April 09, 2019

Uncertainty modeling in affective computing

Asma Ghandeharioun

Brendan Jou

Brian Eoff

Florian Schroff

Uncertainty modeling in affective computing

ABSTRACT

This disclosure describes techniques that capture the uncertainty in machine-vision based affect (emotion) perception. The techniques are capable of predicting aleatoric, epistemic, and annotation uncertainty. Measures of uncertainty are important to safety-critical and subjective assessment tasks such as those found in the perception of affective expressions.

KEYWORDS

- Affective computing
- Emotion perception
- Uncertainty modeling
- Label distribution statistics
- Annotator response time
- Label distribution learning
- Aleatoric uncertainty
- Epistemic uncertainty
- Virtual assistant

BACKGROUND

Current affective computing techniques, e.g., techniques for emotion recognition, apply machine learning models that are indifferent to the stochasticity of the data they fit and are incapable of providing confidence bounds for their predictions. Humans are capable of recognizing emotions in others and also knowing when they are unsure of their recognition; machine learned models have made great strides in the former task, but not the latter. Aside from the subjectivity of emotion perception, there are conditions in which the inputs are too difficult to assess accurately, e.g., blurry images, images captured in low-light conditions, etc. A model that predicts on image inputs falling outside of its data distribution or target output tasks that are subjective in nature should communicate a high level of uncertainty in the prediction. The model must know what it knows and also know what it does not know, and communicate this information alongside its predictions.

DESCRIPTION

Per the techniques of this disclosure, an uncertainty measure is associated with the output of a machine learning based emotion recognizer. With this uncertainty measure, applications that rely on model output, e.g., a virtual assistant, can modulate responses based on the uncertainty in emotion recognition. For example, a virtual assistant can respond thus: "You seem frustrated, *but maybe I'm not reading you correctly*," or "*I'm not quite sure*, but you seem stressed today," or simply, "*I don't know enough* to help in this context." Communicating uncertainty to the user calibrates user expectations, makes virtual assistants more approachable, opens possibilities for human-machine communication narratives, and improves safety-critical subjective assessment tasks.

Various techniques are herein described to model uncertainty in emotion recognition. The techniques are not mutually exclusive, e.g., they can be used concurrently in the same model or application.

Modification of classification loss to learn distributions over logits

This approach identifies aleatoric uncertainty [1], or the variability inherent in the system under observation. Logits are modeled as normal distributions and squashed through a softmax function with Monte Carlo sampling. At the time of inference, the aleatoric uncertainty is the entropy over the class membership probability vector. Aleatoric uncertainty is also referred to as known unknowns.

<u>Parameterizing a model as a samples over learned distributions</u>

This approach identifies epistemic uncertainty [1], also known as model certainty, and quantifies the inability of a model to fully capture the process that generated the data that the model was trained on. The posterior probability being difficult to evaluate, dropout variational inference [5] is used to optimize over the parameters of a mixture of Gaussians. The uncertainty is the entropy over the class membership probability vector. Epistemic uncertainty is also referred to as unknown unknowns.

<u>Uncertainty loss weighting or multi-task learning with an uncertainty branch</u>

In this non-Bayesian approach to modeling uncertainty, an output objective such as a log-loss function with annotator agreement is directly weighted as a measure of uncertainty [2], [3], [4]. The approach also directly predicts the uncertainty as an output objective, e.g., it predicts first-order inter-annotator agreement statistics or explicit uncertainty annotations that annotators provide.

<u>Label distribution statistics (annotator agreement) as a proxy for uncertainty</u>

In this approach [6], variation (divergence) in a set of labels is taken as a notion of uncertainty. For a (classification) problem instance, variation in a set of labels can be predicted , e.g., as a consequence of classification by applying a measure of variation to an estimated distribution over labels. Additionally, variation in a set of labels can be predicted directly from the problem instance. Either way, a variation in a set of labels serves as a proxy for predicted uncertainty.

