Integrating model-based prediction and facial expressions in the perception of emotion

Nutchanon Yongsatianchot and Stacy Marsella

College of Computer and Information Science and Department of Psychology,
Northeastern University, Boston MA 02115, USA

Abstract. Understanding a person's mental state is a key challenge to the design of Artificial General Intelligence (AGI) that can interact with people. A range of technologies have been developed to infer a user's emotional state from facial expressions. Such bottom-up approaches confront several problems, including that there are significant individual and cultural differences in how people display emotions. More fundamentally, in many applications we may want to know other mental states such as goals and beliefs that can be critical for effective interaction with a person. Instead of bottom-up processing of facial expressions, in this work, we take a predictive, Bayesian approach. An observer agent uses mental models of an observed agent's goals to predict how the observed will react emotionally to an event. These predictions are then integrated with the observer's perceptions of the observed agent's expressions, as provided by a perceptual model of how the observed tends to display emotions. This integration provides the interpretation of the emotion displayed while also updating the observer's mental and emotional display models of the observed. Thus perception, mental model and display model are integrated into a single process. We provide a simulation study to initially test the effectiveness of the approach and discuss future work in testing the approach in interactions with people.

Keywords: Emotion perception; Bayesian Inference; Agent-Based Modelling

1 Introduction

Understanding a person's mental state is a key challenge to the design of an Artificial General Intelligence (AGI) that can interact with people. In our everyday life, interpreting and understanding what other people are feeling and thinking is an important task. When you have a conversation with your friends, you want to understand what they are thinking and feeling about a conversation. You may want to continue talking if you infer your friends enjoys it, but you may want to change the topic if you think your friends do not like it. This inference can draw on many sources of information, including the observed behavior such as facial expressions, the situation the observed person is in, and the observer's beliefs about the observed person's goals and beliefs.

One of the important questions regarding these different sources of information is how to integrate them. While you are talking, you observe that you friend frowned. Should you interpret that ambiguous frown as negative reaction to what you are saying or is it rather a sign of concentration showing interest? In addition, how should we use the new observations and inferences to help refine our beliefs about the observed agent's goals and beliefs?

Our interest is in giving a similar capacity to an artificial agent observing another human or artificial agent. This has led us to explore the questions of how predictions from observed agent's models about emotion can be integrated with the perception of facial expression, and how the observer can update the models based on the observation and inference to achieve the true model of observed agent.

A key question here is how emotions relate to expression. Ekman and Izard [4], [8] argue that some facial expressions signal specific *basic emotions*. According to this view, there is a specific way of expressing each basic emotion that is culturally universally recognizable. However, other research [6] [11] has alternatively argued that different cultures and different individuals can express the same emotion differently.

Additionally, Calvo and D'Mello [2] have pointed out the limitation of many existing affect detection systems is that they do not take the context of an emotion evoking situation into account. They have argued on the important of top-down contextually driven predictive models of affect. One type of emotion's theories that makes a prediction about emotion based on context information is appraisal theory. Appraisal theories argue that a person reacts to a situation based on how a person appraises the situation with respect to his or her goals and beliefs. [12] [10] Therefore, when predicting other person's emotion based on context, it is important to take into account the individual difference in term of goals and beliefs.

In this paper, we present an approach to infer on observed agent's emotional states by integrating both top-down predictions about emotional response given how a situation is influencing an observed agent's goal as well as bottom-up facial expression observations of the agent as it expresses that emotional response. This work extends previous work by Alfonso [1] by choosing to leverage ideas of the descriptive Bayesian approach [13] that allow us to capture the individual differences in how the observed agent emotional reacts to a situation and how the observed agent displays that emotional reaction. The descriptive Bayesian approach is an inference approach that allows multiple priors and likelihoods.

