ACTA SCIENTIFIC COMPUTER SCIENCES

Volume 1 Issue 1 November 2019

Editorial

Challenges of Using Deep Learning Methods in Speech Processing and Solutions

Peyman Goli*

Head of the Department of Electrical and Computer Engineering, Khavaran Institute of Higher Education, Mashhad, Iran

***Corresponding Author:** Peyman Goli, Head of the Department of Electrical and Computer Engineering, Khavaran Institute of Higher Education, Mashhad, Iran.

Received: November 05, 2019; **Published:** November 06, 2019

Recently, thanks to the development of computer science in the both software and hardware aspects, speech processing methods have been vastly benefited from deep learning models. Feature engineering is one of the important steps in speech processing methods, so that in the traditional speech processing approaches, choosing and extracting appropriate acoustic features from input speech signals played an important role in the results of the system. In deep learning methods, the useful representations of input signal will be learned by deep layers for the tasks at hand (i.e. classification or regression models), and feature engineering does not need to be manually performed as a complicated preprocessing stage. This important property has significantly improved the performance of deep learning based methods, and made them so favorable in the current speech processing methods.

Deep learning methods have shown a significant performance in speech processing in comparison with the traditional methods. For instance, speech emotion recognition, sound source localization, and speech and speaker recognition using deep classification modes, and speech enhancement and synthesis employing deep regression models have improved the algorithm performance. Various types of deep neural networks (DNNs) have been employed in speech processing methods. For example, deep feedforward layers, convolutional layers, and recurrent neural networks using gated cells (i.e. long short term memories) are DNNs which have been vastly used in speech processing based on their specific applications.

Despite of the great advantages of using deep learning models in speech processing, there are still several major challenges which may limit the usage of deep learning. Providing big data for training deep neural networks is the first challenge in the using of deep learning models, specifically creating labeled data for training deep classification models, which often needs to be supervised by human. Powerful GPUs are the necessary hardware for the training stage, while everyone cannot access them. However, cloud GPUs provided by the international companies, like Google, can be a great alternative. Additionally, there are some problems in the training stage which may negatively affect the training process, and significantly deteriorate the performance of the trained models. Overfitting is the biggest problem in the training stage of deep layers. Generally, all deep neural networks eventually tend to learn the noises and details in the training dataset. This will reduce the system performance on new data. Fortunately, there are several effective solutions to prevent overfitting in the training stage. For example, using big training dataset, determining an appropriate number of hidden layers and neurons, using validation dataset for monitoring training stage, and initializing system parameters in the pre-training stage can overcome overfitting problem. Generalization of the trained models on new conditions is another problem. Since the deep models seriously adapt itself to the training dataset, system generalization may be reduced in various noisy conditions. Providing a relatively complete dataset including different noisy conditions can improve the generalization of trained model on new data. Vanishing gradient is a problem which usually occurs in the gradient based training of some deep neural networks. In this case, the gradient will be vanishingly small, and may prevent the changing of the system parameters during the training stage. Using the gated cells in recurrent neural networks and residual blocks in deep convolutional layers can reduce the negative effects of vanishing gradient problem, in the training stage.

Despite of the big problems in using of deep learning based models, researchers are still so eager to employ deep neural networks in speech processing algorithms. The high performance achieved by deep learning can be superior to the available challenges, and deep learning will dramatically find its way in speech processing methods.

Deep learning has important properties that make it so compatible with speech and signal processing methods. Simplicity, deep learning can automatically perform feature engineering, and replace complex and engineering-heavy preprocessing stage with simple, end-to-end trainable models. Scalability, deep learning is highly amenable to parallelization on system hardware, like GPUs, which make it so compatible with Moore's law. In addition, the capacity of training over small batches of data, allow deep learning models to be trained on datasets of arbitrary size.

Volume 1 Issue 1 November 2019

© All rights are reserved by Peyman Goli.