A survey on modelling of facial affect using Gaussian process domain expert

Jamsheela M K
M.Tech Scholar
Department of ECE, Cochin College of Engineering and Technology, Valanchery, India

***Abstract*-Automated analysis of facial expressions has been gaining significant attention over the past years. Most of existing models for facial behavior analysis rely on generic classifiers, which fail to generalize well to previously unseen data. This is because of inherent differences in source (training) and target (test) data, mainly caused by variation in subjects' facial morphology, camera views, etc. Driven by the necessity for efficient and accurate inference mechanisms explore probabilistic framework of Gaussian processes (GPs). The main goal in our approach to automated analysis of facial expressions is to learn high dimensional mappings between the corresponding facial features and the associated output labels. The model adaptation is facilitated in a probabilistic fashion, by conditioning the target expert on the predictions from multiple source experts. Further exploit the predictive variance of each expert to define an optimal weighting during inference.**

***Index Terms*-Gaussian processes, multi-view facial expression recognition, domain adaptation, GP adaptation.**

I. INTRODUCTION

Facial expressions convey emotions, provide clues about people's personality and intentions, reveal the state of pain, weakness or hesitation, among others. The study and understanding of human facial expressions has been a long standing problem. Face recognition is a biometric method with many applications for its nature of being non-intrusive, natural, and passive. Especially, in applications such as surveillance systems, smart homes, or any application dealing with identifying people from videos, face recognition is the primary biometrics.

Due to its practical importance in medicine, marketing and entertainment, automated analysis of facial expressions has received significant attention over the last two decades [1]. In [2], each output was constrained to have the same variance in its predictions. In this article, we relax this assumption by allowing each output to have a different confidence in the output. In case of AUs, this is a more realistic scenario since the proposed classifier may be more confident in predicting some AUs than the others. The proposed model is a generalization of Gaussian processes (GPs) [3]. In [4] meta-analysis was conducted on the accuracy of predictions of various objective outcomes in the areas of social and clinical psychology from short observations of expressive behavior. Thin slices of behavior provide a great deal of information and permit accurate prediction. First challenge on Facial Expression Recognition and Analysis [5], challenge consists of a FACS Action Unit detection sub challenge and an emotion recognition sub-challenge. This work outlines the data used for the challenge as well as the challenge protocol. Head-pose invariant facial expression recognition that is based on a set of characteristic facial points. To achieve head-pose invariance, propose the Coupled Scaled Gaussian Process Regression (CSGPR) model for headpose normalization [6]. In this model, first learn independently the mappings between the facial points in each pair of (discrete) nonfrontal poses and the frontal pose, and then perform their coupling in order to capture dependences between them. During inference, the outputs of the coupled functions from different poses are combined using a gating function, devised based on the headpose estimation for the query points. The problem of transfer learning is considered in the domain of crowd counting. A solution based on Bayesian model adaptation of Gaussian processes is proposed [7]. This is shown to produce intuitive model updates, which are tractable, and lead to an adapted model (predictive distribution) that accounts for all information in both training and adaptation data. Second Facial Expression Recognition and Analysis Challenge (FERA 2015) [9] dedicated to FACS Action Units detection and intensity estimation on the highly challenging set of data. The dataset for this challenge has been composed using two facial expression databases BP4D and SEMAINE. This is the first time these datasets have been applied to FACS AU analysis except the training and development partitions of the former. The challenge addresses such significant problems of the field as expression intensity estimation as well as robust detection under non-frontal head poses, partial occlusions and environmental factors.

In this article, present an approach that can be used to adapt the context questions where (view) and who (subject), for facial expression recognition (FER) and AU detection, respectively. More specifically, explore the problem of domain adaptation, where the distribution of the features varies across domains, while the output labels remain the same. In the case of the context question 'where', this boils down to adapting the frontal classifier to a non-frontal view using only a small number of expressive images from the target view. Similarly, in the case of the subject adaptation ('who'), the model adaptation is performed by using as few annotated images of target subject as needed to gain in the prediction performance . Thus, aim is to find a data-efficient approach to adapt previously trained generic models for facial behavior analysis, and overcome the burden of computation-wise costly model relearning.

