propose an exploration and evaluation of strategies for personalizing emotion recognition. Our two main contributions are explained below. Firstly, we propose an improvement of the Nearest Class Mean Forest to deal with highly variable data with complex distribution. For such data, the class-conditional distribution is no more unimodal and considering multimodal distribution leads to better accuracy. Secondly, we propose a semi-supervised extension of the former, fully supervised, algorithm. We show that learning incrementally data streams including unlabeled data give promising results.

In section II, we present the original model. In section III, we detail the incremental learning strategies. Section IV is devoted to experiments and thorough analysis of results. Finally, section V gives some conclusions and perspectives.

II. Nearest Class Mean Forest

Before presenting the Nearest Class Mean Forest (NCMF), we present first, the classical random forest and then, the nearest class mean classifier, both being parts of the NCMF.

A. Random Forest

Random Forest (RF) is an ensemble learning method proposed by [2]. It is a popular supervised machine-learning algorithm that has been applied to Facial Expression Recognition (FER) tasks, among others [11] [12]. It is able to minimize both bias and variance. RF combine multiple decision trees trained on a random subset of the training data (bootstrap). Each tree is formed by selecting randomly, in each node, a small group of features to split on (feature bagging). During classification, the observation spreads from the root to one leaf. At the tree level, the decision is the class with the highest class-conditional posterior probability. At the forest level, the decision uses the majority vote rule.

B. Nearest Class Mean Classifier

A nearest class mean (NCM) classifier [3] is a classification model that only stores the centroid of each class of training samples. To classify unseen observations, it uses the nearest class mean rule.

C. NCMF

It is a combination of the previous concepts of RF and NCM. But some differences are to be noticed. A class bagging occurs during training: only a random subset of available classes is considered in each node. The splitting decision function is also modified (cf. Fig. 1). The optimal pair of centroids is chosen in each node according to the following criterion:

$$c_p^* \leftarrow \underset{c_p \in Pairs}{argmax} I(D_n, Ks_n, c_p) \tag{1}$$

where D_n is the data available in the current node n, Ks_n a random sampled set of classes, $c_p = \{c_i, c_j\}$ is a pair of centroids from $Pairs$ that contains all the possible pairs of classes. The pair of centroids c_p^* that maximizes the Information Gain I [15] is selected. Fig. 1 shows the splitting process. During training (a), sample is sent to the left or right child given its nearest centroid (c_i or c_j). The same process happens for testing and incremental learning (b).

III. Incremental learning strategies in NCMF

A. Previous strategies and limits

Update Leaf Statistics (*ULS*) and Incremental Growing Tree (*IGT*) have been introduced in [1]. They were proposed in a context of big data, where updating the forest rather than relearning it from scratch is justified in terms of computational load. Each incremental sample is propagated within the trees until it reaches a leaf. The ULS strategy only updates the distributions in the tree leaves. Thus, when incremental data appears, the distributions evolve and therefore the predictions are likely to change. The IGT strategy replaces the leaf by a new splitting node. This strategy is needed when class-conditional probabilities are getting closer.

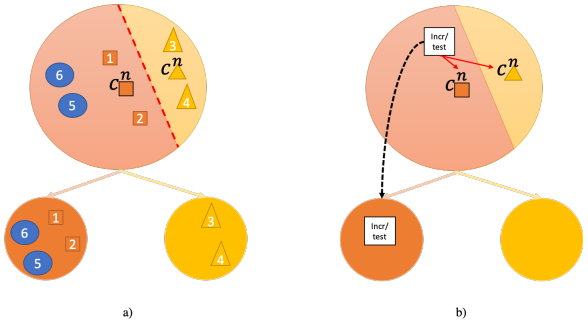

Fig. 1. Splitting decision in an NCMF node.

The main issue with this incremental procedure is the variability of the data between the initial learning and the incremental learning. In-depth analysis of splitting process at the node level shows that sometimes, incremental data is wrongly oriented due to its proximity to the centroid of another class, even if a centroid of its class is present in the node. This phenomenon is classical when data are highly variable: a single centroid cannot represent all the data and another centroid needs to be created. In statistical words, that means that the data distribution comes to be multimodal. That is why we propose to improve the IGT strategy by checking cluster quality.

