A DNA Computing-Inspired Silicon Chip for Pattern Recognition

Joo-Kyung Kim[1], Byung Soo Kim[2], Oh Hyuk Kwon[2], Seung Kon Hwang[2], Jung-Woo Ha[1], Chan-Hoon Park[1], Duck Jin Chung[2], Chong Ho Lee[2], Jaehyun Park[2], and Byoung-Tak Zhang[1]

[1] Biointelligence Laboratory, School of Computer Science and Engineering, Seoul National University, Seoul 151-744, Korea
[2] School of Information and Communication, Inha University, Incheon 402-751, Korea
E-mail: btzhang@bi.snu.ac.kr

A molecular evolutionary learning model has been proposed in a previous work that uses only the primitive DNA-computing operations, such as hybridization and duplication, to automatically build or "evolve" a pattern classifier from a set of training examples [1]. We have evaluated the feasibility of this probabilistic library model (PLM) by testing its performance on a digit recognition data set of 3760 patterns (10 fold cross validation) of 8×8 image. Software simulations of the PLM model using a version of the molecular algorithm described in [1] achieved a classification accuracy of 93 %, outperforming in its best performance the standard machine learning algorithms, such as multilayer perceptrons and decision trees which obtained 91% and 85% for this data set, respectively.

Motivated by the promising result and by the fact that the simulated molecular algorithm consists of the simple, parallel memory operations (such as, storing, matching, selecting, and copying the binary feature vectors), we implemented the PLM model for pattern recognition in a silicon memory chip utilizing the massive parallelism offered by the hardware. The architecture of the PLM chip consists of four major blocks: the reference data block, the compare block, the accumulation block, and the voter block. The reference block makes the library elements (feature vectors) and masks the learning data patterns by matching against the library elements. The compare unit performs pattern matching of the test data with the library elements, and transfers the matching result to the accumulation block. The accumulation and voter blocks select the best-matched pattern. These processes are performed in a parallel fashion, emulating the DNA hybridization process (though the detailed physico-chemical reactions are not simulated in the current design). The proposed architecture was implemented using Verilog HDL and simulated by Modelsim. The result of detailed timing simulations shows that one full matching is executed in 3.83 microsec using Xilinx Virtex-4LX FPGA [2], which is an improvement of the simulation speed by order of 10^7 (compared with the software simulation taking order of minutes). It is worth emphasizing again that this implementation basically uses memory access technologies such as fetching and counting, and does not need digital signal processors (DSPs) performing precise numerical calculations, which is usually the case for solving pattern recognition problems based on conventional machine learning algorithms (rather than our DNA-based molecular algorithms).

References

1. Zhang, B.-T. and Jang, H.-Y., "A Bayesian algorithm for in vitro molecular evolution of pattern classifiers", *Lecture Notes in Computer Science* (DNA11), 3384:458-467, 2005.
2. Xilinx Corp., *Virtex-4 User Guide*, www.xilinx.com, Jan. 2007.