DYNAMIC MULTI-RATER GAUSSIAN MIXTURE REGRESSION INCORPORATING TEMPORAL DEPENDENCIES OF EMOTION UNCERTAINTY USING KALMAN FILTERS

Ting Dang\textsuperscript{1,2}, Vidhyasaharan Sethu\textsuperscript{1}, Eliathamby Ambikairajah\textsuperscript{1,2}

\textsuperscript{1} School of Electrical Engineering and Telecommunications, UNSW, Australia
\textsuperscript{2} ATP Research Laboratory, DATA61 (CSIRO), Australia
1. Continuous Emotion Prediction
2. Inter-rater Variability
3. Dynamic multi-rater GMR
4. Experimental Results
5. Conclusion
Continuous Emotion Prediction

- **Emotion Representation**

- **Categorical Representation**
  - Happy, anger, sad, etc.

- **Dimensional Representation**
  - Affective attribute: arousal, valence
Continuous Emotion Prediction

![Diagram showing the process of continuous emotion prediction](image)

**Training Phase**
- Pre-processing
- Feature Extraction
- Modelling
- Regression Model

**Test Phase**
- Pre-processing
- Feature Extraction
- Prediction

**Graph**
- **Y-axis**: Valence
- **X-axis**: Time (s)
- Line graph showing valence over time
Continuous Emotion Prediction

![Image of emotion prediction graph and diagram]

- **Valence** axis ranges from -0.1 to 0.5.
- **Time/s** axis ranges from 0 to 1.
- Three raters (Rater 1, Rater 2, Rater 3) are shown with their valence curves.
- Average curve is plotted.

**Training Phase**:
- Pre-processing
- Feature Extraction
- Modelling
- Regression Model

**Test Phase**:
- Pre-processing
- Feature Extraction
- Prediction

Speech with known labels flows through the training and test phases.

Speech with unknown labels flows through the test phase, where it is used for prediction.
Inter-rater Variability

- Averaging ratings ignore the discrepancies between raters.
- Intense emotions are easier to recognize while the subtle emotions are more ambiguous.
- Other factors (i.e. recording conditions) may affect rater’s judgements.
Inter-rater Variability

3 raters

Average

Valence Rating

Time/s

Conventional prediction of mean ratings
• Gaussian assumption of label distribution may not be true

• Multi-rater Gaussian mixture regression (GMR) does not consider temporal dependencies
Inter-rater Variability

3 raters

Average

Valence Rating

Multi-rater GMR

Dynamic Multi-rater GMR

Kalman filter $P(y_{t+1} \mid y_t)$

Time/s
Dynamic multi-rater GMR

- Incorporation of both forward and backward Kalman filters into multi-rater GMR to account for the temporal dependencies in both directions.
- Label distribution given by GMM instead of single Gaussian.
- Measure to quantify uncertainty from predicted distribution (GMM).
Gaussian Mixture Regression (GMR)

- **GMR model**
  - Joint distribution of feature vectors and labels

\[ \lambda(z) = P(x, y) \]

- **Probability distribution**
  - Training vectors are generated by concatenating the feature vector and *mean rating*

- **Incorporation of uncertainty**
  - Training vectors are generated by concatenating the feature vector and *individual annotation*
Gaussian Mixture Regression (GMR)

- Predicted label distribution
  - Predicted label distribution (GMM)
  - Approximated label distribution (Gaussian)

Dominant mixture component to approach the label distribution
Gaussian Mixture Regression (GMR)

Plot of uncertainty of emotion predictions

- Standard deviation of six raters correlates with the predicted uncertainty of emotion

Limitations

- The assumption of Gaussianity over label distribution may not hold true
- GMR does not model temporal dependencies between frames
Dynamic multi-rater GMR

- Adopting predicted GMM distribution directly
- Kalman filter is adopted to explore the temporal dependencies

Noisy observations of underlying label distribution (predictions independent of other frames)

Incorporating temporal dependencies
Dynamic multi-rater GMR

- Adopting predicted GMM distribution directly
- Kalman filter is adopted to explore the temporal dependencies
- Vector representation of GMM distributions is adopted by Kalman filter

\[
\mathbf{v}_t = [\bar{w}_{1t}, \ldots, \bar{w}_{M_1t}, \bar{\mathbf{u}}_{1t}^T, \ldots, \bar{\mathbf{u}}_{M_1t}^T, \text{Vec}(\bar{\mathbf{\Sigma}}_{1t})^T, \ldots, \text{Vec}(\bar{\mathbf{\Sigma}}_{M_1t})^T]^T
\]

\[
\mathbf{s}_t = [w_{1t}, \ldots, w_{M_2t}, \mathbf{u}_{1t}^T, \ldots, \mathbf{u}_{M_2t}^T, \text{Vec}(\mathbf{\Sigma}_{1t})^T, \ldots, \text{Vec}(\mathbf{\Sigma}_{M_2t})^T]^T
\]
Dynamic multi-rater GMR