B. Supervised Incremental Growing Tree Correction based on Calinski-Harabasz criterion (IGTC)

We can observe when updating a node (NCM classifier with two centroids $\{c_i, c_j\}$ that a sample x of class i can be misdirected (i.e goes towards to the "wrong" child node, corresponding to centroid c_j). To deal with, we propose the Incremental Growing Tree Correction (IGTC) function that selects one strategy among two: Update Centroid (UC) and Add new Centroid (AC). We use the Calinsky-Harabasz index to decide which is the accurate one. This index (2) is the weighted ratio of inter-group variance I_B to intra-group variance I_W. We use it as a measure of the clustering quality: the higher the value of the index, the better the clusters are defined.

$$s_{ch} = \frac{I_B}{I_W} \frac{N-k}{k-1} \tag{2}$$

In order to take into account the potential multimodality in incremental data, we make two hypotheses, leading to two strategies.

- UC strategy: the class-conditional density remains unimodal after incrementation. For a new sample x of class i, we update the centroid c_i using (3).

$$c_i' = c_i + w_{ij} x \tag{3}$$

The weight w_{ij} is the prior confusion probability of the baseline classifier for the corresponding classes i and j. It

is given by the confusion matrix (see Fig. 3). We assume here that the more confusion the initial model make, the more it needs to be modified. Then, we compute a first index s_{ch1}.

- AC strategy: The class-conditional distribution becomes multimodal. Here, the new sample x becomes a new centroid of class i (in this node) and we can compute another index s_{ch2}.

Depending on both indexes, we decide to apply the UC strategy if s_{ch1} is higher than s_{ch2}. Otherwise, we apply the AC strategy.

C. Semi-supervised IGTC

In the educational context in which we are working, the dataset at our disposal is mostly composed of children or teenagers faces. Obviously, it is fully unlabeled. Training models on adult faces is quite easy thanks to benchmarks emotional datasets (see section IV-A). However those models do not get acceptable results when tested on children faces due to high morphological and behavioral changes throughout life. That is why we are working on lifelong learning that could be (1) initially trained on adult faces and (2) able to personalize itself to a group of individuals with different ages. Adapting our algorithm to unsupervised data is necessary. That is why we focus now on the situation when the received data contains a given ratio of unlabeled data. To this end, we propose three approaches using our incremental models:

- **One Pass Incrementation (OPI)**: We use the baseline model to give a pseudo-label to all unlabeled data. Then, we use all that data to increment the model in a supervised way using strategy IGTC.
- **Two Pass Incrementation (TPI)**: We first increment using the labeled data. Then, we use the incremented model to give a pseudo label to unlabeled data. Finally, we increment one last time the model with the pseudo labeled samples.
- **Continuous Incrementation (CI)**: We act in the same way as we receive the data. As long as observations are labeled, we increment the model. Then, we use it to predict pseudo-labels for unlabeled instances when met. Model increments finally on them and so on (see Fig.2). Here, data is streaming and model continuously increments itself.

IV. Experiments

A. Data sets

1) Compound Facial Expressions of Emotion (CFEE): contains 230 subjects with one image for each of the 22 categories present in the dataset: 6 basic emotions, 15 compound emotions (i.e. a combination of two basic emotions), and neutral expression [4]. For each subject, we selected 7 images with the six basic emotions and the neutral face. Thus, 1285 images are retained to train the NCMF baseline classifier and 322 for evaluation.