In the present problem instance, e.g., emotion classification, the number of labels assigned by annotators is used as a measure of uncertainty. For example, if a number of annotators agree that a certain image is that of a happy face, then the uncertainty is low. If annotators of an image rate it variously as displaying emotions ranging across, e.g., happiness, sadness, anger, surprise, etc., then the uncertainty is high. The techniques of this disclosure directly predict label distribution statistics, thereby predicting uncertainty in the detected emotion.

Annotator response time (RT) as a proxy for uncertainty

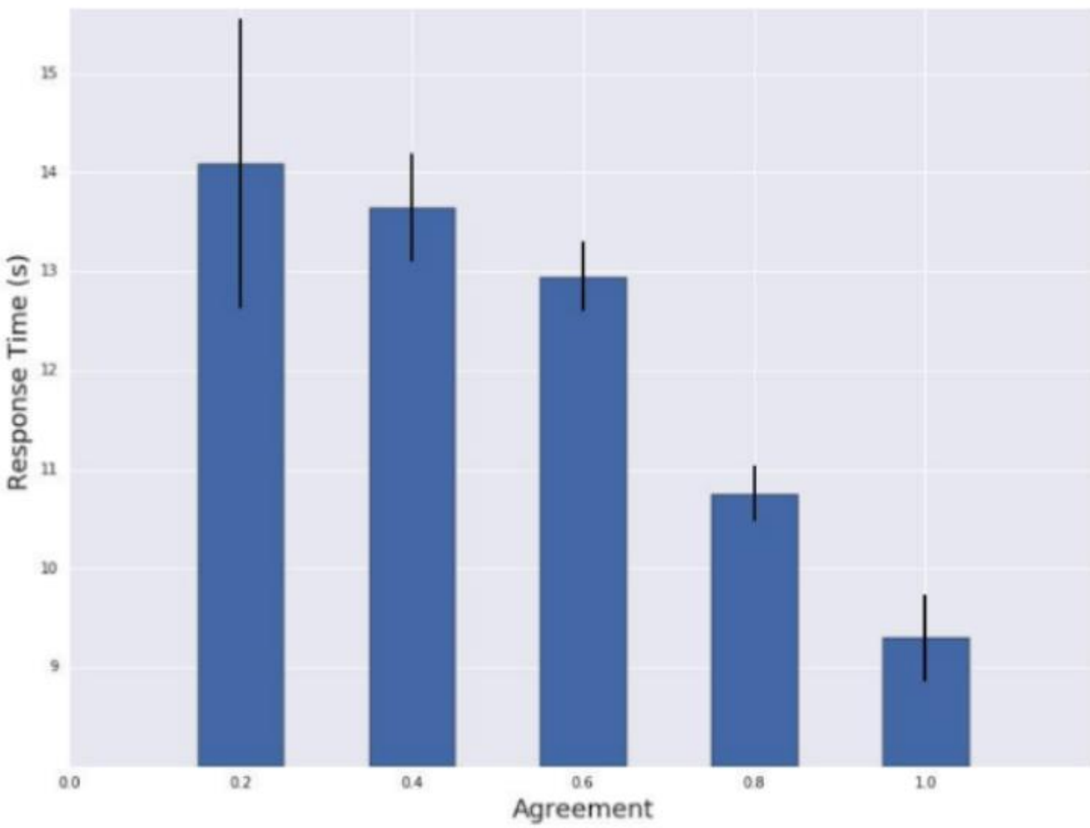

Fig 1: Histogram of annotator response time versus annotator agreement

Strong correlation is found between response time (RT) and inter-annotator agreement, as illustrated in Fig. 1. These observations indicate that the annotator response time is a suitable proxy for uncertainty in emotion recognition.

For example, if an image takes a short time, e.g., less than 1.5 seconds on average, for an annotator to classify it, then the uncertainty in classification is low. If an image takes a long time, e.g., tens to hundreds of seconds on average, for annotators to classify it, then the uncertainty in classification is high. The techniques of this disclosure predict response time statistics, e.g., mean and variance, in a multi-task setting with mean-squared loss. The techniques may return the logarithm of the variance to make the uncertainty prediction more stable.

