Dynamic Probabilistic Graph Convolution for Facial Action Unit Intensity Estimation Supplemental Material

Tengfei Song[1], Zijun Cui[2], Yuru Wang[3*], Wenming Zheng[1*], and Qiang Ji[2]
[1]Southeast University
[2]Rensselaer Polytechnic Institute
[3]NorthEast Normal University
songtf@seu.edu.cn, cuiz3@rpi.edu, wangyr915@nenu.edu.cn,
wenming_zheng@seu.edu.cn, qji@ecse.rpi.edu

1. The generation of the posterior distribution $p(\mathcal{G}|\mathcal{D})$

1.1. The proof about obtaining the posterior distribution $p(\mathcal{G}|\mathcal{D})$ by Metropolis Hasting MCMC sampling

To clarify why Metropolis Hasting MCMC sampling can draw the distribution of the posterior distribution $p(\mathcal{G}|\mathcal{D})$, we provide more details. For the posterior distribution $p(\mathcal{G}|\mathcal{D})$ of $\mathcal{G}$, we assume that there are K kinds of different graph structures, $i.e.$, $\{\mathcal{G}_k\}_{k=1}^K$ and each kind of graph structure is a random variable. We sample N graphs $\{\mathcal{G}_n\}_{n=1}^N$ and calculate the probability for each kind of graph structure, $i.e.$, $p(\mathcal{G}_k|\mathcal{D}) = \frac{m_k}{N}$, in which m_k is the frequency that $\mathcal{G}_k$ appears during the sampling.

The N sampled $\{\mathcal{G}_n\}_{n=1}^N$ construct a Markov Chain and we have the transition probability $p(\mathcal{G}_n = \mathcal{G}_{k'}|\mathcal{G}_{n-1} = \mathcal{G}_k) = \mathcal{T}(\mathcal{G}_{k'}|\mathcal{G}_k)$ from $\mathcal{G}_k$ to $\mathcal{G}_{k'}$ with $\mathcal{G}_k \neq \mathcal{G}_{k'}$. For the n-th sampling step, the probability for being in state $\mathcal{G}_{k'}$ can be expressed as:

$$\begin{aligned}p(\mathcal{G}_n = \mathcal{G}_{k'}) &= p(stay \quad at \quad \mathcal{G}_{k'}) + p(move \quad to \quad \mathcal{G}_{k'}) \\ &- p(move \quad from \quad \mathcal{G}_{k'}) \\ &= p(\mathcal{G}_{n-1} = \mathcal{G}_{k'}) + \sum_{\mathcal{G}_{k'} \neq \mathcal{G}_k} p(\mathcal{G}_{n-1} = \mathcal{G}_k)\mathcal{T}(\mathcal{G}_{k'}|\mathcal{G}_k) \\ &- \sum_{\mathcal{G}_{k'} \neq \mathcal{G}_k} p(\mathcal{G}_{n-1} = \mathcal{G}_{k'})\mathcal{T}(\mathcal{G}_k|\mathcal{G}_{k'}) \\ &= p(\mathcal{G}_{n-1} = \mathcal{G}_{k'}) + \sum_{\mathcal{G}_{k'} \neq \mathcal{G}_k} [p(\mathcal{G}_{n-1} = \mathcal{G}_k)\mathcal{T}(\mathcal{G}_{k'}|\mathcal{G}_k) \\ &- p(\mathcal{G}_{n-1} = \mathcal{G}_{k'})\mathcal{T}(\mathcal{G}_k|\mathcal{G}_{k'})]. \qquad (11)\end{aligned}$$

Without the summation term, there is an equilibrium distribution:

$$p(\mathcal{G}_n = \mathcal{G}_{k'}) = p(\mathcal{G}_{n-1} = \mathcal{G}_{k'}) \equiv p_{eq}(\mathcal{G}_{k'}). \qquad (12)$$

