

## Стратегии обучения клеточного коллективного классификатора с конкурентной популяционной динамикой и биологически обоснованной моделью отклика

А. А. Сутягин✉, О. И. Канаков

Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского  
Россия, 603950, Н.Новгород, проспект Гагарина, 23

E-mail: ✉alealesutyagin@gmail.com, okanakov@rf.unn.ru

Поступила в редакцию 08.05.2026, принята к публикации 26.06.2026,  
опубликована онлайн 07.07.2026

**Аннотация.** Цель исследования состоит в подтверждении результативности стратегии обучения коллективного классификатора, основанной на популяционной динамике в режиме конкуренции с сосуществованием, управляемой обучающими примерами, применительно к классификатору, отклик которого описывается моделью синтетической геной цепи, а также в сопоставлении стратегий обучения, основанных на режимах конкуренции с единственным победителем и с сосуществованием. **Методы.** Используются модели конкурентной популяционной динамики, ранее предложенные для описания процесса обучения коллективных классификаторов, в форме систем обыкновенных дифференциальных уравнений. Отклик классификатора описывается ранее предложенной моделью синтетической геной цепи. Результативность обучения характеризуется вероятностью правильного ответа обученного классификатора, оцениваемой по ответам на тестовые примеры. Сопоставляются результаты для двух стратегий обучения на примере двух задач классификации при разных способах определения порога классификации (по аналитической оценке и путём подбора) и при изменении параметров классификатора. **Результаты.** Стратегия обучения, основанная на конкуренции с сосуществованием, даёт лучший или аналогичный результат в сравнении с обучением в режиме конкуренции с единственным победителем, в зависимости от условий задачи классификации и процесса обучения. При этом первая стратегия, в отличие от второй, не требует использования внешнего сортировщика клеток и своевременного прерывания процесса обучения и допускает квазиоптимальную аналитическую оценку порога классификации. Это достигается ценой усложнения системы, ограничения на количество типов клеток в составе классификатора и замедления процесса обучения. **Заключение.** Коллективный классификатор, обучаемый в режиме конкуренции с сосуществованием, представляет собой искусственную экологическую систему с конкуренцией, управляемой обучающими примерами. Такая стратегия обучения сложна в плане биологической реализации, однако является шагом на пути к созданию самообучаемых многоклеточных классификаторов.

**Ключевые слова:** Конкуренция, популяционная динамика, классификатор, синтетическая биология.

**Благодарности.** Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации (проект № FSWR-2026-0017).

**Для цитирования:** Сутягин А. А., Канаков О. И. Стратегии обучения клеточного коллективного классификатора с конкурентной популяционной динамикой и биологически обоснованной моделью отклика // Известия вузов. ПНД. 2026. DOI: 10.18500/0869-6632-003225. EDN: CUWGDB

Статья опубликована на условиях Creative Commons Attribution License (CC-BY 4.0).

## Learning strategies for a cellular collective classifier with competitive population dynamics and biologically motivated response model

A. A. Sutyagin<sup>✉</sup>, O. I. Kanakov

National Research Lobachevsky State University of Nizhny Novgorod, Russia  
23, Gagarina pr., Nizhniy Novgorod 603950, Russia

E-mail: ✉alealesutyagin@gmail.com, okanakov@rf.unn.ru

Received 08.05.2026, accepted 26.06.2026, available online 07.07.2026

**Abstract.** *Purpose.* The aim of this study is to verify the effectiveness of the learning strategy for collective classifiers based on population dynamics in the coexistence regime of competition driven by training samples, when applied to a classifier with response described by a model of a synthetic gene circuit, and to compare the learning strategies based on the winner-take-all and coexistence competition regimes. *Methods.* We use the models of competitive population dynamics, which were suggested previously for describing collective classifier training, in the form of ordinary differential equation systems. The classifier response is described by a previously suggested model of a synthetic gene circuit. The efficiency of training is characterized by the trained classifier's correct answer probability, which is estimated by its responses to testing samples. We compare the results for the two learning strategies, using two classification problems as examples, with different ways to determine the classification threshold (namely, by an analytical expression and by optimization), and varying classifier parameters. *Results.* The coexistence competition learning strategy produces a better or similar result as compared to the winner-take-all strategy, dependent upon the specifics of the classification problem and of the learning process. The former strategy, unlike the latter, does not require neither the use of an external cell sorter nor terminating the learning process in proper time, and admits a quasioptimal analytical estimate of the classification threshold. This is achieved at the cost of an increased complexity of the system, imposing a limitation on the quantity of cell types constituting the classifier, and slowing down the learning process. *Conclusion.* A collective classifier trained by competitive dynamics in the coexistence regime is essentially an artificial ecological system with competition driven by training samples. This learning strategy is complicated in terms of biological implementation, but it provides a step towards creating self-learnable multicellular classifiers.

