Name: Jiangwen Sun Email: jsun@odu.edu

Annotating Genetic Variants by Training Deep Neural Networks to Predict Multiple Epigenomic Events in Brain via Effective Multi-task Learning

Mohammad Shiri, Sanjeeva Dodlapati, Jiangwen Sun

Old Dominion University

Deep neural networks trained for predicting cellular events from DNA sequence have become emerging tools to help elucidate biological mechanisms underlying associations identified in genome-wide association studies. To enhance training, multi-task learning (MTL) has been commonly exploited in previous works where trained networks were needed for multiple profiles differing in either event modality or cell type. All existing works adopted a simple MTL framework where all tasks share a single feature extraction network. Such a strategy leads to substantial negative transfer, reduced performance for a large portion of tasks compared to single-task learning. Existing methods for addressing such negative transfer are generally with very limited scalability. In this study, we developed a highly scalable task grouping framework to address negative transfer by only jointly training tasks that are potentially beneficial to each other. Our approach exploits the network weights associated with task-specific classification heads that can be cheaply obtained by one-time joint training of all tasks. We evaluated the framework with a dataset aggregated from ENCODE and NCBI GEO that included a total of 367 profiles with 86, 90, and 191, respectively for chromatin accessibility, transcription factor binding, and histone modification in varying cells/tissues related to human central nervous system. The results demonstrated the effectiveness of the framework. The performance of most models is considered good with average AUCs by event types ranging from 0.872 to 0.966. This, together with the large variation in their estimated effects among genetic variants using trained models indicated the great potential of our work.