To express individual differences in how agent's emotionally reacts, we use an Appraisal Theory of emotion. To model differences in expression of emotion, we draw on the concept that people have "display rules" [11] that mediate how they express emotion. First, we argue that appraisal is operating top-down and acts as prior in Bayesian inference making probabilistic predictions about observed agent's emotion from context. Second, we can group individual difference in facial expression into the group of display rules for each emotion which allows us to infer emotion from facial expression. Finally, we also seek to model not only

inference of emotion, but also how observations and inferences could be used to update observed agent's models of an observed agent's goal and display rules.

In the rest of the paper, we first discuss the proposed method. We illustrate how the descriptive Bayesian approach captures individual difference, how appraisal theory can be used to predict emotion given situation, and how facial expressions can be grouped using display rules. Then, we describe in detail the mathematic behind our approach. After that, we explain the simulation to test the proposing method and the result of simulations. A simulation study was designed to demonstrate that our method could converge to the observed agent's true model and display rules, and could predict observed agent's emotion more accurately by using both agent's model with context and display rules. At the end, we discuss the implication of the work and future work.

2 Method

2.1 Expressing individual difference in Bayesian Inference

In a standard Bayesian model, the learner's inferences are described by Bayes' rule as following:

$$\Pr(h|x) = normalize(\Pr(x|h)\Pr(h|H))$$

where x represents the data available to the learner, h is a hypothesis that generates the data, and H is the set of all hypotheses available to the learner. In this setting, we need to know and constrain the prior and likelihood beforehand. Tauber et al. [13] proposed a descriptive Bayesian approach in which Bayes' rule could be expressed with multiple priors and likelihoods. In the descriptive approach, there could be multiple possible choices of prior and likelihoods, and learner's inferences are also conditioned on all possible prior and likelihoods. This approach argues that the learner's prior should not be perceived as fixed by some expectation about the thing to learn, and the likelihoods need not to correspond to any specific theory or model of how data are generated.

For our work, we apply the idea of multiple priors and likelihoods to capture the individual difference as the following. Given the same event or context, different person could experience different emotion based on his or her goals and beliefs used to evaluate the event. As a result, different models act as possible different priors of emotion. Similarly, there could be many different ways to express the same emotion based on a display rule, so display rules act as possible different likelihoods for a specific emotion. Therefore, the descriptive Bayesian approach allows us to capture individual differences in emotion expression in terms of different display rules as multiple likelihoods and different models as multiple priors.

2.2 Appraisal Theory and Theory of Mind

In order to predict the emotion based on context and agent's model, we use appraisal theory of emotion. Generally, appraisal theories argue that emotion

arises from a process of a subjective assessment of the relation between the event and a person's goal. [12] In this work, we use the appraisal theory proposed by Ortony, Clore, and Collins or OCC model of emotion [10], [3]. Briefly, OCC model is an appraisal theory that focuses only on the structure of situation, and does not involve any process of appraisal. This is suitable for our purpose since all we want in our simulation is a distribution of possible emotion from a given situation, and not the underlying processes. OCC model specifies the features of the prototypical situations represented by each kind of emotion, and separates emotions into three groups - emotion that focuses on event, agent or object. Note that we could replace OCC model with any other appraisal theory as long as it provides a reasonable way to obtain a distribution of emotion given context and agent's model. Further, we assume an observing agent can appraise events from the perspective of the observed. In particular, the observer has beliefs about observed agent's goals, what is sometimes referred to as a Theory of Mind [14]. (We are assuming agent architectures that can model other agents [9].)

2.3 Display rules

The display rules in this work are influenced by Safdar et al. work [11], and dialect theory [6]. In essence, display rule modifies the expression of emotion. Safdar et al. proposes seven different possible behavioral responses: amplify, deamplify, neutralize, masque by displaying another emotion, qualify by combining the actual emotion with another emotion, and express exactly without modification. In addition, Elfenbein et al [6]. have shown that different culture has a different way of displaying the same emotion similar to different dialects in language.

Combining these two ideas, three different display rules of each emotion were designed for simulation study. We define a display rule as a set of action units (AU) [5] associated with the probability that it will be expressed. It can be thought of as a function that takes in an emotion and generates facial expression. See Display rules in simulation section for more detail.