**Keywords:** Competition, population dynamics, classifier, synthetic biology.

**Acknowledgements.** This work was supported by the Ministry of Science and Higher Education of the Russian Federation (project No. FSWR-2026-0017).

**For citation:** Sutyagin AA, Kanakov OI. Learning strategies for a cellular collective classifier with competitive population dynamics and biologically motivated response model. *Izvestiya VUZ. Applied Nonlinear Dynamics.* 2026. DOI: 10.18500/0869-6632-003225

*This is an open access article distributed under the terms of Creative Commons Attribution License (CC-BY 4.0).*

### Введение

Концепция коллективного классификатора<sup>1</sup> была предложена в [1] и развивалась в последующих работах [2–4]. Основная идея состоит в создании обучаемой системы, представляющей собой ансамбль достаточно простых элементов разных типов с постоянными параметрами, каждый из которых обеспечивает некоторую нелинейную функцию отклика на один или несколько входных сигналов (признаков). Элементы каждого типа идентичны между собой, но вид функции отклика различается между элементами разных типов. Отклик ансамбля определяется как сумма откликов всех его элементов. Неидентичность

<sup>1</sup>В англоязычной литературе в этом значении используется термин *distributed classifier* — буквальный перевод «распределённый классификатор».

элементов даёт возможность изменять суммарную функцию отклика ансамбля путём изменения его состава (количества элементов каждого типа) без изменения параметров самих элементов. Путём сопоставления суммарного отклика с некоторым пороговым значением, такой ансамбль может быть использован для решения задачи двухклассовой классификации (отнесения классифицируемого объекта к одному из двух классов на основании одного или нескольких наблюдаемых признаков [5, 6]), чем и обусловлено наименование «коллективный классификатор» для такого ансамбля [1].

Решающим правилом классификатора называют функцию, отображающую множество значений входных данных (признаков) на множество классов. Если известны условные распределения вероятностей на множестве значений признаков по условию принадлежности классифицируемого объекта к каждому из классов, то минимальная вероятность ошибки классификации обеспечивается решающим правилом Байеса [5]. Если же эти распределения неизвестны, но дано множество «обучающих примеров» (пар, каждая из которых состоит из значения признаков объекта и «метки класса», указывающей на априори известную принадлежность объекта к одному из классов), то процесс подбора параметров классификатора для максимизации его эффективности (в каком-либо смысле — например, вероятности правильного ответа) на основании обучающих примеров называют обучением классификатора. В качестве параметров коллективного классификатора, подбираемых в процессе обучения, могут выступать численности типов клеток, входящих в его состав (без подстройки внутренних параметров элементов), что минимизирует сложность отдельного элемента и даёт возможность реализации такого классификатора средствами синтетической биологии в виде популяции живых клеток, снабжённых синтетическими генными структурами [1]; везде далее под «клеткой» понимается элемент ансамбля.

Интерес к исследованию коллективных классификаторов мотивируется возможными приложениями в области создания интеллектуальных биосенсоров, а также тем обстоятельством, что принцип коллективного классификатора во многом заимствует механизмы обучения и функционирования адаптивных иммунных систем у позвоночных (см. обзоры литературы в [1, 3]); таким образом, биологический коллективный классификатор можно рассматривать как обучаемый синтетический (искусственный) аналог иммунной системы, способный к классификации условий среды, что, в свою очередь, создаёт возможность для адаптивной реакции на изменения этих условий. Так, в работе [7] был разработан внутриклеточный коллективный классификатор, элементами которого являются синтетические генные цепи, для распознавания ракового состояния клетки.

