jh190066-DAH

*篠崎隆宏（東京工業大学　工学院　情報通信系）*

# Deep Neural Network Optimization Based on Dual Inheritance Theory and Its Application

## Goal:

An efficient automatic development framework of deep neural network (DNN) is needed that jointly optimizes the network configurations and the parameters. Our goal is to provide such a method by applying the dual inheritance theory in evolutional biology to an engineering black-box optimization method

## Evolution based DNN optimization and it limitations:

We have previously proposed an evolutionary algorithm based automatic development framework of DNN
[Shinozaki+ ICASSP 2015]

- It successfully optimized meta-parameter settings of network structure and training conditions
- The optimization efficiency and the achieved performance were limited

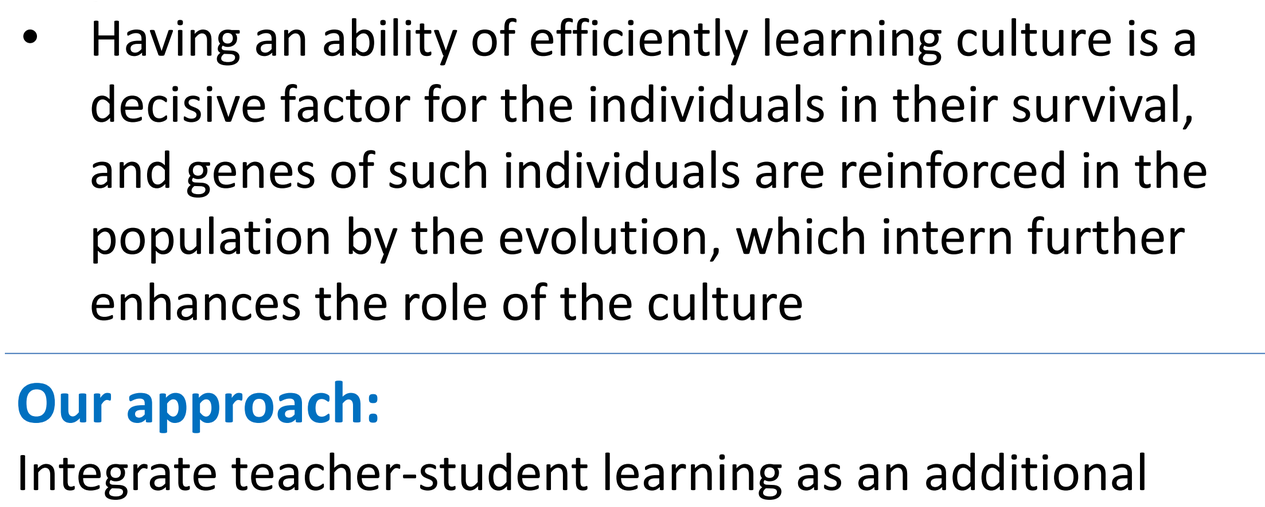
## Dual inheritance theory:

Dual inheritance theory in evolutionary biology explains the outstanding intelligence of human beings as the result of the synergistic effect of the gene-culture interaction

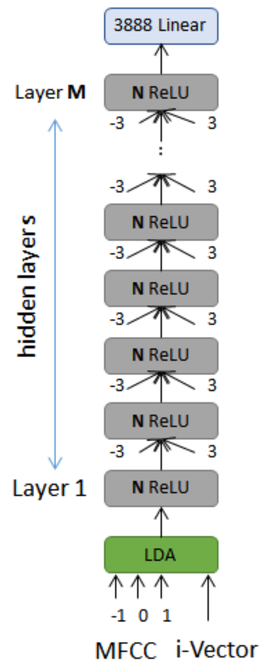
- In human beings, knowledge of ancestors is inherited by descendants as culture, which provides an additional information pass to the gene inheritance process
- Having an ability of efficiently learning culture is a decisive factor for the individuals in their survival, and genes of such individuals are reinforced in the population by the evolution, which intern further enhances the role of the culture

## Our approach:

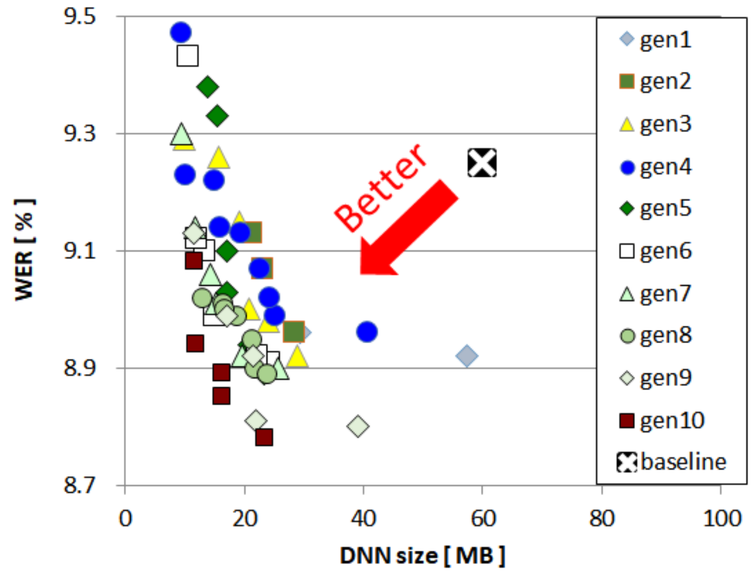
Integrate teacher-student learning as an additional information pass between generations to the evolutionary optimization framework of DNN




Proposed algorithm




(a) Initial structure




(b) Evolution result. Baseline is the initial DNN. Lower DNN size saves memory and computation, and lower word error rate (WER) indicates better performance.

Pareto based multi-objective evolution of DNN speech recognition system

## Analogy between human evolution and DNN development:

- There is a double optimization structure in DNN development, i.e., the design of the network configuration and the estimation of connection weights
- The former corresponds to the design of the brain by a gene, and the latter is parallel to the leaning by an individual

- The objective is to maximize the expected gene performance by optimizing gene distribution parameter $\theta$ as in conventional evolution strategy
- However, the individuals have an additional information pass as the culture in addition to the gene inheritance

$$\hat{\theta} = argmax_\theta \int h(x) P_\theta(x) dx$$

$x$: gene, $h(x)$: performance, $P_\theta(x)$: gene distribution

## Current research status:

Finished initial implementation and started experiments using a speech recognition system as the task