2.4 Calculation

Inferring Emotions

$$\forall e \in E : \Pr(e|X,c) = norm(\Pr(X|e,c)\Pr(e|c)) \tag{1}$$

$$= norm\Big(\prod_{x\in X} \Pr(x|e,c)\big(\sum_{m\in M} \Pr(e|m,c)\Pr(m|c)\big)\Big) \tag{2}$$

$$= norm\Big(\prod_{x\in X} \big(\sum_{d_e\in D_e} \Pr(X|e,d_e,c)\Pr(d_e|e,c)\big)\big(\sum_{m\in M} \Pr(e|m,c)\Pr(m)\big)\Big) \tag{3}$$

$$= norm\Big(\prod_{x\in X} \big(\sum_{d_e\in D_e} \Pr(X|d_e)\Pr(d_e)\big)\big(\sum_{m\in M} \Pr(e|m,c)\Pr(m)\big)\Big) \tag{4}$$

The equations above represent the way to infer the emotion of an observed agent given facial expression and context. *norm* stands for normalize. E is a

probabilistic distribution of emotion, and e is a category of emotion. An example of E is the following: $E = \{\text{angry} : 0.1, \text{happy} : 0.5, \text{sad} : 0.1, \text{no emotion} : 0.3\}$ where the number is the probability that the observer expect agent to experience that emotion. X is a set of action units represents facial expression, and x is an individual action unit. D_e is a set of display rule for emotion e and d_e is an individual display rule for emotion e. M is a probabilistic distribution of observer agent's possible models of the observed agent, and m represents each possible model. Lastly, c is context which contains the information about the situation that is eliciting the emotion. In the case of display rules of each emotion d_e, in our simulation, they are defined to be specific for a given context so they are already taken context into account. See simulation section for full description and examples of observed agent's goal, display rules, and context.

The first equation, $\Pr(e|X, c)$ is expressed using Bayes' rule. In order to calculate $\Pr(e|c)$, we express it in term of multiple possible models, m. Basically, the observer has multiple mental models of observed agent that could be used to infer observed agent's emotion from context. $\Pr(e|m, c)$ is calculated based on OCC model which takes both model and context, and returns a probabilistic distribution of emotion. For $\Pr(m|c)$, we assume that model is independent from context, which results in $\Pr(m)$. Multiple possible models represent multiple possible priors in the descriptive Bayesian approach.

We express $\Pr(X|e, c)$ in term of multiplication of $\Pr(x|e, c)$ for all $x \in X$. Here, we assume that each action unit is independent. $\Pr(X|e, c)$ can be further expressed in term of multiple display rules of a given emotion, d_e. Again, this is similar to the idea of the descriptive Bayesian approach in which we could have multiple likelihood functions. The first term $\Pr(x|d_e, e, c)$ is the likelihood that x will be generated from d_e. A display rule d_e is a subset of both context and emotion so $\Pr(x|d_e, e, c)$ can be reduced to $\Pr(x|d_e)$. Since we define d_e specific for a given context c and emotion e, $\Pr(d_e|e, c)$ can be reduced to $\Pr(d_e)$.

After the calculation, an emotion that has a highest probability or a maximum a posteriori (MAP) of $\Pr(e|X, c)$ is the prediction of emotion that the observed agent experiences.

Updating the distribution of models

$$\Pr(m|e) = norm(\Pr(e|m) \Pr(m)) \tag{5}$$

$$\Pr(m)_{new} = \sum_{e \in E} (\Pr(m|e) \Pr(e) + \Pr(m)_{old}(1 - \Pr(e))) \tag{6}$$

Equation 5 calculates posterior probability of m given e using Bayes' rule. $\Pr(e|m)$ is calculated using OCC model similar to how we calculate prior in equation 3, but only for one model. Note that we omit context in the above equations but we use it to calculate $\Pr(e|m)$.