Математически, коллективный классификатор полностью определяется способом задания функций отклика отдельных элементов ансамбля и «стратегией обучения» — способом подстройки состава ансамбля при обучении классификатора. В работах [1–3] были рассмотрены различные стратегии обучения, основанные на селекции элементов с помощью некоторого сортировщика, являющегося внешним по отношению к ансамблю<sup>2</sup>. Простейшая «жёсткая» стратегия [2] подразумевает безусловное удаление элемента из ансамбля, если он даёт ответ, противоречащий обучающему примеру. Такая стратегия подходит только для классов, не имеющих пересечений в пространстве признаков. В свою очередь, «мягкая» стратегия [1, 2] основана на вероятностном удалении элемента (с возможным замещением его новым элементом для сохранения полной численности) в зависимости от его ответа на обучающий пример и подходит для классов, описываемых перекрывающимися вероятностными распределениями. Этот исходный вариант мягкой стратегии обучения порождает конкурентную популяционную динамику численностей типов клеток в режиме конкурен-

---

<sup>2</sup>Если элементом ансамбля является живая клетка, то для этой цели может использоваться техника флуоресцентной сортировки (англ. fluorescence activated cell sorting, FACS) [1].

ции с единственным победителем [1]. Это означает, что асимптотический результат такой стратегии всегда представляет собой ансамбль, состоящий из клеток единственного типа, что сильно ограничивает возможности обучения и противоречит самой концепции коллективного классификатора. Таким образом, за исключением простейших случаев, для достижения оптимального результата требуется прерывание процесса обучения в некоторый момент, не известный заранее [1]. Модификация мягкой стратегии обучения, предложенная в [3], решает эту проблему за счёт перехода от конкуренции с единственным победителем к конкуренции с сосуществованием. Показано [3], что асимптотическим результатом такой модифицированной стратегии при определённых допущениях является обученный классификатор, решающее правило которого в определённом смысле аппроксимирует (оптимальное) правило Байеса. Это продемонстрировано, в том числе, на конкретном примере задачи классификации, где один из классов характеризуется бимодальным вероятностным распределением на пространстве признаков, а другой — унимодальным. В работе [4] был рассмотрен вариант стратегии обучения в режиме конкуренции с сосуществованием, не требующий внешнего сортирующего устройства и обеспечиваемый собственной популяционной динамикой ансамбля ценой усложнения элемента (введением внутренней динамики элемента в форме «подсистемы обучения», влияющей требуемым образом на конкурентную популяционную динамику в ансамбле).

В работах [1,2] были предложены варианты реализации элемента коллективного классификатора в виде живой клетки, требуемый отклик которой на внешний сигнал обеспечивается синтетической генной цепью — искусственно синтезированным фрагментом ДНК кода, который встраивается в клетку средствами синтетической биологии (обычно в виде плазмид — кольцеобразных молекул ДНК) и реализует требуемую функциональность клетки поверх её собственной генной системы, продолжающей нормально функционировать и поддерживающей жизнеспособность клетки. При этом в качестве признака (входа классификатора) может выступать любой параметр среды, воспринимаемый сигнальной системой клетки (концентрации химических веществ, температура, освещённость и т. п.); в качестве выхода классификатора для определённости рассматривалась концентрация зелёного флуоресцентного белка GFP (англ. green fluorescent protein), широко используемого в экспериментах в качестве «репортёра» (выходного сигнала, легко наблюдаемого и измеряемого экспериментатором). В математических моделях классификатора [1, 2] функция отклика клетки задавалась биологически обоснованным образом, исходя из модели предложенной генной цепи.

С другой стороны, модифицированные варианты стратегии обучения, основанные на конкуренции с сосуществованием, были исследованы в работах [3, 4] только для случая функций отклика, составляющих базис линейной интерполяции (сетку эквидистантных треугольных пиков), что удобно для анализа и оправданно при абстрактной постановке задачи, но не соответствует биологической реализации классификатора. Проблема реализации конкурентной стратегии обучения в режиме сосуществования с учётом биологически обоснованного вида функций отклика до сих пор не исследовалась.

Целью настоящей работы является восполнение этого пробела: исследование коллективного классификатора с подсистемой обучения [4] при замене треугольных функций отклика на биологически обоснованные колоколообразные функции [1]. Тем самым проверяется применимость механизма обучения [4] к реалистичным биологическим функциям отклика. Кроме того, использование одинаковых функций отклика клеток даёт возможность прямого сопоставления стратегий обучения [1] и [4]; для этого в примерах задач классификации рассматриваются классы, описываемые унимодальными и бимодальными распределениями, аналогичные рассмотренным в [1].

