Trustworthy Localized Corrections-guided Mutual Learning for Multi-View Learning

Qiuran Li[1,3], Yi Luo[2,3], Yan Sun[3], Tong Wu[2,3], Aiguo Chen[3,4,†]

[1] *School of Computer Science and Engineering, University of Electronic Science and Technology of China, China;*
[2] *School of Information and Software Engineering, University of Electronic Science and Technology of China, China*;
[3] *Laboratory of Intelligent Collaborative Computing, University of Electronic Science and Technology of China, China;*
[4] *Ubiquitous Intelligence and Trusted Services Key Laboratory of Sichuan Province*

TABLE I
DETAILS OF EACH DATASET

Dataset	Size	K	Dimensionality
Caltech101	8677	101	4096/4096
CUB	11788	10	1024/300
HandWritten	2000	10	240/76/216/47/64/6
PIE	680	68	484/256/279
Scene15	4485	15	20/59/40

I. PROOF 1

Proof. An upper bound, denoted by M, exists upon convergence. The inequation is given by:

$$0 \leq \left(log(S_f) - log(S_{ww})\right)^2 \leq M$$

Assume that $S_f \geq S^{ww}$, which then obtain the following formula:

$$0 \leq log(S_f) - log(S_{ww}) \leq \sqrt{M}$$

$\Delta_S = S_f - S_{ww}$ is calculated as follows:

$$\begin{aligned}\Delta_S &= S_{ww}\left(e^{log(S_f)-log(S_{ww})} - 1\right)\\ &\leq S_{ww}\left(e^{\sqrt{M}} - 1\right)\end{aligned}$$

Similarly, the lower bound of Δ_S is calculated as:

$$\Delta_S \geq S_{ww}\left(1 - e^{-\sqrt{M}}\right)$$

In summary, the scope of Δ_S is determined by the following inequation:

$$S_{ww}\left(1 - e^{-\sqrt{M}}\right) \leq \Delta_S \leq S_{ww}\left(e^{\sqrt{M}} - 1\right)$$

□

II. DATASETS

Details of each dataset is shown in Table I. **Caltech101** [1] comprises 8766 images across 11 categories, each represented by two views. **CUB** [2] comprises 11788 images, each paired with textual descriptions, spanning 200 different categories of birds. **HandWritten** [3] comprises 2000 samples, evenly distributed across 10 classes corresponding to the digits '0' to '9', with each class containing 200 samples. **PIE** [4] comprises 680 facial images from 68 distinct subjects. Three feature types, intensity, LBP, and Gabor, are extracted from each image. **Scene15** [5] comprises 4485 images spanning 15 categories, encompassing both indoor and outdoor scenes. Three types of features GIST, PHOG, and LBP are extracted from the dataset.

†Corresponding author: Aiguo Chen.
Email: qiuranli@std.uestc.edu.cn (Qiuran Li),
agchen@uestc.edu.cn (Aiguo Chen).

Algorithm 1 TLCML algorithm

/*Training*/
Input: Multi-View dataset: $D_N^M = \{\{\boldsymbol{x}_n^m\}^M, \boldsymbol{y}_n\}_N$
Output: Parameters of model

Initialize the parameters of model.
while not converged **do**
  for m=1:M **do**
    $\boldsymbol{e}^m \leftarrow$ view-specific MENN;
    Subjective opinion $\leftarrow$ Dirichlet distribution;
    Obtain the loss of view-specific MENN with Eq. 3;
  end for
  Concatenate all of evidence;
  Obtain joint opinion with TML module;
  Obtain the joint loss of with Eq. 3;
  Obtain the consistency loss with Eq. 7;
  Obtain the boundary loss with Eq. 9;
  Obtain the overall loss with Eq. 11;
  Update the parameters;
end while

/*Test*/
Calculate the decisions and corresponding uncertainty.

III. ALGORITHM

The detailed procedure of our method had been comprehensively outlined in Algorithm 1.

REFERENCES

[1] Fei-Fei Li, Marco Andreeto, Marc'Aurelio Ranzato, and Pietro Perona, "Caltech 101," 4 2022.

[2] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, "cub," Tech. Rep. CNS-TR-2011-001, California Institute of Technology, 2011.

[3] Robert Duin, "Multiple Features," UCI Machine Learning Repository, 1998, DOI: https://doi.org/10.24432/C5HC70.

[4] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and S. Baker, "Multi-pie," 12 2013.
[5] Nouman Ali and Bushra Zafar, "15-Scene Image Dataset," 8 2018.