보행자 시점의 멀티 도메인 교차로 분류 문제에서의 도메인 제거 손실 함수 비교

Marcella Astrid[1,2], Muhammad Zaigham Zaheer[1,2], 이승익 [1,2]
[1] 과학기술연합대학원대학교, [2] 한국전자통신연구원

{marcella.astrid, mzz}@ust.ac.kr, the_silee@etri.re.kr

Comparisons of Domain−Removal Losses in Multi−Domain Pedestrian−View Intersection Classification

Marcella Astrid[1,2], Muhammad Zaigham Zaheer[1,2], Seung−Ik Lee[1,2]
[1]University of Science and Technology
[2]Electronics and Telecommunications Research Institute

Abstract

For pedestrian−view intersection classification model that can run on multi−domain input, i.e., indoor and outdoor, we adopt method from Kim *et al.*, [1]. The training objective is to not extract domain related features by using negative entropy loss on domain prediction. However, with similar goal, Alvi *et al.* [2] utilize confusion loss, i.e., cross entropy to uniform distribution. In this work, we compare the two losses on our multi−domain pedestrian−view intersection classification model.

Ⅰ. Introduction

To assist navigation of small robots, several researchers [3, 4] propose pedestrian−view intersection classification methods. However, these models are implemented only on one domain, i.e., outdoor domain. To incorporate both indoor and outdoor domains, we adopt method from Kim *et al.* [1] in order to remove the domain information in the extracted features. In this paper, we focus on comparing two losses, i.e., negative entropy and confusion loss, utilized in multi−domain training to reduce domain information in the extracted features.

Ⅱ. Methodology

We adopt model from Kim *et al.* [1] to train our multi−domain intersection classification model. As seen in Fig. 1, the architecture consists of two branches. The first branch is utilized to predict intersection categories $\hat{\boldsymbol{y}} \in \mathbb{R}^2$ and the second branch is utilized for domain prediction $\hat{\boldsymbol{z}} \in \mathbb{R}^2$. The training objective is to prevent the base network $\mathcal{F}$ to extract domain information while still extracting the features related to the intersection classification. To correctly classify the intersection, we train the model to minimize cross entropy loss between prediction $\hat{\boldsymbol{y}}$ and its ground truth $\boldsymbol{y}$. In addition to the loss, to remove the domain from the features extracted by the $\mathcal{F}$, Kim *et al.* [1] incorporate negative entropy loss which has to be minimized by $\mathcal{F}$ as:

$$\min_{\mathcal{F}} \alpha \left(\sum_{i=1}^{2} \hat{z}_i \log \hat{z}_i \right), \tag{1}$$

where α is weighting hyperparameter. In order to minimize this loss, $\mathcal{F}$ should not extract any domain information so that the entropy of $\hat{\boldsymbol{z}}$ is maximized, i.e., unconfident $\hat{\boldsymbol{z}}$ prediction. The loss value and its gradient are visualized in Fig. 2(a).

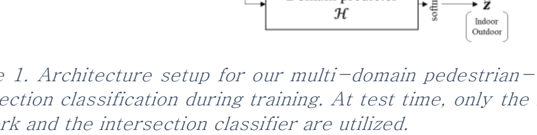

Figure 1. Architecture setup for our multi−domain pedestrian−view intersection classification during training. At test time, only the base network and the intersection classifier are utilized.

Instead of negative entropy, with similar goal to the negative entropy loss, some researchers [2] utilize confusion loss, i.e., cross entropy loss with respect to the uniform distribution as:

$$\min_{\mathcal{F}} \alpha \left(-\sum_{i=1}^{2} 0.5 \log \hat{z}_i \right). \tag{2}$$

This loss is minimized if the domain predictor predicts 0.5 confidence for each indoor and outdoor class, i.e., it is unable to predict the domain from the extracted features. The loss value and its gradient can be seen in Fig. 2(b). In this work, we focus on comparing Eq. (1) and (2) in our pedestrian-view intersection classification model.

Additionally, following Kim *et al.* [1], we also add an adversarial training for the domain prediction. $\mathcal{F}$ maximizes cross entropy between domain prediction $\hat{\boldsymbol{z}}$ and its ground truth $\boldsymbol{z}$. On the other hand, $\mathcal{H}$ minimizes the loss. During test, only $\hat{\boldsymbol{y}}$ is considered for accuracy measurements.