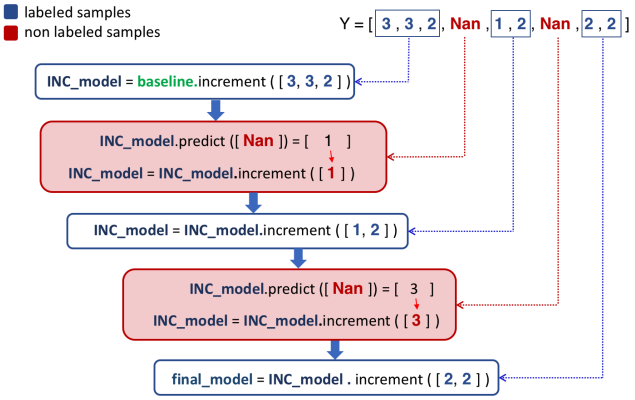

Fig. 2. Continuous Incrementation

2) Extended Cohn-Kanade CK+: is the most popular database in the field of emotion recognition. It contains 327 image sequences of deliberate and spontaneous facial expressions of 123 subjects [5]. A sequence lasts about 20 images and always starts with a neutral expression, then progresses to a specific expression up to a peak of intensity (apex) which is labeled using the Facial Action Coding System (FACS) [6]. So, 902 images (only basic emotions) are retained to perform incremental learning and same for evaluation.

B. Feature extraction

Most systems usually rely on FACS [6]. FACS encodes the movements of specific facial muscles called action units (AUs). The production of an AU has a temporal evolution, typically modeled with four temporal segments [7]. We will focus here only on neutral, and apex segments due to the databases we use in this study (e.g. CFEE database only includes static images). Neutral means no signs of muscular activity. Apex is the frame where the intensity reaches a maximum level. We use OpenFace [1], an open source software developed by [14], to detect 17 AUs and their intensities (integer values from 0 to 5).

C. Experimental protocol

To begin, a NCMF model is trained on the whole CFEE dataset. This model will be called *NCMF baseline*.

Then, we use CK+ dataset for incrementation and evaluation. In each sequence, images 0 and n (neutral and apex) are used for incrementation while images 1 and $n-1$ are used for evaluation. Thus, if the model has been personalized to an individual (using the apex image), it should perform better in labeling subtle expressions from the same individual.

To show the ability of the model to personalize on a group of subjects, the database is split into S slots (we chose 8 slots in our study). Subjects are affected to one single slot.

For each slot we use the incremental strategies (cf. section III) on a copy of the NCMF baseline. This will be called incremented NCMF i (related to the Id of the slot). Next

[1]https://github.com/TadasBaltrusaitis/OpenFace

step is the evaluation of the incremented NCMF i on its associated test slot (containing the same subjects but not the same images). So, there are S incremented models. The performance measure is the mean accuracy on all slots.

Note that in the experiments of this paper, the NCMF *class bagging* parameter has been deactivated. As the number of classes is low, we decided to consider all the available classes.

D. Experimental results

Baseline training: To build the *NCMF baseline* on CFEE training set , the out-of-bag error was used to select the optimal number of trees in the forest. The best trade-off accuracy versus computational load was a forest with 50 trees. Fig. 3 details the confusion matrix computed on the test part of CFEE.

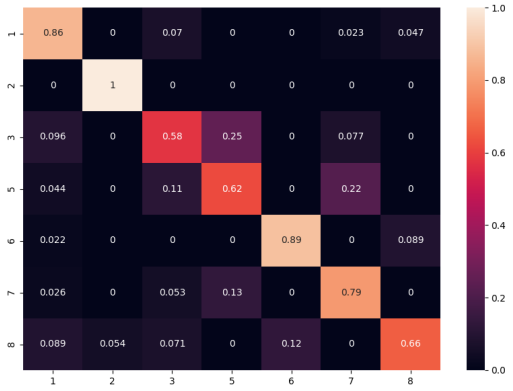

Fig. 3. Confusion matrix on CFEE test (normalized on rows). Labels are 1:Neutral, 2:Happy, 3:Sad, 5:Anger, 6:Surprise, 7:Disgust, 8:Fear

Supervised incremental learning: Fig. 4 displays the performances (mean accuracy on all slots) obtained by the baseline model and the incremented model depending on the strategy used and the number of trees in the forest. It is here just for comparison. The incrementation based on the ULS strategy improves performance but we can notice that the IGT strategy performs better. Finally, the IGTC strategy has the greatest impact, regardless of the number of trees. As for the initial training, the incremental forest with 50 trees obtains the best performances.

Fig. 5 details the average recognition rate per class obtained on the different slots of CK+ test, before and after incremental learning. In any case when they are not equal, it is IGTC that outperforms the recognition of the different labels. One can also notice the interest of using incremental methods to improve emotion recognition in situations of within class variability due to cross dataset evaluation. Focusing on classes "sad" and "angry", we can observe that incremental learning using IGTC improves class-conditional accuracy drastically. The reason being that the IGTC method uses the NCMF confusions on the training set. The most striking confusions are mainly on the classes Sad, Anger and Fear (see Fig. 3).