The contributions of this work can be summarized as follows:

- This is the first work in the field of facial behavior modeling that can simultaneously perform adaption to multiple outputs (i.e., AUs).
- Proposed model exploits the variance in the predicted expression in order to utilize a measure of confidence for weighting the importance of each expert.
- Data efficient since it can perform the adaptation using only a small number of target labeled data.
- Prediction mechanism based on the weighted combination of the source and target experts acts as a guard against negative transfer, allowing the model to explore the full capacity of the appropriate domain.

A. Domain Adaptation in Facial Behavior Analysis

An important issue for the facial behavior analysis, and, in particular, the analysis of AUs, remains the poor generalizability to previously unseen data / contexts. A widely used algorithm for adaptation is the kernel mean matching (KMM), which directly infers resampling weights by matching training and test distributions. The other method, employed the KMM to learn person-specific AU detectors. This is attained by modifying the SVM cost function to account for the mismatch in the distribution between source and target domain, while also adjusting the SVM's hyper-plane to the target test data. Although effective, this transductive learning approach is inefficient since for each target subject a new classifier has to be relearned during inference.

In this work, consider supervised approach that needs only a small amount of annotated data from target domain to perform the adaptation and also define both target and source experts, assuring that the performance of the resulting classifier is not constrained by the distribution of the source data, as in unsupervised adaptation approaches.

B. Domain Adaptation

Domain adaptation is a well-studied problem in machine learning. In general, the adaptation problem stems from the change in the distributions of the input features and/or output labels between the two domains. The goal of domain adaptation is to match the differing distributions in order to learn a machinery that works sufficiently well on the target data. The adaptation can be performed either in an unsupervised or a (semi-)supervised setting, based on the availability of labeled target domain data. The (semi-)supervised setting is more appropriate to target task, since the available labels can be used to enhance the classification performance.

Compared to the aforementioned work, proposed approach defines a target specific expert, which is then combined with the source domain experts. The benefit of this is that the resulting classifier is not limited by the distribution of the source data.

II. Gaussian Process Domain Experts (GPDE)

In the following, domain adaptation is introduced to the framework of GPs and also include a methodology for obtaining a universal classifier with good generalization abilities and capable of modeling domain specific attributes.

C. Gaussian process

A Gaussian process is a generalization of the multivariate Gaussian distribution to an infinite number of dimensions (random variables). A sample from a Gaussian process is a random function f that models the relationship between two variables, that is$f: X \rightarrow Y$. X and Y are usually corresponding multivariate instances $X = \{xi\}_{i=1}^{N}$ and$Y = \{yi\}_{i=1}^{N}$, with $xi \in R^{q}$and $yi \in R^{D}$.

- GPs, as a fully probabilistic framework, can naturally provide a well calibrated uncertainty in their predictions.
- Due to their non-parametric nature, GPs allow us to specify various types of covariance functions that can capture complex data structures.
- Prior knowledge can be easily introduced during the learning of latent variables using GPs.

D. Problem formulation

Consider a supervised setting for domain adaptation with access to a large collection of labeled source domain data, S, and a smaller set of labeled target domain data, T. Let X and Y be the input (features) and output (labels) spaces, respectively. $X^{(s)} = \{x_{ns}^{(s)}\}_{ns=1}^{Ns}$, $X^{(t)} = \{x_{nt}^{(t)}\}_{nt=1}^{Nt}$ and Nt ≪ Ns. In order to avoid the burden of learning approximate solutions with GP classification, formulate the predictions as a regression problem where:

$$y_{nv}{}^{(v)} = f^{(v)}\left(x_{nv}{}^{(v)}\right) + \epsilon^{v} \text{ , where } \epsilon^{v} \sim N(0, \sigma_v^2 I)$$