In the inference part, we infer emotion in term of probabilistic distribution so there is uncertainty associated with our inference. For example, the observer may infer that observed agent experience happy with some probability p. We need to take uncertainty of evidence into account when we update a distribution

of model. Equation 6 represents how we use posterior probability from equation 5 to update the probability of model, m, accounting for uncertainty of evidence. There are two possible cases - either observed agent experiences emotion e with probability $\Pr(e)$ or does not experiences it with probability $1-\Pr(e)$. If observed agent experience e, we update $\Pr(m)$ based on the posterior probability as in the first part, $\Pr(m|e)\Pr(e)$, in equation 6. If observed agent does not experience e, we keep $\Pr(m)$ the same as in the second part, $\Pr(m)(1-\Pr(e)$, in equation 6. We update $\Pr(m)$ using every emotion e in E.

Updating the distribution of display rules of each emotion

$$\Pr(d_e|X) = norm(\Pr(X|d_e)\Pr(d_e)) \tag{7}$$

$$\Pr(d_e)_{new} = \Pr(d_e)_{old}(1-\Pr(e)) + \Pr(d_e|X)\Pr(e) \tag{8}$$

Equation 7 expresses the posterior probability of display rules of emotion e, d_e, given X using Bayes rule. Equation 8 is similar to equation 6 in which we takes into account the probability of emotion when we updating the probability of d_e. For display rule, unlike model that we takes into account all emotions, we only consider emotion e that corresponds to d_e.

3 Simulation

In order to demonstrate the method, a simulation study was designed. There are two things we want to test. First, by using situational context and observed agent's facial expression, our method, starting with uninform distribution of model and display rules, could converge to the true observed agent's model and display rules. Second, after converge, our method could use both model and display rule to predict observed agent's emotion correctly, better than using model alone, using display rules alone, and using neither of them.

3.1 Context and Model

The simulation is the following. At each time step, the observer observes a target agent receives a payment from the boss. The boss can either give the observed agent extra money, or deduct money from a payment. The upper bound is 6000, and the lower bound is -6000. The goal of an observed agent is to earn a baseline payment. Since the goal is just a reference point, we can set it to be 0. An observed agent can have different expectation on what the extra money should be. In this simulation, there are three different expectations - no expectation (0), low expectation (+2000), and high expectation (+4000).

OCC theory uses a threshold to determine whatever a person will experience any emotion or not. However, in this work, we want to express it in probabilistic terms. Therefore, to calculate the probability of an emotion using OCC and context, we use a logistic function with the expectation as the mid-point.

According to OCC model, one of the mechanisms that makes an agent to experience different emotions from the same event is determined by the component the agent focuses on. In this simulation, an agent can focus on event or agent causing the event. We define two different types of focus. The first type of agent is likely to focus more on an event while the second type of agent is likely to focus more on an agent causing the event. Combining three different expectations and two different foci, there are 6 possible models of observed agent in our simulation

In summary, we simulate the situation that can please or displease the observed agent according to the agent's goal, and the observed agent can focus on the event itself or another agent that causes it. As a result, according to OCC, there are four different kinds of emotion - happy, sad, grateful and angry. However, we group happy and grateful together as a positive emotion labeled happy, because gratitude does not normally show up in facial expression literatures. Therefore, we are left with happy, sad, angry, and no emotion.

To illustrate OCC model, consider the following example, an observed agent, with low expectation (2000) and likely to focus on event (0.8), receive 2000 extra money. The probability that agent will be happy is $0.6 \times 0.8 = 0.48$, where 0.6 is the probability of feeling displeased calculating from logistic function and 0.8 is the probability that an agent will focus on event. The probability that agent will be grateful (happy) is $0.6 \times 0.2 = 0.12$, where 0.6 is the same as happy case and 0.2 is the probability that an observed agent will focus on agent that causes the event. Therefore, an observed agent will be happy with probability 0.6, and no emotion with probability 0.4.