We have a sufficient (but not necessary) condition to keep the equilibrium distribution $p_{eq}(\mathcal{G}_{k'})$ is

$$p_{eq}(\mathcal{G}_k)\mathcal{T}(\mathcal{G}_{k'}|\mathcal{G}_k) = p_{eq}(\mathcal{G}_{k'})\mathcal{T}(\mathcal{G}_k|\mathcal{G}_{k'}). \qquad (13)$$

During the sampling, if we may have

$$p(\mathcal{G}_k)\mathcal{T}(\mathcal{G}_{k'}|\mathcal{G}_k) > p(\mathcal{G}_{k'})\mathcal{T}(\mathcal{G}_k|\mathcal{G}_{k'}), \qquad (14)$$

we impose the acceptance probability $\alpha(\mathcal{G}_{k'}|\mathcal{G}_k)$ to accept $\mathcal{G}_{k'}$ and try to close the equilibrium distribution with

$$p(\mathcal{G}_k)\mathcal{T}(\mathcal{G}_{k'}|\mathcal{G}_k)\alpha(\mathcal{G}_{k'}|\mathcal{G}_k) = p(\mathcal{G}_{k'})\mathcal{T}(\mathcal{G}_k|\mathcal{G}_{k'})\alpha(\mathcal{G}_k|\mathcal{G}_{k'}). \qquad (15)$$

If $p(\mathcal{G}_k)\mathcal{T}(\mathcal{G}_{k'}|\mathcal{G}_k) > p(\mathcal{G}_{k'})\mathcal{T}(\mathcal{G}_k|\mathcal{G}_{k'})$, we want to suppress the transition to $\mathcal{G}_{k'}$ and maximize the transition to $\mathcal{G}_k$. We should set the $\alpha(\mathcal{G}_k|\mathcal{G}_{k'}) = 1$ and the acceptance probability $\alpha(\mathcal{G}_{k'}|\mathcal{G}_k)$ will be

$$\alpha(\mathcal{G}_{k'}|\mathcal{G}_k) = \frac{p(\mathcal{G}_{k'})\mathcal{T}(\mathcal{G}_k|\mathcal{G}_{k'})}{p(\mathcal{G}_k)\mathcal{T}(\mathcal{G}_{k'}|\mathcal{G}_k)}. \qquad (16)$$

If $p(\mathcal{G}_k)\mathcal{T}(\mathcal{G}_{k'}|\mathcal{G}_k) < p(\mathcal{G}_{k'})\mathcal{T}(\mathcal{G}_k|\mathcal{G}_{k'})$, we want to maximize the transition to $\mathcal{G}_{k'}$ such that we set

$$\alpha(\mathcal{G}_{k'}|\mathcal{G}_k) = 1. \qquad (17)$$

So, if we want to keep the distribution of $\mathcal{G}_k$ converge to an equilibrium distribution, we should have the acceptance probability

$$\alpha(\mathcal{G}_{k'}|\mathcal{G}_k) = \begin{cases} \frac{p(\mathcal{G}_{k'})\mathcal{T}(\mathcal{G}_k|\mathcal{G}_{k'})}{p(\mathcal{G}_k)\mathcal{T}(\mathcal{G}_{k'}|\mathcal{G}_k)} & \text{if} \quad p(\mathcal{G}_k)\mathcal{T}(\mathcal{G}_{k'}|\mathcal{G}_k) > p(\mathcal{G}_{k'})\mathcal{T}(\mathcal{G}_k|\mathcal{G}_{k'}) \\ 1 & \text{otherwise,} \end{cases} \qquad (18)$$

which is the same as

$$\alpha(\mathcal{G}_{k'}|\mathcal{G}_k) = \min\{1, \frac{p(\mathcal{G}_{k'})\mathcal{T}(\mathcal{G}_k|\mathcal{G}_{k'})}{p(\mathcal{G}_k)\mathcal{T}(\mathcal{G}_{k'}|\mathcal{G}_k)}\}. \qquad (19)$$