Kalman filter

- $v_t$ is treated as the observation of label distribution and $s_t$ is the underlying distribution that depends on the long-term dynamics

\begin{align*}
  s_t &:= \text{underlying label distribution} \\
  v_t &:= \text{noisy observation of } s_t \\

  s_t &= Fs_{t-1} + w_{t-1} \quad (\text{noise } w_{t-1} \sim N(0, Q)) \\
  v_t &= Hs_t + r_t \quad (\text{noise } r_t \sim N(0, R))
\end{align*}

- During training phase, parameters of Kalman filters ($F$, $Q$, $H$ and $R$) are estimated where the observations $v_t$ and the ground truth $s_t$ are known.

- During test phase, Kalman filters are utilised to predict the label distribution $\hat{s}_t$ based on the GMR prediction $v_t$ and the prediction of previous frames $\hat{s}_{t-1}$
Dynamic multi-rater GMR

- Forward and backward Kalman filter

- Forward

- Backward

- Final label prediction

\[
\hat{s}_t = \alpha s_t^{KF1} + (1 - \alpha) s_t^{KF2}
\]

\[
\hat{s}_t \quad [w, u, \Sigma]
\]
Measures of Uncertainty

Probabilistic uncertainty volume \( PUV \)

- Probabilistic uncertainty volume estimates the local variability of a distribution

\[
PUV_t = \int f(y) \, dy, \quad f(y) = \begin{cases} 
1, & P(y_t) > \theta \\
0, & P(y_t) \leq \theta 
\end{cases}
\]

- Given threshold \( \theta \), \( PUV_1 \) for a broad GMM (high uncertainty in left side) is larger than \( PUV_2 \) for a narrow GMM (low uncertainty in right side)
System Evaluation

- System evaluation focuses on the comparison between predicted and underlying label distributions.

**Predicted by system**

**Inferred from annotations (multiple raters)**
Evaluation Metrics

- Underlying label distribution (GMM) is time-dependent and estimated in the label space by 6 annotations.
Evaluation Metrics

- Underlying label distribution is time-dependent and estimated in the label space by 6 annotations
- Probabilistic uncertainty volume $PUV$ is estimated for the predicted and underlying label distribution respectively for each frame

➢ Correlation coefficient (CC)

- Pearson’s correlation coefficient between probabilistic uncertainty volume estimated from the predicted and the underlying label distribution
- A higher CC indicates better predicted label distributions
Evaluation Metrics

- Underlying label distribution is time-dependent and estimated in the label space by 6 annotations

KL divergence

- KL divergence estimates the similarity between the predicted and the underlying label distributions
- A smaller KL divergence indicates better predicted label distributions
- Median and 25th and 75th percentiles of KL divergence over entire test dataset are estimated (boxplot)
Experimental Settings

➢ Training phase

Feature Extraction (650)

PCA (40)

Features (80)

$X_i = [x_i^T, \Delta x_i^T]^T$

Labels (2)

$Y_{ik} = [y_{ik}^T, \Delta y_{ik}^T]^T$, $k = 1, \ldots, 6$

GMM (82)

Prediction

$v_i = P(y_i \mid x_i, \lambda^{(x,y)})$

Kalman filter

'Ground truth'

$s_i = P(y_i)$

Kalman model $F, Q, H, R$

➢ Test phase

Feature Extraction (650)

PCA (40)

Features (80)

$X_i = [x_i^T, \Delta x_i^T]^T$

GMM $\lambda^{(x,y)}$

Prediction

$v_i = P(y_i \mid x_i, \lambda^{(x,y)})$

Kalman model $F, Q, H, R$

Predicted label distribution

$\hat{s}_i = P(\hat{y}_i)$

* Github: https://github.com/TingDang90/Dynamic-multi-rater-GMR
Experimental Settings

- Database: RECOLA (6 annotations)
- Features: 5 functionals applied to 130 LLDs
- PCA: 40 dimensions
- Delays: 2s for arousal and 4s for valence
- GMM mixture number: [2,4,8]
- Linear coefficient of Kalman filter: [0, 1] with a step increase of 0.1

- Baseline:
  --- Multi-rater GMR system
    i. CC between the PUV of predicted Gaussian and PUV of underlying label distribution
    ii. KL between the predicted Gaussian and the underlying label distribution (GMM)
Experimental Results

» CC between predicted and true PUV

- CC between the PUV of the predicted and underlying label distributions (GMM)

- Incorporating temporal dependencies benefits uncertainty prediction, especially for valence

![Bar chart](chart1.png)

(a) arousal

![Bar chart](chart2.png)

(b) valence
Experimental Results

KL divergence between predicted and underlying label distributions

- KL between the predicted and underlying label distribution (GMM) is computed.