Label distribution learning (LDL) to predict uncertainty

	emotion	score
0	Happiness	0.50
1	Sadness	0.25
2	Surprise	0.10
3	Anger	0.05
4	Fear	0.05
5	Disgust	0.05

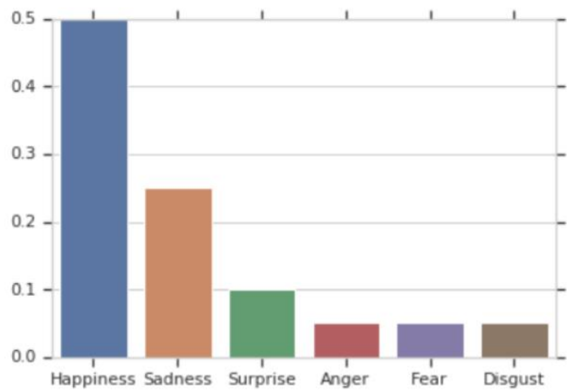

Fig. 2: An image and emotive labels associated with the image, along with their scores

Under label distribution learning (LDL), the emotion detector is trained to output the distribution of labels based on a symmetric distribution metric such as Jeffrey’s divergence. Modeling facial expressions as distributions rather than binary classifications, improves performance and efficacy.

Combining LDL and RT to predict uncertainty

Uncertainty can be predicted by combining LDL and RT as follows. Loss based on response time is attenuated to make difficult examples count more [7] and to prevent the propagation of uncertainty [1].

In this manner, inferring both class prediction scores and confidence measures, per the techniques described herein, enables the rejection of samples which are on the tails of a learned model distribution. For example, instead of requiring a model to perform well on a task for recognizing user frustration under poor lighting conditions, the confidence measure can simply indicate that the model is not certain of the output. The techniques also enable the rejection of adversarial inputs, e.g., malicious inputs that are intentionally fed to try to get the model to produce unexpected outputs. Explicitly communicating model confidence, e.g., by having the model indicate not having sufficient information or assessing that the risk of being wrong is unacceptably high can make the user experience more empathetic. Simultaneously predicting annotator uncertainty, either in the form of response time or agreement statistics, accelerates convergence and improves performance.

CONCLUSION

This disclosure describes techniques that capture the uncertainty in machine-vision based affect (emotion) perception. The techniques predict aleatoric, epistemic, and annotator uncertainty, e.g., in the form of response time or agreement statistics. Annotator uncertainty serves as a proxy for uncertainty. Measures of uncertainty are important to safety-critical and subjective assessment tasks such as those found in the perception of affective expressions.

REFERENCES

[1] Kendall, Alex, and Yarin Gal. “What uncertainties do we need in bayesian deep learning for computer vision?” In *Advances in neural information processing systems*, pp. 5574-5584. 2017.

[2] Deng, Jun, and Björn Schuller. “Confidence measures in speech emotion recognition based on semi-supervised learning.” In *Thirteenth Annual Conference of the International Speech Communication Association*. 2012.

[3] Zhang, Zixing, Jun Deng, Erik Marchi, and Björn Schuller. “Active learning by label uncertainty for acoustic emotion recognition.” In *Proceedings INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association, Lyon, France*. 2013.

[4] Eyben, Florian, Martin Wöllmer, and Björn Schuller. “A multitask approach to continuous five-dimensional affect sensing in natural speech.” *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, no. 1 (2012): 6.

[5] Gal, Yarin, and Zoubin Ghahramani. “Dropout as a bayesian approximation: Representing model uncertainty in deep learning.” In *International Conference on Machine Learning*, pp. 1050-1059. 2016.

[6] Raghu, Maithra, Katy Blumer, Rory Sayres, Ziad Obermeyer, Sendhil Mullainathan, and Jon Kleinberg. “Direct Uncertainty Prediction with Applications to Healthcare.” *arXiv preprint arXiv:1807.01771* (2018).

[7] Lin, Tsung-Yi, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. "Focal loss for dense object detection." In *Proceedings of the IEEE international conference on computer vision*, pp. 2980-2988. 2017.