3.2 Display rules

A display rule for each emotion composes of a list of action units (AU) with a probability that it will show up on the face. This probability is $\Pr(x|d_e)$ in our equation. The list of AUs that we use in our simulation is the following: AU 1 - inner brow raiser, AU 4 - brow lowerer, AU 5 - upper lid raiser, AU 6 - cheek raiser, AU 12 - lip corner puller, AU 15 - lip corner depresser, AU 23 - lip tighter, and AU 25 - lips part. One example of display rule happy could be AU6, AU12, and AU25 with all of them having a probability 0.9, and the rest of action units with a probability 0.1. This means if an agent has this display rule, it is very likely that when an agent feels happy, AU6, AU12, and AU25 will show up while other action units likely to not show. Another example of happy rule could be AU6 and AU12 with both having a probability 0.25 representing a display rule of happy that unlikely to express smile. In the simulation, we define 3 display rules for each emotion, so there are 81 combinations of display rules.

3.3 Experiment

We run the simulation for each different possible combination of model and display rules. At each time step, the amount of extra money is randomly generated that an observed agent received. A distribution of observed agent's emotion is

generated based on the money and agent's model using OCC. Then one emotion is randomly generated from the distribution, and used to generate a set of action unit based on a observed agent's display rule of the emotion. Once both situational context and a set of shown action units are generated, we feed them to 4 different methods listing below to generate the prediction.

The first method which is the proposed method uses both situational context and facial expression to generate a prediction. In other word, it uses both observer agent's models of the observed goals and display rules (M and D). For the starting distribution of observed agent's model, every model is equally likely, so it has the same probability. For the starting distribution of display rules, the probability of the high display rule is 0.5 while the probability of other two rules is 0.25. Before testing the performance of this method, we first run a simulation on the same setup for 200 time steps to let the observer learns agent's model and display rules before testing in the simulation with other methods.

For the rest of the method, we do not train them. Instead, we provide them with agent's true model or display rules, or pre-defined display rules. The second method only uses context with true model of agent (M only), and ignore agent's facial expression. This method only applies OCC to a given situation and chooses emotion with highest probability to be a prediction of agent's emotion. Basically, this method only calculates $\Pr(e|m_{true}, c)$ or prior in equation 1 and uses the result to infer agent's emotion.

The third method uses only facial expression with true display rules for all emotion (D only), and ignore situational context. Essentially, this method only calculates $\Pr(X|e, c, D_{true})$ which is similar to likelihood in equation 3, and uses the result to infer agent's emotion.

The fourth method uses only facial expression, but with a high probability (or typical prototype) display rule for all types of emotion. In essence, it discards agent's model and display rules (No M and D). This method is similar to the third method but using a high probability display rule rather than the true display rule.

The simulation runs encompass 486 different agents, based on 6 different goal models times 81 different combination of display rules. For each of these agents, 100 simulations were run. Each simulation run encompassed an initial training session of 200 steps, followed by an evaluation phase of 500 steps. We calculate the accumulated error in predicting the emotion over these 500 steps for each method. If the method predicts observed agent's emotion correctly, then the error is 0. If it does not predict correctly, then the error is 1.

4 Simulation Results

On average, the proposed method took 105 time steps to converge (or need about 105 observations to converge) in which we define to be when the probability of one of the model is higher than 0.95. It fails to converge to the true observed agent's model only 1.78% of the time, but it always converges to the true observed agent's display rules for each emotion.

Table 1. Results of simulations. M stands for model and D stands for display rules. The error is the error in predicting the observed agent's emotion.

Error	M and D	M only	D only	No M and D
Maximum	0.1177	0.2540	0.2516	0.4078
Minimum	0.0213	0.1770	0.0486	0.0486
Mean	0.0637	0.2220	0.1425	0.1933
Standard Derivation	0.035	0.036	0.073	0.12

Table 1 shows the error in predicting an observed agent's emotion for each method. On average, the proposed method yields 6.37% error with SD = 0.035, while "M only" yields 22.2% error with SD = 0.036, "D only" yields 14.25% error with SD = 0.073, and "No M and D" yields 19.33% error with SD = 0.12. The proposed method yields a maximum error at 11.77% when, after training, it does not converge to the true model so it cannot predict emotion accurately. The minimum error for both "D only" and "No M and D" is only at 4.86% when the true display rules of observed agent are high probability display rules. It is important to note that the simplicity of simulation may have an effect on these errors.

