라벨 스무딩을 활용한 치은염 이진 분류기 캘리브레이션

이상현

전주대학교부설 국제영재아카데미

Calibration for Gingivitis Binary Classifier via Epoch-wise Decaying Label-Smoothing

Sanghyun Lee

Global Prodigy Academy (GPA)

E-mail : sanghyunlfa18@gpa.ac.kr

ABSTRACT

Future healthcare systems will heavily rely on ill-labeled data due to scarcity of the experts who are trained enough to label the data. Considering the contamination of the dataset, it is not desirable to make the neural network being overconfident to the dataset, but rather giving them some margins for the prediction is preferable. In this paper, we propose a novel epoch-wise decaying label-smoothing function to alleviate the model over-confidency, and it outperforms the neural network trained with conventional cross entropy by 6.0%.

키워드

Binary Classifier, Data Augmentation, Label-smoothing, Epoch-wise decaying

Ⅰ. Introduction

Rapid development of technologies, especially from communication engineering and computer science field are followed by advent of new paradigms in many other areas that benefits from such cutting-edge technologies. Healthcare system, one of the most correlated area, is currently getting digitized over time. For instance, wearable devices that measures blood pressure and blood sugar enabled people suffering from diabetes to monitor their condition in real-time manner. Accordingly, new paradigm in healthcare system is slowly rising to the surface; telemedicine.

Still, majority of the area in the medicine are yet to be dealt with telemedicine. Dentistry however, is considered as a candidate to be applied in an initial stage of telemedicine, due to its nature where patients exhibit obvious exterior symptoms in most of the cases. Thus, number of companies are currently delving into developing and presenting the novel architecture in dental telemedicine system.

It is regarded trivial that machine learning is being integrated in telemedicine system. However, in supervised machine learning, the ground-truth label is required. Considering the scarcity of the data-annotators in the field, since these qualified data annotators should be doctors and dentists, it is quite unrealistic to conduct a dataset fully with data that are thoroughly labeled by experts. This indicates the innate contamination of the dataset.

In this paper, we propose epoch-wise decaying label-smoothing to mitigate such contamination and to avoid the model being overconfident on the dataset. We choose gingivitis classification task for this work, as it is very common disease for the modern people and can be determined by examining the gum image. We first conducted a dataset for gingivitis classification by web crawling. To enlarge the number of the data, data augmentations are applied, considering a realistic situation when dental telemedicine system is successfully deployed.

Ⅱ. Data Augmentation

Number of gingivitis images acquired via web crawling are too small. Majority of them are ill-labeled, hence we had a human expert(*a dentist*) to pick some well-labeled images among them. We

Algorithm 1: Data Augmentation

Given: W: Web Crawler;
H: Human Expert, $DataAug$: Data Augmentation
$D \leftarrow W(t)$;
$\mathcal{D} \leftarrow H(D)$;
$\hat{\mathcal{D}} \leftarrow DataAug(\mathcal{D})$;
$\mu, \sigma \leftarrow mean(\hat{\mathcal{D}}), std(\hat{\mathcal{D}})$;
$m = \sigma \times 2$;
$Dataset = []$;
for $i \leftarrow 0$ **to** $len(\hat{\mathcal{D}})$ **do**
  if $\hat{\mathcal{D}}(i) > \mu - m$ **and** $\mu + m > \hat{\mathcal{D}}(i)$;
  $Dataset.append(\hat{\mathcal{D}}(i))$;
return $Dataset$

Algorithm 1. Data Augmentation

were aware that such process is not perfect to fully avoid the contamination since the images are not taken from identical conditions, thus varying illumination and camera positions can even fake the human expert. After human expert picks out the data that are eligible to be used as data to train the neural network, we apply data augmentations.

Functions for data augmentation includes color-jittering, random scaling, flipping, rotating and adding random Gaussian noise. These augmentations will be increase the robustness of the neural network when deployed in the applications. Users who wish to identify if they have gingivitis or not should take the photo of their gum and upload them on the application. Here, the photo, or image from the users will vary in angles, sizes, and even colors due to different lightning. Therefore, we believe these 5 types of data augmentation methods are adequate for this task. Nonetheless, extreme data distortion or data modification can occur in adverse effects. To roughly eliminate the outliers, we calculated the mean and standard deviation of the acquired augmented dataset. Then, we cut-off the marginal values from each side and eventually acquired the dataset that will be used for training and testing the neural network. Systematic process of dataset acquirement is described in Algorithm 1. Dataset is then divided randomly into 8:2 ratio to configure training and testing dataset.

