



## **Deep Learning Program Opening Workshop August 12-16, 2019**

### **SPEAKER TITLES/ABSTRACTS**

**Shih-Kang Chao**  
University of Missouri

“Training DNN with Dynamic SMD”

Stochastic gradient descent (SGD) is a popular algorithm that can handle extremely large data sets due to its low computational cost at each iteration and low memory requirement. However, a major drawback of SGD is that it does not adapt well to the underlying structure of the solution, such as sparsity. Many variations of SGD have been developed based on the concept of stochastic mirror descent (SMD). In this paper, we develop diffusion approximation to the "dynamic" SMD with constant step size, using the local Bregman divergence. The dynamic SMD allows the regularizer of SMD to vary with time. The diffusion approximation results shed light on how to fine-tune an  $l_1$ -norm based SMD algorithm which zeros inactive coefficients without suffering from bias. Numerical analysis with sparse principal component analysis and neural networks will be shown.