## 1. Методика

**1.1. Решающее правило коллективного классификатора.** Следуя обозначениям работ [3, 4], рассмотрим ансамбль из  $s$  типов клеток, численности которых полагаются достаточно большими, чтобы описывать их непрерывными переменными  $n_i \geq 0$ ,  $i = 1 \dots s$ . Каждый тип клеток характеризуется функцией отклика  $f_i(x) \geq 0$ , описывающей отклик любой клетки  $i$ -го типа на входное значение признака  $x$ . Суммарный отклик  $F(x)$  всех клеток, составляющих коллективный классификатор, и соответствующее решающее правило  $C(x)$  тогда задаются выражениями [4]

$$F(x) = \sum_{i=1}^s n_i f_i(x), \quad C(x) = \begin{cases} +, & F(x) > \Theta, \\ -, & F(x) \leq \Theta, \end{cases} \quad (1)$$

где символы  $+$  и  $-$  обозначают классы (соответственно именуемые «положительным» и «отрицательным»), а  $\Theta$  — порог классификации, определяемый по окончании обучения.

**1.2. Модель конкуренции с подсистемой обучения.** Как показано в [4], необходимая для обучения классификатора подстройка численностей типов клеток  $n_i$  может быть обеспечена конкурентной популяционной динамикой, управляемой последовательностью обучающих примеров и описываемой системой обыкновенных дифференциальных уравнений

$$\dot{n}_i = n_i \left( 1 - \frac{1}{N_c} \sum_{j=1}^s n_j \right) - k_i \frac{n_i^2}{N_c}, \quad (2a)$$

$$\dot{k}_i = \begin{cases} -\tau_{\text{off}}^{-1} k_i, & \text{вне процесса обучения,} \\ \tau_l^{-1} f_i(x_m) (k^+ - k_i), & \text{если } c_m = +, \\ \tau_l^{-1} f_i(x_m) (k^- - k_i), & \text{если } c_m = -, \end{cases} \quad i = 1 \dots s. \quad (2b)$$

В уравнении популяционной динамики (2a) параметр  $N_c > 0$  влияет на суммарную равновесную численность ансамбля и потому именуется «ёмкостью среды»<sup>3</sup>. Обозначения, входящие во второе уравнение (2b), определены ниже. Слагаемое в правой части уравнения (2a), содержащее  $k_i$ , описывает «внутривидовую» (по экологической терминологии) конкуренцию между клетками каждого типа, интенсивность которой определяется величинами  $k_i$  для каждого типа клеток в отдельности. Эта конкуренция не зависит от клеток других типов в ансамбле. Часть уравнения (2a), которая остаётся при  $k_i = 0$ , описывает конкуренцию «за общий ресурс», в которой все клетки участвуют одинаковым образом, независимо от их типа<sup>4</sup>. Время в модели безразмерно и измеряется в единицах характерного времени популяционной динамики (которое, таким образом, в (2a) принято за единицу для сокращения числа параметров без потери общности).

Сначала рассмотрим популяционную динамику (2a) как автономную, при постоянных  $k_i > 0$ . Тогда подсистема (2a) имеет состояние равновесия  $n_i = n_i^* > 0$ , полностью определяемое значениями  $k_i$ , расположенное в положительном ортанте фазового пространства и

<sup>3</sup>Учитывая, что изменение  $N_c$  эквивалентно перемасштабированию всех  $n_i$ , а для классификатора имеют значение лишь соотношения между численностями  $n_i$ , значение  $N_c$  не важно — можно без потери общности положить, например,  $N_c = 1$ .

<sup>4</sup>Клетки каждого типа участвуют в обоих видах конкуренции, поэтому правая часть  $i$ -го уравнения (2a) содержит два слагаемых, пропорциональных  $n_i^2$ .

глобально устойчивое в нём [4]:

$$n_i^* = hN_c k_i^{-1}, \quad h = \left(1 + \sum_{l=1}^s k_l^{-1}\right)^{-1}. \quad (3)$$

В случае, когда все  $k_i = 0$ , динамика (2a) переходит в вырожденный режим конкуренции с одинаковой конкурентоспособностью всех типов клеток, при котором система имеет устойчивое инвариантное многообразие  $\sum_i n_i = N_c$ , целиком состоящее из состояний равновесия; вне этого многообразия динамика численностей  $n_i$  сохраняет их относительные пропорции [4]. Поскольку именно относительные пропорции численностей типов клеток определяют ответ коллективного классификатора, динамика системы в вырожденном режиме конкуренции не влияет на результат обучения. Тем не менее, для полноты отметим, что в случае, если до обращения в нуль всех  $k_i$  система находилась в состоянии равновесия (3), то в качестве асимптотического состояния  $n_i = n_i^{**}$ , к которому система приходит в результате вырожденного режима конкуренции, имеем

$$n_i^{**} = \frac{N_c}{\sum_l k_l^{-1}} k_i^{-1}, \quad \sum_{i=1}^s n_i^{**} = N_c, \quad (4)$$