*Corresponding author

In this paper, we have the acceptance probability with $\mathcal{G}_{k'}$ and $\mathcal{G}_k$ are neighbor graphs

$$p^{acc}(\mathcal{G}_{n+1} = \mathcal{G}_{k'}|\mathcal{G}_n = \mathcal{G}_k) = \min\{1, \frac{p(\mathcal{G}_{n+1}|\mathcal{D})}{p(\mathcal{G}_n|\mathcal{D})}\frac{|\mathcal{N}(\mathcal{G}_n)|}{|\mathcal{N}(\mathcal{G}_{n+1})|}\}, \tag{20}$$

in which we have the transition probabilities

$$p(\mathcal{G}_{n+1} = \mathcal{G}_{k'}|\mathcal{G}_n = \mathcal{G}_k) = \mathcal{T}(\mathcal{G}_{k'}|\mathcal{G}_k) = \frac{1}{|\mathcal{N}(\mathcal{G}_n)|} \tag{21}$$

$$p(\mathcal{G}_n = \mathcal{G}_k|\mathcal{G}_{n+1} = \mathcal{G}_{k'}) = \mathcal{T}(\mathcal{G}_k|\mathcal{G}_{k'}) = \frac{1}{|\mathcal{N}(\mathcal{G}_{n+1})|.} \tag{22}$$

We can't directly calculate the $p(\mathcal{G}_{n+1}|\mathcal{D})$, but we can calculate the $p(\mathcal{D}|\mathcal{G}_{n+1})$ and $p(\mathcal{D}|\mathcal{G}_n)$. And then, we can get the $\frac{p(\mathcal{G}_{n+1}|\mathcal{D})}{p(\mathcal{G}_n|\mathcal{D})}$ since $\frac{p(\mathcal{G}_{n+1}|\mathcal{D})}{p(\mathcal{G}_n|\mathcal{D})} = \frac{p(\mathcal{D}|\mathcal{G}_{n+1})}{p(\mathcal{D}|\mathcal{G}_n)}$. Based the on the acceptance probability to generate samples of graph structure, we can get the converged posterior distribution $p(\mathcal{G}|\mathcal{D})$.

1.2. Pseudo code of the generation of $p(\mathcal{G}|\mathcal{D})$

Algorithm 1 Obtain $p(\mathcal{G}|\mathcal{D})$ by Metropolis Hasting MCMC sampling

Require: AU intensity annotations $\mathcal{D}$
Ensure: N samples $\{\mathcal{G}_n\}_{n=1}^N$

```
Initial graph structure G_1;
for n = 1 to N :
    Generate a new sample G_{n+1} based on proposal probability Equ (3);
    Calculate the marginal likelihood p(D|G_{n+1}) by Equ (2);
    Calculate the acceptance probability p^{acc}(G_{n+1}|G_n) by Equ (4);
    if accept:
        G_{n+1} = G_{n+1};
    else:
        G_{n+1} = G_n;
    end
end
```

2. The visualization of AU changes

The visualization of images To show how the actual facial expression looks like, we show some specific examples on FERA2015 in Figure 8. From Figure 5 (a), we can see that the graph structure with the highest probability contains four links, i,e. the link between AU6 and AU10, the link between AU12 and AU14, the link between AU6 and AU12 and the link between AU12 and AU17. All the four people in Figure 8 have the happy expression and they show different AU changes from low intensity to high intensity, which can reflect the dependencies the graph structure with the highest probability in Figure 5 (a). For Figure 7 (a), AU6 and AU10 change simultaneously from low intensity to high intensity. Figure 7 (b) shows the dependencies between AU12 and AU14. Figure 7 (c) shows the dependencies between AU6 and AU12. Figure 7 (d) shows the dependencies between AU12 and AU17.

Table 3. AU correlations from anatomy defined in FACS[1].

AU correlation	AUs
positive	(1,2), (4,7), (4,9), (6,12), (9,17), (15,24), (17,24), (23,24)
negative	(2,6), (2,7), (12,15), (12,17)

3. The AU correlations in FACS

The AU correlations defined in FACS [1] are shown in Table 3. We employed these AU correlations as the prior knowledge to construct the graph for DPG-PK.