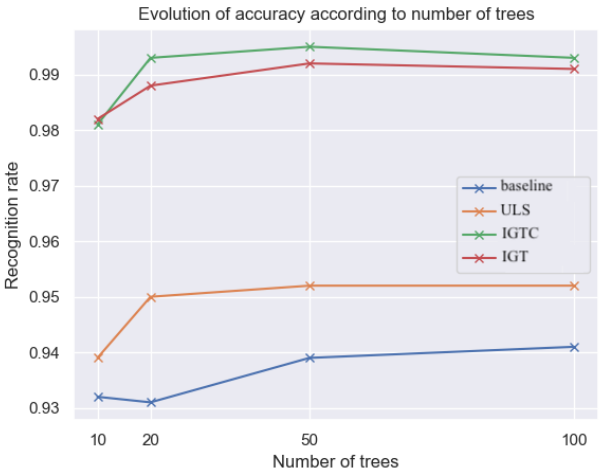

Fig. 4. Performances of incremented models with respect to the number of trees (CK+ test)

The weight being then more important for these classes (cf. section III-B), we can observe a clear improvement of the performances of the NCMF which then focused its attention on the good separation of these classes during the incremental phase.

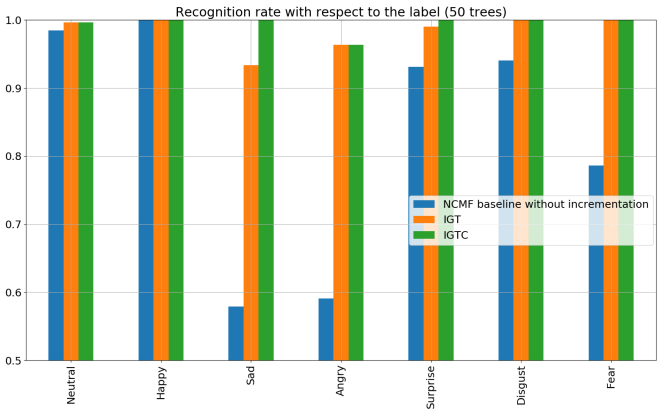

Fig. 5. Recognition rate per emotion (CK+ test)

These incremented models (IGTC on each slot) were also evaluated on the other slots (see Table I). It can be observed that the incremented models give the best performance on their respective slot. This result was expected since the protocol was carried out in order to customize each model to its subjects. Overall, we note that the incremental models improve or equal in most cases the performance of the baseline on other slots (see the underlined scores). It can also be noted that the baseline and incremented models have the same accuracy on CFEE test (0.76 ± 0.01). This suggests that there is no catastrophic forgetting.

Semi-supervised incremental learning: Results of semi-supervised experiment can be found in Table II. Due to the results previously presented in the supervised part, we

focused here on an NCMF of 50 trees incremented with IGTC method. It is interesting to see that, even when the ratio of unlabeled data is high, incrementation leads to an improved mean accuracy (NCMF baseline performance on CK+ test is 0.939 ± 0.026). However, the three strategies we proposed (OPI, TPI and CI) get quite similar results.

TABLE I
INTER-SLOTS PERFORMANCES WITH IGTC MODELS

	s_0	s_1	s_2	s_3	s_4	s_5	s_6	s_7
i_0	**0.99**	0.97	0.93	0.97	0.94	0.90	0.97	0.91
i_1	0.94	**1.00**	0.93	0.96	0.94	0.87	0.97	0.88
i_2	0.95	0.97	**1.00**	0.96	0.94	0.90	0.97	0.91
i_3	0.94	0.97	0.93	**1.00**	0.95	0.90	0.98	0.91
i_4	0.95	0.97	0.93	0.98	**1.00**	0.88	0.98	0.89
i_5	0.95	0.97	0.92	0.95	0.94	**0.98**	0.96	0.91
i_6	0.95	0.97	0.93	0.95	0.94	0.89	**1.00**	0.90
i_7	0.97	0.97	0.92	0.97	0.95	0.91	0.97	**0.99**

