Distributed classifier based on genetically engineered bacterial cell cultures

Andriy Didovyk\textsuperscript{1}, Oleg I. Kanakov\textsuperscript{2}, Mikhail V. Ivanchenko\textsuperscript{3}, Jeff Hasty\textsuperscript{1,4,5}, Ramón Huerta\textsuperscript{1,∗}, Lev Tsimring\textsuperscript{1,∗}

\textsuperscript{1} BioCircuits Institute, University of California San Diego, La Jolla CA, USA
\textsuperscript{2} Department of Radiophysics, Lobachevsky State University of Nizhniy Novgorod, Nizhniy Novgorod, Russia
\textsuperscript{3} Department for Bioinformatics, Lobachevsky State University of Nizhniy Novgorod, Nizhniy Novgorod, Russia
\textsuperscript{4} Department of Bioengineering, University of California San Diego, La Jolla CA, USA
\textsuperscript{5} Molecular Biology Section, Division of Biological Science, University of California San Diego, La Jolla CA, USA

* E-mail: rhuerta@ucsd.edu, ltsimring@ucsd.edu

Abstract

We describe a conceptual design of a distributed classifier formed by a population of genetically engineered microbial cells. The central idea is to create a complex classifier from a population of weak or simple classifiers. We create a master population of cells with randomized synthetic biosensor circuits that have a broad range of sensitivities towards chemical signals of interest that form the input vectors subject to classification. The randomized sensitivities are achieved by constructing a library of synthetic gene circuits with randomized control sequences (e.g. ribosome-binding sites) in the front element (Fig. 1). The training procedure consists in re-shaping of the master population in such a way that it collectively responds to the “positive” patterns of input signals by producing above-threshold output (e.g. fluorescent signal), and below-threshold output in case of the “negative” patterns. The population re-shaping is achieved by presenting sequential examples and pruning the population using either graded selection/counterselection or by fluorescence-activated cell sorting (FACS). We demonstrate the feasibility of experimental implementation of such system computationally using a realistic model of the synthetic sensing gene circuits (Fig. 2).

Figure 1. A modular genetic circuit proposed for implementing a distributed genetic classifier. Sensing and response functionalities are split into separate modules. In the first module (sensor), an inducible promoter drives the expression of the transcription factor \( U \) in response to the applied signaling molecule \( X \). The response function of the promoter is chosen to be monotonic (see inset). In the second module (reporter), another inducible promoter drives the expression of a reporter (GFP) in response to induction by \( U \). The promoter is activated by intermediate concentrations of \( U \) and inhibited by high concentrations of \( U \). Thus the resulting response function of the entire two-promoter circuit to the concentration of signaling molecule is bell-shaped for the relevant values of the circuit parameters as shown on the inset.
Figure 2. Classification results for the data set drawn from one bimodal and one unimodal distributions. (A) Evolution of the classifier performance for $\gamma = 0.1$ (“hard” learning; blue) and $\gamma = 1$ (“soft” learning; red), population size $N_c = 10^4$ cells. The classifier performance versus cell population size $N_c$ (B) or GFP fluorescence readout noise $\sigma$ (C); $\gamma = 1$ in (B) and (C), $N_c = 10^4$ in (C). The median and interquartile range of the distribution of the classifier performance calculated from $10^3$ different stochastic realizations are shown in (A)–(C), readout noise $\sigma = 1/35$ in (A) and (B). (D)–(I) Evolution of the parameters of the ensemble of cells before and after training – an example trajectory. The parameters used are $\gamma = 1$, $N_c = 10^4$, $\sigma = 1/35$. It illustrates the shift in the distribution of parameters due to the training process of elimination of cells. The distribution of RBS/promoter strengths $m_u$ before training (D) and after 200 training iterations (E). (F) Normalized GFP fluorescence of the ensemble of cells $f(x)$ (blue) after 200 training iterations, log-normal distribution generating positive (green) and negative (red) class examples. (G) Evolution of the classifier performance in this realization. Evolution of $m_u$ distribution (H) and normalized cumulative GFP fluorescence $f(x)$ (I).