Gaussian processes (GPs) extend multivariate Gaussian distributions to infinite dimensionality. Formally, a Gaussian process generates data located throughout some domain such that any finite subset of the range follows a multivariate Gaussian distribution. Now, the n observations in an arbitrary data set, y = {y1,……..,yn}, always be imagined as a single point sampled from some multivariate (n-variate) Gaussian distribution, after enough thought. Hence, working backwards, this data set can be partnered with a GP. Thus GPs are as universal as they are simple. Very often, it's assumed that the mean of this partner GP is zero everywhere. What relates one observation to another in such cases is just the covariance function, k (x; x'). A popular choice is the 'squared exponential'

$$K(x, x') = \sigma_f^2 \exp(-\frac{1}{2l^2} ||x - x'||^2)$$

Where the maximum allowable covariance is defined asσ_f^2, this should be high for functions which cover a broad range on the y axis. If $x \cong x'$, then $K(x, x')$ approaches this maximum, meaning f(x) is nearly perfectly correlated with $f(x')$. This is good for function to look smooth, neighbours must be alike. Now if x is distant from x', we have instead$K(x, x') \cong 0$, Where $\{l, \sigma_f\}$ are the kernel hyper-parameters. The regression mapping can be fully defined by the set of hyper-parameters $\theta = \{l, \sigma_f, \sigma_v\}$.

E. GP Adaptation

A straightforward approach to obtain a model capable of performing inference on data from both domains is to assume the existence of a universal latent function with a single set of hyper-parametersθ.

a. Train a GP on the source data with marginal likelihood $P\left(X^{(s)}, Y^{(s)}, \theta\right)$ to learn the hyper-parameters θ.
b. Use the obtained posterior distribution of the source data, as a prior for the GP of the target data $P\left(X^{(t)}, Y^{(t)}, D^{(s)}, \theta\right)$.
c. Correct the posterior distribution to account for the target data$D^{(t)}$as well.

III. CONCLUSION

The facial expressions analysis using Gaussian process exploits successfully the nonparametric probabilistic framework of GPs to perform domain adaptation for both multi-class and multi-label classification of human facial expressions and also this approach defines a target expert to model domain-specific attributes. It also reduce the effect of negative transfer. As a purely probabilistic model, GPDE explores also the variance in the predictions. The other methods are purely on basic classifiers and also which fail to generalize well to previously unseen data.

REFERENCES

[1] M. Pantic, "Machine analysis of facial behaviour: Naturalistic and dynamic behaviour," Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 364, no. 1535, pp. 3505–3513, 2009.
[2] S. Eleftheriadis, O. Rudovic, M. P. Deisenroth, and M. Pantic, "Gaussian process domain experts for model adaptation in facial behavior analysis," IEEE Conf. on Computer Vision & Pattern Recognition, Workshops, 2016.
[3] C. Rasmussen and C. Williams, Gaussian processes for machine learning. MIT press Cambridge, MA, 2006, vol. 1.
[4] N. Ambady and R. Rosenthal, "Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis." APA Psychological Bulletin, vol. 111, no. 2, p. 256, 1992.
[5] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, "The first facial expression recognition and analysis challenge," in IEEE Int'l Conf. on Automatic Face and Gesture Recognition, 2011, pp. 921–926.
[6] O. Rudovic, M. Pantic, and I. Patras, "Coupled Gaussian processes for pose-invariant facial expression recognition," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 35, no. 6, pp. 1357–1369, 2013.
[7] B. Liu and N. Vasconcelos, "Bayesian model adaptation for crowd counts," in IEEE Int'l Conf. on Computer Vision, 2015, pp. 4175–4183.
[8] M. F. Valstar, T. Almaev, J. M. Girard, G. McKeown, M. Mehu, L. Yin, M. Pantic, and J. F. Cohn, "FERA 2015 - second facial expression recognition and analysis challenge," in IEEE Int'l Conf. on Automatic Face and Gesture Recognition, vol. 6, 2015, pp. 1–8.
[9] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 31, no. 1, pp. 39–58, 2009.
[10] Y.-L. Tian, T. Kanade, and J. F. Cohn, "Facial expression analysis," in Handbook of face recognition, 2005, pp. 247–275.
[11] M. Pantic, "Machine analysis of facial behaviour: Naturalistic and dynamic behaviour," Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 364, no. 1535, pp. 3505–3513, 2009.
[12] J. M. Girard, J. F. Cohn, L. A. Jeni, S. Lucey, and F. D. la Torre, "How much training data for facial action unit detection?" in IEEE Int'l Conf. on Automatic Face and Gesture Recognition, 2015.