5 Discussion and Future work

In this work, we propose a method to infer observed agent's emotion from prediction about emotional response and facial expression observations, and a way to update the observer's model of observed agent's goals and display rules that are needed to make the inference. To test the proposed method, a simulation study was created. The results of simulation show that the proposed model converges to the true model and display rules almost all the time. It also does better than a method with model alone, with display rules alone, and with only a high probability display rule.

There are several important problems that still need to be addressed. A key problem is how to acquire the information. In case of facial expression, some studies report success in accurately reading action units on the face [7]. For events, in a specific setting such as game or classroom, acquiring the relevant information needed for appraisal theory to predict emotion is feasible. For example, if an agent gets an answer wrong in the exercise, it is displeased event. Another problem is how much each information source contributes to help inferring observed agent's emotion. For example, facial expression may be a better predictor for happiness, but affective prosody may be a better predictor when it comes to angry or sad.

The next important step in our work is to validate our method with real humans. In our simulation, OCC is used to model the observed agent's emotional reaction, but if the observed is a human then OCC may not be an accurate model of the emotion elicitation process. Therefore, in order to further test our method, we need to replace simulated observed agents with human subjects, and let the

system try to predict human emotions based on various types of event that could elicit them.

All in all, this work demonstrates how to capture individual difference in descriptive Bayesian approach, and the way to update observer agent's distribution of models and display rules of observed agent to yield more accurate models. Moreover, this work argues for the importance of context, goals and display rules to make an accurate emotion inference.

References

1. Alfonso, B., Pynadath, D.V., Lhommet, M., Marsella, S.: Emotional perception for updating agents' beliefs. In: Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on. pp. 201–207. IEEE (2015)
2. Calvo, R., D'Mello, S., et al.: Affect detection: An interdisciplinary review of models, methods, and their applications. Affective Computing, IEEE Transactions on 1(1), 18–37 (2010)
3. Clore, G.L., Ortony, A.: Psychological construction in the occ model of emotion. Emotion Review 5(4), 335–343 (2013)
4. Ekman, P., Friesen, W.V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W.A., Pitcairn, T., Ricci-Bitti, P.E., et al.: Universals and cultural differences in the judgments of facial expressions of emotion. Journal of personality and social psychology 53(4), 712 (1987)
5. Ekman, P., Rosenberg, E.L.: What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press (1997)
6. Elfenbein, H.A., Beaupré, M., Lévesque, M., Hess, U.: Toward a dialect theory: cultural differences in the expression and recognition of posed facial expressions. Emotion 7(1), 131 (2007)
7. Happy, S., Routray, A.: Automatic facial expression recognition using features of salient facial patches. Affective Computing, IEEE Transactions on 6(1), 1–12 (2015)
8. Izard, C.E.: Innate and universal facial expressions: evidence from developmental and cross-cultural research. (1994)
9. Marsella, S.C., Pynadath, D.V., Read, S.J.: Psychsim: Agent-based modeling of social interactions and influence. In: Proceedings of the international conference on cognitive modeling. vol. 36, pp. 243–248. Citeseer (2004)
10. Ortony, A., Clore, G.L., Collins, A.: The cognitive structure of emotions. Cambridge university press (1990)
11. Safdar, S., Friedlmeier, W., Matsumoto, D., Yoo, S.H., Kwantes, C.T., Kakai, H., Shigemasu, E.: Variations of emotional display rules within and across cultures: A comparison between canada, usa, and japan. Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement 41(1), 1 (2009)
12. Smith, C.A., Lazarus, R.S.: Emotion and adaptation. (1990)
13. Tauber, S., Navarro, D.J., Perfors, A., Steyvers, M.: Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory
14. Whiten, A.: Natural theories of mind: Evolution, development and simulation of everyday mindreading. Basil Blackwell Oxford (1991)