Biweekly Meeting

Juhyeong Kim

SKKU Visual Computing Lab

wngud0811@naver.com

August 12, 2019

- Uncertainty aware audiovisual activity recognition using deep Bayesian variational inference(CVPR workshop 2019)

Importantce of uncertainty Quantification

- Deep learning models may fail in the case of noisy data or out-of-distribution data.

Train time

rabbit: **0.8** / wolf: 0.2

rabbit: 0.3 / wolf: **0.7**

Test time

rabbit: 0.1 / wolf: **0.9**

1. **Impications**
    - SoftMax cannot capture confidence of model.
    - In these situations, uncertainty quantification is needed.
2. **Frequentist DNN vs Bayesian DNN**
    - Frequentist DNN has low computaitional costs at training time and inference time.
    - Bayesian DNN can quantify uncertainty by calculating predictive distribution.

Uncertainty aware audiovisual activity recognition using deep Bayesian variational inference(CVPR workshop 2019)

- Solve activity recognition problem by combining standard DNN and Bayesian DNN.
- Utilize information from multiple source(Video, Audio).
- Apply Bayesian learning framework while maintaining scalability.
- Quantify uncertainty and detect out-of-distribution example.

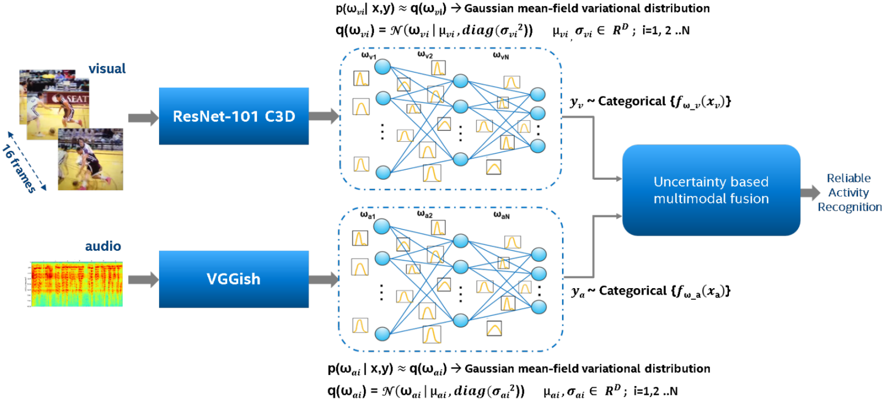
Method

1. Use pretrained model to extract deterministic feature in video domain and audio domain and
2. Remove last layer of pretrained models from each domain.
3. Add three Variational layers for each models.
   $q_\theta(w) = \mathbf{N}(w|\mu, \sigma^2)$

Method

4. Train variational layer with ELBO:
   $L^v = -\mathbb{E}_{q_\theta(w)}[\log p(y|x,w)] + KL[q_\theta(w)||p(w)]$
5. Fine-tune deterministic models with ML Loss:
   $L^d = -\sum_c y_c \log \hat{y_c}$
6. Combine prediction and uncertainty result from two information sources to make final descision.
7. If prediction from one source has high uncertainty, only use information from sources with lower uncertainty.

Method



$p(\omega_{vi} | x,y) \approx q(\omega_{vi})$ → Gaussian mean-field variational distribution
$q(\omega_{vi}) = \mathcal{N}(\omega_{vi} | \mu_{vi}, diag(\sigma_{vi}^2))$ $\mu_{vi}, \sigma_{vi} \in R^D$; i=1, 2 ..N
visual
16 frames
ResNet-101 C3D
ω_{v1}
ω_{v2}
ω_{vN}
y_v ~ Categorical $\{f_{\omega\_v}(x_v)\}$
Uncertainty based multimodal fusion
Reliable Activity Recognition
audio
VGGish
ω_{a1}
ω_{a2}
ω_{aN}
y_a ~ Categorical $\{f_{\omega\_a}(x_a)\}$
$p(\omega_{ai} | x,y) \approx q(\omega_{ai})$ → Gaussian mean-field variational distribution
$q(\omega_{ai}) = \mathcal{N}(\omega_{ai} | \mu_{ai}, diag(\sigma_{ai}^2))$ $\mu_{ai}, \sigma_{ai} \in R^D$; i=1,2 ..N

Evaluation

1. Unceratinty Metric
   - $p(y^*|x^*, D) = \int p(y^*|x^*, w) q_\theta(w)\, dw$
   - $p(y^*|x^*, D) \approx \frac{1}{T}\sum_{i=1}^{T} p(y^*|x^*, w), w_i \sim q_\theta(w)$
   - Bayesian Active Learning by Disagreement(BALD) $= H(y^*|x^*, D) - \mathbb{E}_{p(w|D)}[H(y^*|x^*, w)]$
     $H(y^*|x^*, D) = -\sum_{i=0}^{K-1} p_{i\mu} \log p_{i\mu}$

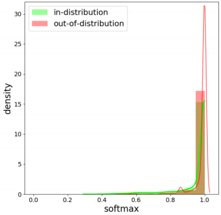
2. Generalization Performance

	certain	uncertain
accurate	n_{ac}	n_{au}
inaccurate	n_{ic}	n_{iu}

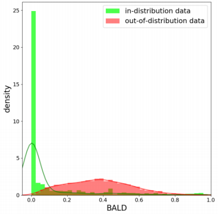
- Accuracy vs Uncertainty(AvU) $= \frac{n_{ac}+n_{ic}}{n_{ac}+n_{au}+n_{ic}+n_{iu}}$

Result

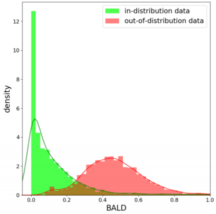
- Histogram comparing in- and out-distribution data predictions.



(a) DNN
confidence measure

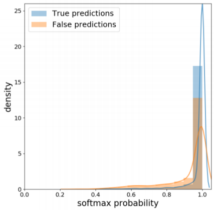


(b) Bayesian DNN (MC Dropout)
uncertainty measure

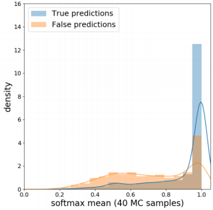


(c) Bayesian DNN (Stochastic VI)
uncertainty measure

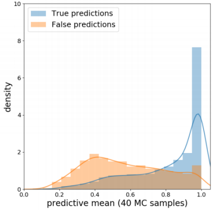
- Histogram comparing True prediction and wrong predictions.



(a) DNN model



(b) Bayesian DNN (MC Dropout) model



(c) Bayesian DNN (Stochastic VI) model

Result

- Generalization Performace comparision

Model	Top1 (%)	Top5 (%)
Vision		
DNN	52.65	79.79
Bayesian DNN (MC Dropout)	52.88	80.10
Bayesian DNN (Stochastic VI)	53.3	81.20
Audio		
DNN	34.13	61.68
Bayesian DNN (MC Dropout)	32.46	60.97
Bayesian DNN (Stochastic VI)	35.80	63.40
Audiovisual		
DNN	56.61	79.39
Bayesian DNN (MC-Dropout)	55.04	80.34
Bayesian DNN (Stochastic VI)	**58.2**	**83.8**

Conclusion

- Combination of deterministic and variational layer in single model works well.
- Using Bayesian DNN, we can quantify uncertainty and detect out-distribution data.
- By considering uncertainty, model can automatically decide whether it will use information from specific sources.
- MC-dropout may yield bad quality of uncertainty.