где значения  $k_i$  те же, что в (3) (то есть до их обнуления). Заметим также, что принадлежность состояния равновесия (4) к устойчивому инвариантному многообразию, целиком состоящему из состояний равновесия, означает, что эти состояния являются устойчивыми по Ляпунову, но не являются асимптотически устойчивыми.

Простая функциональная связь (3) или (4) между установившимся состоянием популяционной динамики и величинами  $k_i$  обеспечивает возможность целенаправленного формирования установившегося состава численностей типов клеток в ансамбле в процессе обучения классификатора через посредство коэффициентов внутривидовой конкуренции  $k_i$  (которые в полной системе (2a,b) являются переменными).

В свою очередь, «подсистема обучения» (2b) устанавливает зависимость динамики переменных  $k_i$  от последовательности обучающих примеров. В уравнении (2b)  $\tau_l$  — масштаб времени процесса обучения (в единицах характерного времени популяционной динамики, см. выше комментарий к уравнению (2a)),  $\tau_{\text{off}}$  — масштаб времени выхода из процесса обучения,  $k^+$  и  $k^-$  — константы. В качестве исходных данных для обучения выступает обучающая последовательность  $\{\pi_m\}_{m=1}^{N_{\text{learn}}}$ , состоящая из примеров  $\pi_m = (x_m, c_m)$ , где  $x_m$  — значение признака,  $c_m \in \{+, -\}$  — метка класса. Обучающие примеры предъявляются системе последовательно. Для этого время в процессе обучения разбивается на равные интервалы с длительностью  $T_s$ , в течение каждого из которых «предъявляется» один очередной обучающий пример, а полное время обучения составляет  $N_{\text{learn}}T_s$ . Значения параметров  $x_m$  и  $c_m$  в правой части (2b) в каждый момент времени в процессе обучения берутся из соответствующего примера  $\pi_m$ , предъявляемого на текущем интервале времени; по окончании предъявления всех обучающих примеров в уравнении (2b) используется вариант правой части «вне процесса обучения». Таким образом, правая часть уравнения (2b) зависит от времени разрывным образом. Если рассматривать последовательность обучающих примеров как случайную, то система (2a,b) в процессе обучения является стохастической, а её решение (как  $k_i(t)$ , так и  $n_i(t)$ ) представляет собой случайный процесс.

Уравнение (2b) линейно по  $k_i$ , поэтому на каждом интервале времени предъявления очередного ( $m$ -го) обучающего примера решение для переменной  $k_i(t)$  экспоненциально стремится к значению  $k^+$  или  $k^-$  (в зависимости от метки класса  $c_m$  текущего примера),

причём скорость стремления (коэффициент в экспоненте по времени) пропорциональна отклику  $f_i(x_m)$ ; если  $f_i(x_m) = 0$ , то значение  $k_i$  во время предъявления данного примера не изменяется. По завершении процесса обучения все  $k_i$  экспоненциально устремляются к нулю на временном масштабе  $\tau_{\text{off}}$ , что переводит популяционную динамику в вырожденный режим конкуренции, как описано выше. Обращение переменных  $k_i$  в нуль происходит не мгновенно, однако мы полагаем параметр  $\tau_{\text{off}} \ll 1$  малым в сравнении с временным масштабом вырожденной конкуренции, в силу чего далее пользуемся формулой (4), полученной в предположении мгновенного обнуления величин  $k_i$  по окончании обучения.

Смысл подсистемы обучения состоит в том, чтобы значение  $k_i$  для каждого типа клеток определялось обучающими примерами, относящимися именно к этому типу (то есть дающими существенно ненулевой отклик  $f_i(x_m)$ ), а также в эффективном «усреднении» таких обучающих примеров на времени порядка  $\tau_l$ . Это усреднение подразумевает, что в случае равного (в среднем) количества дающих ненулевой отклик  $f_i(x_m)$  обучающих примеров из обоих классов, математическое ожидание процесса  $k_i(t)$  должно в силу симметрии сходиться в процессе обучения к величине  $(k^+ + k^-)/2$ , а при преобладании примеров из положительного (отрицательного) класса должно смещаться в направлении  $k^+$  ( $k^-$ ).