i refers to incremented models. s refers to slots

TABLE II
ACCURACY OF IGTC METHOD ACCORDING TO RATIO OF UNLABELED DATA

Method	OPI	TPI	CI
5%	0.991 ± 0.008	0.991 ± 0.010	0.990 ± 0.012
10%	0.990 ± 0.009	0.991 ± 0.010	0.990 ± 0.012
20%	0.984 ± 0.017	0.984 ± 0.016	0.986 ± 0.012
50%	0.971 ± 0.011	0.969 ± 0.024	0.972 ± 0.023
75%	0.957 ± 0.019	0.960 ± 0.019	0.957 ± 0.019

V. CONCLUSION

We present here an adaptation of the Nearest Class Mean Forest to emotion recognition. This algorithm is well suited for incremental learning. That is why we used it to improve a baseline model (trained on CFEE data set) on the benchmark CK+ data. The aim of the incrementation is to personalize the model on one (or a set of) individual(s), in order to reduce identity bias. We have evaluated two strategies (namely ULS and IGT). We improved the latter by using multimodal clustering in the internal nodes of tree (IGTC) and showed that accuracy improves. We also showed that it is possible to learn new data in a semi-supervised way.

It should be interesting to consider other approaches in the context of semi-supervised learning [13] with still using an incremental approach. We have already shown that internal measure of conflict between the forest trees can help to give a pseudo-label to unlabeled data.

It is also planned for future work to characterize non-prototypical expressions such as engagement, boredom, confusion and frustration that are more likely to occur in the educational context in which we are working.

ACKNOWLEDGMENT

This work has been partially supported by the French National Agency (ANR) in the frame of its FRQC program (TEEC, project number ANR-16-FRQC-0009-03).

REFERENCES

[1] Ristin, M., Guillaumin, M., Gall, J., Van Gool, L., "Incremental learning of random forests for large-scale image classification," IEEE transactions on pattern analysis and machine intelligence, 2015, vol. 38, no 3, p. 490-503.
[2] Breiman, L., "Random forests," Machine learning, 2001, vol. 45, no 1, p. 5-32.
[3] Hastie, T., Tibshirani, R., and Friedman, J., "The Elements of Statistical Learning: Data Mining, Inference, and Prediction," Springer, 2001.
[4] Tao, S.Y.; Martinez, A.M. "Compound facial expressions of emotion," Natl. Acad. Sci. 2014, 111, E1454–E1462.
[5] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I., "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," In : 2010 ieee computer society conference on computer vision and pattern recognition-workshops. IEEE, 2010. p. 94-101.
[6] Ekman, R., "What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)," Oxford University Press, USA, 1997.
[7] Sariyanidi, E., Gunes, H., Cavallaro, A., "Automatic analysis of facial affect: A survey of registration, representation, and recognition," IEEE transactions on pattern analysis and machine intelligence, 2014, vol. 37, no 6, p. 1113-1133.
[8] Ko, B. C., "A brief review of facial emotion recognition based on visual information," sensors, 2018, vol. 18, no 2, p. 401.
[9] Li, S., Deng, W., "Deep facial expression recognition: A survey," IEEE Transactions on Affective Computing, 2020.
[10] Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., Wermter, S., "Continual lifelong learning with neural networks: A review," Neural Networks, 2019, vol. 113, p. 54-71.
[11] Dapogny, A., Bailly, K., Dubuisson, S., "Pairwise conditional random forests for facial expression recognition," In : Proceedings of the IEEE international conference on computer vision. 2015. p. 3783-3791.
[12] Wei, W., Jia, Q., Chen, G., "Real-time facial expression recognition for affective computing based on Kinect," In : 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA). IEEE, 2016. p. 161-165.
[13] Zhou, Z. H., "A brief introduction to weakly supervised learning," National science review, 2018, vol. 5, no 1, p. 44-53.
[14] Baltrušaitis, T., Robinson, P., Morency, L. P., "Openface: an open source facial behavior analysis toolkit," In : 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2016. p. 1-10.
[15] Quinlan, J. Ross. "Induction of decision trees." 1986, vol. 1, no 1, p. 81-106.