УДК 004.896

Алгоритмы уменьшения размерностей моделей для анализа звуковых событий методом дистилляции знаний

Алхаддур А. (Университет ИТМО)
Научный руководитель – доктор технических наук, Матвеев Ю.Н.
(Университет ИТМО)

The goal of this study is to investigate the use of knowledge distillation (KD) to improve the performance of models for sound event detection. The peculiarity of this task is considered in choosing the best way to distill the model in each environment and for each subset of events.

Introduction. Sound event detection is a challenging task which is causally linked to many factors such as the complexity of audio signals, overlapping events, environmental variability, limited labeled data, and the large number of possible sound classes. Traditional approaches to this problem involve training large, complex models that are capable of accurately recognizing a wide range of sound events. While these models can achieve good performance, they often require a large amount of computational resources and may be difficult to deploy in resource-constrained environments. One way to address these challenges is to use knowledge distillation, which is a technique that allows a smaller, more efficient model to be trained to mimic the behavior of a larger, more accurate model. The general idea behind knowledge distillation is to transfer the knowledge learned by the larger model to the smaller model, so that the smaller model can achieve similar or even improved performance while being more computationally efficient.

Main Part. Our study proposes a novel approach to distill sound event detection (SED) models by using environment-specific information to create specialized models. We used knowledge distillation and other training approaches like transfer learning to train a smaller, more efficient SED model. To evaluate the performance of the models, we measured their accuracy and computational efficiency, based on model size and FLOPs. We compared our distilled model with a model trained from scratch and a model distilled without considering the deployment environment. Our findings showed that the distilled model achieved comparable or even better results than the model trained from scratch, while having the same architecture. Our study demonstrates the effectiveness of using environment-aware knowledge distillation to distill specialized models and improve the efficiency and accuracy of SED models. Our approach can be particularly useful in resource-constrained environments where computational efficiency is crucial, and where there are specific sound classes that need to be detected accurately.

Conclusion. We addressed the challenges of training SED models, and demonstrated how knowledge distillation can be used to reduce model size and required computations, and proposed a method for selectively distilling SED models based on a subset of events of interest.

References

1. Gou J. et al. Knowledge distillation: A survey //International Journal of Computer Vision. – 2021. – Т. 129. – С. 1789-1819.
2. Mesaros A. et al. Sound event detection: A tutorial //IEEE Signal Processing Magazine. – 2021. – Т. 38. – №. 5. – С. 67-83.

3. Turpault N. et al. Sound event detection in domestic environments with weakly labeled data and soundscape synthesis. – 2019.

Алхаддур А. (автор) Подпись

Матвеев Ю.Н. (научный руководитель) Подпись