Следуя [3, 4], рассмотрим случай, когда функции отклика клеток  $f_i(x)$  являются «узкими» в сравнении с  $w_+(x)$  и  $w_-(x)$  — функциями условной плотности вероятности признака  $x$  по условию принадлежности объекта к классу «+» и «-», — в том смысле, что для них выполняется приближённое соотношение<sup>5</sup>

$$\frac{\int g(x)f_i(x)dx}{\int f_i(x)dx} \approx g(x_i), \quad (5)$$

где  $g(x) = w_+(x)$  или  $g(x) = w_-(x)$ ;  $x_i$  — некоторое «центральное значение», которое приписывается соотношением (5) каждой функции отклика  $f_i(x)$  [4]. Для определённости будем считать, что обучающие примеры выбираются равновероятно из обоих классов. Тогда математические ожидания процессов  $k_i(t)$  со временем сходятся к стационарным значениям  $\hat{k}_i$ , для которых известна оценка [4]

$$\hat{k}_i \approx \frac{k^+ + k^-}{2} + \frac{k^+ - k^-}{2} \cdot \frac{w_+(x_i) - w_-(x_i)}{w_+(x_i) + w_-(x_i)}, \quad (6)$$

где в качестве частных случаев имеем  $\hat{k}_i \approx (k^+ + k^-)/2$  при  $w_+(x_i) = w_-(x_i)$ ;  $\hat{k}_i \approx k^+$  при  $w_+(x_i) \gg w_-(x_i)$ ;  $\hat{k}_i \approx k^-$  при  $w_+(x_i) \ll w_-(x_i)$  — как и обсуждалось в предыдущем абзаце, причём если  $k^+ < k^-$  (что далее предполагается), то всегда  $\hat{k}_i \in [k^+, k^-]$ .

При выполнении условия  $T_s \ll \tau_l$  дисперсия флуктуаций переменных  $k_i(t)$  сходится к малой величине порядка  $T_s/\tau_l$  [4]. В этом пределе случайные процессы, являющиеся решением (стохастической) системы (2а,б), можно приближённо заменить их математическими ожиданиями и рассматривать популяционную динамику как детерминированную. Тогда установившиеся значения численностей типов клеток  $n_i = n_i^{**}$  в асимптотическом состоянии после обучения даются выражением (4), куда вместо  $k_i$  подставляется оценка установившегося математического ожидания  $k_i = \hat{k}_i$  из (6).

Если функции отклика удовлетворяют соотношению

$$f_i(x_j) = A \delta_{ij}, \quad (7)$$

<sup>5</sup> являющееся приближённым вариантом формулировки «фильтрующего свойства» дельта-функции и становящееся точным, если  $f_i(x) = C\delta(x - x_i)$ , где  $C$  — некоторая константа

где  $A$  — некоторая константа,  $\delta_{ij}$  — символ Кронекера, то для суммарного отклика обученного классификатора (1) имеем

$$F(x_i) = An_i^{**} = A \frac{N_c}{\sum_l \hat{k}_l^{-1}} \hat{k}_i^{-1}. \quad (8)$$

Байесовское решающее правило (в предположении равной априорной вероятности классов, что везде далее подразумевается) состоит в том, что на основании значения признака  $x$  классификатор должен выбрать класс «+», если  $w_+(x) > w_-(x)$ , и класс «−» в противном случае. Области решений байесовского классификатора в пространстве признака разделяются корнем (корнями) уравнения  $w_+(x) = w_-(x)$ . Полагая количество типов клеток достаточно большим, а сетку значений  $\{x_i\}$  достаточно плотной, рассмотрим ближайшее к этому корню (к одному из корней) значение  $x_{i^*} \in \{x_i\}$ , для которого тогда приблизительно выполняется  $w_+(x_{i^*}) \approx w_-(x_{i^*})$ . Тогда из (6) имеем  $\hat{k}_{i^*} \approx (k^+ + k^-)/2$ , а из (8)

$$\Theta \approx F(x_{i^*}) \approx A \frac{N_c}{\sum_l \hat{k}_l^{-1}} \frac{2}{k^+ + k^-}, \quad (9)$$