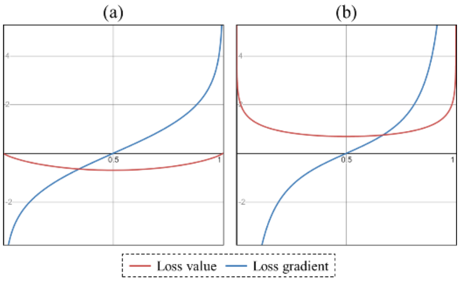

Figure 2. Graphs representing loss values and their gradients with respect to the prediction: (a) $y = x\log x + (1-x)\log(1-x)$ *similar to negative entropy in Eq. (1), (b)* $y = -0.5\log x - 0.5\log(1-x)$ *similar to confusion loss in Eq. (2). In this graph, x-axis is confidence of indoor class* $\hat{z}_1$ *and since the final output is softmax, confidence of outdoor class is* $\hat{z}_2 = 1 - \hat{z}_1 = 1 - x$.

III. Experiments

For indoor dataset, we collected 2,020 and 791 images for non-intersection and intersection categories, respectively. The dataset is sampled from various YouTube videos recorded in multiple scenes. For outdoor dataset, we collected 2,092 and 2,713 images for non-intersection and intersection categories, respectively. The dataset is recorded in urban areas and parks. Each indoor and outdoor dataset is first divided into train, validation, and test sets with ratio of 5:2:3. Then, we combine the train and validation set of both domains as the overall training and validation sets. However, we test our model in each domain separately.

For our model, we utilize ImageNet-pretrained ResNet-18 [5]. The $\mathcal{G}$ and $\mathcal{H}$ branches start from the third residual block. Fine-tuning using our dataset is conducted using Adam, learning rate 10^{-5}, weight decay 10^{-6}, and batch size 64. The highest validation accuracy out of 500 training epochs are selected for testing. We run each experiment five times and report the average.

For training, the input image is first cropped into random size between range of 0.8 to 1.0 from the original size then resized to 224×224. Additionally, we add random horizontal flip, contrast, brightness, sharpness, and color balance augmentation. Finally, we normalize the input values to range 0 to 1. For preprocessing during validation and test, we only resize and normalize the input.

Table 1 shows the accuracy comparisons between models trained using negative entropy loss and models trained using confusion loss on various α settings. With higher values of α, model trained using negative entropy tends to perform better than using confusion loss. This phenomenon can be caused by the gradients. As seen in Fig. 2, confusion loss tends to have steeper gradient than the negative entropy loss which makes it more susceptible with higher α.

Table 1. Test accuracy comparisons between models trained on negative entropy and confusion loss on different weighting α.

	Negative Entropy (Eq. (1))		Confusion Loss (Eq. (2))	
α	Indoor	Outdoor	Indoor	Outdoor
1	**80.07**	**68.62**	69.61	57.16
0.1	**80.67**	**68.90**	73.99	63.73
0.01	**80.92**	**68.08**	80.75	67.64
0.001	**81.03**	67.85	80.82	**68.25**
0.0001	79.93	67.60	**80.78**	**68.37**

III. Conclusion

In this paper, we compare two losses, i.e., negative entropy and confusion loss, utilized in multi-domain setting of pedestrian-view intersection classification. The results show that using a higher weighting factor, negative cross entropy is more stable compared to the confusion loss.

ACKNOWLEDGMENT

This work was supported by the ICT R&D program of MSIT/IITP. [2019-0-01309, Development of AI Technology for Guidance of a Mobile Robot to its Goal with Uncertain Maps in Indoor/Outdoor Environments].

참고문헌

[1] B. Kim, H. Kim, K. Kim, S. Kim, and J. Kim, "Learning not to learn: Training deep neural networks with biased data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9012-9020.

[2] M. Alvi, A. Zisserman, and C. Nellåker, "Turning a blind eye: Explicit removal of biases and variation from deep neural network embeddings," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018, pp. 0-0.

[3] M. Astrid, M. Z. Zaheer, J.-H. Lee, J.-Y. Lee, and S.-I. Lee, "What do pedestrians see?: Visualizing pedestrian-view intersection classification," in *2020 20th International Conference on Control, Automation and Systems (ICCAS)*. IEEE, 2020, pp. 769-773.

[4] M. Astrid, J.-H. Lee, M. Z. Zaheer, J.-Y. Lee, and S.-I. Lee, "For safer navigation: Pedestrian-view intersection classification," in *2020 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2020, pp. 7-10.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.