III. Loss Functions

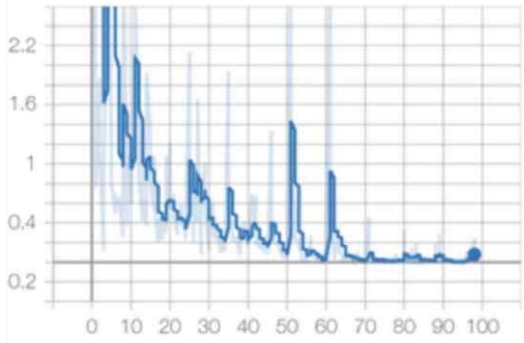

Figure 1. Loss graph for cross-entropy loss

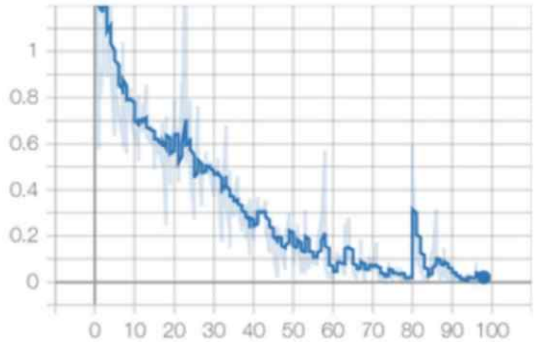

Figure 2. Loss graph for epoch-wise decaying label-smoothing loss

Among the numerous neural network models that are known to be effective in image classification tasks, we had chose ResNet34[1] architecture as backbone network. It is established in the literature that cross-entropy loss is well functioning loss function for classification tasks. Accordingly, we first trained the network with cross-entropy loss. Figure 1 shows the loss graph for testing in case

$$\hat{L} = (L \times (1 - \alpha) + \alpha / c) \times (1 + E(i) / k)$$

Equation 1. Epoch-wise decaying label-smoothing

of cross-entropy loss. Quantitatively, we can observe that high peaks appear during the testing. Equation 1 shows epoch-wise decaying label-smoothing function we had proposed to mitigate the phenomenon where network being overconfident on the dataset. Here, $\hat{L}$ denotes new one-hot label, L denotes original one-hot label, α denotes smoothing factor, c denotes number of classes, $E(i)$ denotes current epoch, and lastly, k denotes decaying factor. Unlike conventional label smoothing function[2] that lacks epoch-wise decaying part in Equation 1, we believe that since the neural network learns the

Figure 3. Validation performances

power to distinguish the true and false gradually, performance can be boosted by adding epoch-wise decaying factor to increase the label value as training progresses.

Figure 2 illustrates the power of epoch-wise decaying label-smoothing loss. It shows greater stability compared to Figure 1 during testing. To compared the power of our epoch-wise decaying label-smoothing function, we evaluated the validation performance of the neural networks that are trained with (1) cross-entropy loss (2) label-smoothing loss (without decaying) with smoothing factor α as 0.01 (3) label-smoothing loss (without decaying) with α as 0.02 (4) epoch-wise decaying smoothing loss (**Ours**) with α as 0.02 and decaying factor k as 100 (5) epoch-wise decaying smoothing loss (**Ours**) with α as 0.02 and decaying factor k as 80. As depicted in Figure 3, Ours shows the most outperforming results compared to others. When compared to neural network trained with cross-entropy loss, Ours achieved 6.0% increase.

Ⅳ. Label-smoothing

Despite the network showing reasonable performance on validation dataset, we need further examination in order to make sure that the neural network did rely on common-sense knowledge to classify the data. Grad-CAM[3] is an well known work in the area of XAI(explainable AI) to visualize the local region of the image where the neural network inferred to be a key pattern of a certain class. As shown in the Figure 4 and 5, the neural network trained with cross-entropy loss failed to locate the bleeding and swollen region, whereas the neural network in Figure 4 successfully capture the such region at the bottom area.

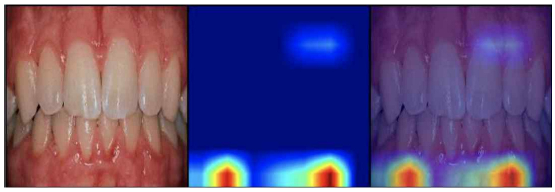

Figure 4. Grad-CAM visualization of neural network trained with epoch-wise decaying label smoothing loss

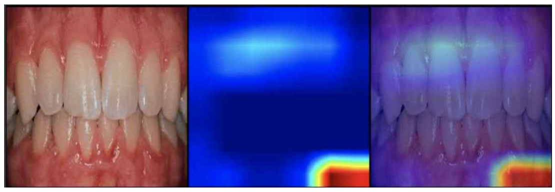

Figure 5. Grad-CAM visualization of neural network trained with cross entropy loss

Ⅴ. Conclusion

In this paper, we propose a novel loss function for gingivitis classification task, coined as epoch-wise decaying label-smoothing function. Our loss function outperforms the neural network trained with conventional cross entropy loss by large margin. Also, we developed a sophisticated strategy to conduct a dataset for the task. Moreover, by visualizing the performance of the neural network, we examined the result more thoroughly. Yet, we have not dealt with every combination of hyper-parameters in our loss function. For further experiments, we are planning to discover a relationship between smoothing factor and decaying factors we have suggested in the loss function.

References

[1] K. He, X. Zhang, S. Ren, J. Sun, “Deep Residual Learning for Image Recognition”, *arxiv 2015*

[2] G. Pereyra, G. Tucker, J. Chorowski, L. Kaiser, G. Hinton, “Regularizing Neural Networks by Penalizing Confident Output Distribution”, *ICLR 2017*

[3] R. R. Selvaraju, M. Cogswell, A. Das, R. Vendantam, D. Parikh, D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based localization”, *ICCV 2017*