References

[1] Paul Ekman and Wallace V Friesen. *Manual for the facial action coding system*. Consulting Psychologists Press, 1978. 2

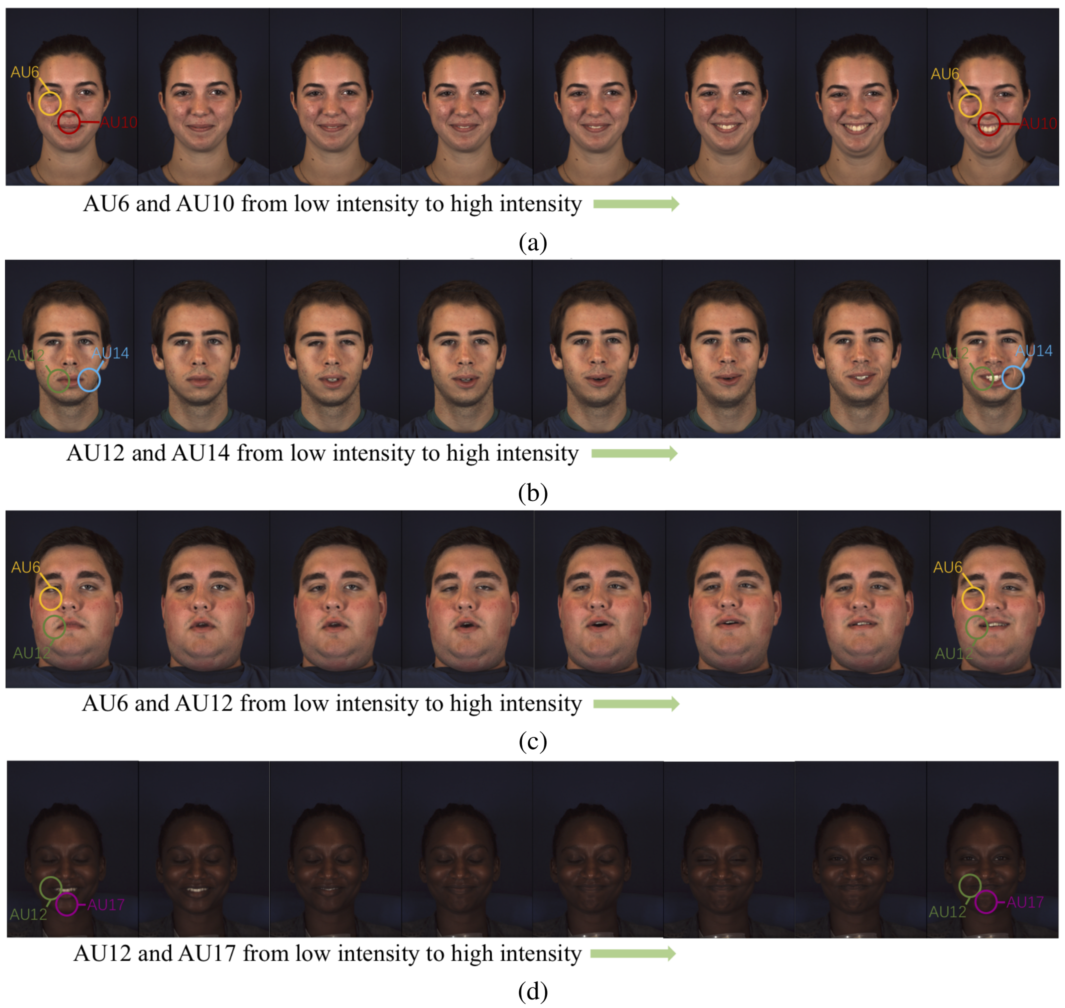

Figure 8. The visualizatoin of AU change on FERA2015 (a) AU6 and AU10. (b) AU12 and AU14. (c) AU6 and AU12. (d) AU12 and AU17.