- The proposed system leads to more reliable and smoothed distribution prediction.
Conclusion

• A dynamic multi-rater GMR to predict emotion uncertainty by considering the temporal dependencies is proposed, which is achieved by applying Kalman filters.

• Probabilistic uncertainty volume is introduced as a measure to quantify uncertainty of emotion predictions (GMM).

• The statistics of KL divergence between predicted and underlying label distributions indicate that incorporating temporal dependencies between frames leads to a smoother change in the label distributions.

• Future work will focus on relaxing linearity assumption about the evolution of emotion label distributions.
Reference


Thank you
Thresholds of Probabilistic Uncertainty Volume

- Thresholds $\theta_i$ are defined in terms of percentiles of all the probabilities calculated by fitting the test features to the GMM models.

- The optimal threshold $\theta_i$ is determined experimentally based on the system performance.
CC between PUV from predicted and underlying distributions

Optimal threshold $\theta_i$ for arousal

![Graph showing correlation between PUV estimated from predicted and underlying label distributions vs. threshold in percentiles. The correlation starts at 0.45 for the 80th percentile and increases to around 0.5 for the 98th percentile, indicating an optimal threshold.](image-url)
Smoothness of PUV from underlying label distribution

- Prediction
- Ground truth

- Smoothed Ground truth
- Smoothed Prediction

Mean rating

Frame
KL divergence

• Symmetric KL divergence is utilised, with a larger KL divergence indicating a greater separation between them.

• Specifically, a Monte-Carlo estimate of the symmetric KL divergence proposed in [11] is utilised to quantify the separation between two distributions.

\[
I_{SKL}(P_1, P_2) = \frac{1}{2} \left| \int_x P_1(x) \ln \frac{P_1(x)}{P_2(x)} \, dx + \int_x P_2(x) \ln \frac{P_2(x)}{P_1(x)} \, dx \right| \tag{3.1}
\]
Experimental Results

KL divergence between predicted and underlying label distributions

<table>
<thead>
<tr>
<th></th>
<th>Arousal</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proposed</td>
<td>Baseline</td>
</tr>
<tr>
<td>Mean</td>
<td>0.1439</td>
<td>1.6872</td>
</tr>
<tr>
<td>SD</td>
<td>0.1818</td>
<td>7.2714</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Valence</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proposed</td>
<td>Baseline</td>
</tr>
<tr>
<td>Mean</td>
<td>0.2085</td>
<td>1.8628</td>
</tr>
<tr>
<td>SD</td>
<td>0.2044</td>
<td>1.1236</td>
</tr>
</tbody>
</table>

--- Baseline means the KL divergence calculated between predicted and underlying GMM distributions.

--- The proposed system leads to more reliable and smoothed distribution prediction.
KL divergence

Baseline  Proposed  Baseline  Proposed
Arousal    Valence
Uncertainty Prediction using Kalman filters
Figure. 25-75% quartile plotted as error bar, with 6 true annotations overlaid. 
utterance 2 in dev set
Yellow: predicted GMM(ESN); Cyan: assumed ‘ground truth’; Green: predicted GMM(Kalman filter)
Left: utterance 4 in dev set;
Kalman filter

\[ P(s_t | s_{t-1}) = N(s_t; Fs_{t-1}, Q) \]

\[ P(v_t | s_t) = N(v_t; Hs_{t-1}, R) \]

\[ F = (A^T A + \lambda I)^{-1} A^T B \]

\[ Q = \text{cov}(B - AF) \]

\[ H = (C^T C + \lambda I)^{-1} C^T D \]

\[ R = \text{cov}(D - CH) \]
CC between the standard deviation (SD) of predicted Gaussians and PUV (ground truth)

<table>
<thead>
<tr>
<th></th>
<th>Arousal</th>
<th>Valence</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 mix</td>
<td>0.0050</td>
<td>0.008</td>
</tr>
<tr>
<td>4 mix</td>
<td>0.3726</td>
<td>0.075</td>
</tr>
<tr>
<td>8 mix</td>
<td>0.4632</td>
<td>0.1243</td>
</tr>
<tr>
<td>cc</td>
<td>0.2392</td>
<td>0.0512</td>
</tr>
</tbody>
</table>
(a) arousal

(b) valence