где первый знак приближённого равенства отражает тот факт, что величина  $F(x_{i^*})$  по построению представляет собой значение функции отклика коллективного классификатора вблизи границы, разделяющей области решений байесовского классификатора в пространстве признаков, то есть, иными словами, оценку оптимального порога классификации  $\Theta$  в (1). Для упрощения выражения (9) дополнительно предположим, что хотя бы для одного типа клеток  $i^{**}$  выполняется условие  $w_+(x_{i^{**}}) \ll w_-(x_{i^{**}})$ , тогда из (6) имеем  $\hat{k}_{i^{**}} \approx k^- \approx \max_l \{\hat{k}_l\}$ , а из (8)

$$F(x_{i^{**}}) \approx A \frac{N_c}{\sum_l \hat{k}_l^{-1}} \frac{1}{k^-} \approx \min_l F(x_l). \quad (10)$$

Вводя ещё одно предположение  $k^+ \ll k^-$  и сравнивая между собой выражения (9) и (10), получаем оценку оптимального значения порога [4]

$$\Theta = 2 \min_l F(x_l). \quad (11)$$

В рамках сделанных предположений, решающее правило обученного классификатора (1) с порогом (11) аппроксимирует байесовское решающее правило — в том смысле, что предельное значение суммарной функции отклика классификатора  $F(x)$  при  $x \in \{x_i\}$  обращает решающее правило (1) в байесовское, а ответ классификатора для промежуточных значений признака  $x \notin \{x_i\}$  определяется конкретным видом функций отклика клеток  $f_i(x)$ .

**1.3. Функции отклика клеток.** В работе [4] была продемонстрирована применимость описанного механизма для обучения коллективного классификатора на примере набора функций отклика, заданных в виде базиса линейной интерполяции на равномерной сетке

$$f_i(x) = \max \left\{ 1 - \left| \frac{x - x_i}{\Delta} \right|, 0 \right\}, \quad x_i = x_1 + (i - 1)\Delta, \quad (12)$$

чем обеспечивается выполнение свойства (7), а значения суммарной функции отклика  $F(x)$  при  $x \notin \{x_i\}$  представляют собой линейную интерполяцию значений в узлах сетки  $F(x_i) = n_i$ . Такой выбор математически удобен, однако не отвечает биологической реализации элемента классификатора в виде клетки с синтетической геной структурой.

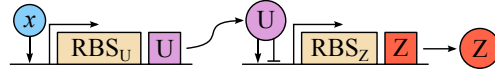


Рис. 1. Схема генной сети, реализующей колоколообразный отклик клетки:  $x$  — входной химический сигнал;  $\text{RBS}_U$  — сайт связывания рибосомы сенсорного гена;  $U$  — промежуточный транскрипционный регулятор;  $\text{RBS}_Z$  — сайт связывания рибосомы репортерного гена;  $Z$  — выходной сигнал.

Fig. 1 Scheme of the genetic circuit implementing a bell-shaped cell response:  $x$  is the input chemical signal;  $\text{RBS}_U$  is the ribosome binding site of the sensor gene;  $U$  is the intermediate transcriptional regulator;  $\text{RBS}_Z$  is the ribosome binding site of the reporter gene;  $Z$  is the output signal.

В настоящей работе функция отклика (12) заменяется на биологически обоснованную, соответствующую генетической схеме из [1], представленной на рис. 1. В этой схеме сенсорный промотор, управляющий синтезом белка  $U$  (промежуточного транскрипционного регулятора), монотонно активируется внешним химическим сигналом  $x$  (который является входом классификатора); синтез белка  $Z$  (выходного сигнала классификатора, например GFP) управляется регулятором  $U$  немонотонным образом: активируется при промежуточных концентрациях регулятора  $U$  и репрессируется (подавляется) при высоких, что достигается посредством соответствующего промотора (в качестве которого может выступать, например, промотор  $P_{RM}$  фага лямбда [1]).

Динамика концентраций белков, обеспечиваемая такой генной сетью, описывается уравнениями [1]

$$\dot{u} = r_u(x) - \mu_u u, \quad \dot{z} = r_g(u) - \mu_g z, \quad (13)$$

где  $u$  и  $z$  — концентрации регулятора  $U$  и «выходного» белка  $Z$ ,  $\mu_u$  и  $\mu_g$  — скорости их деградации, а скорости синтеза  $r_u(x)$ ,  $r_g(u)$  задаются функциями Хилла [1]

$$r_u(x) = m_u \cdot \frac{\alpha A_u^{p_u} + x^{p_u}}{A_u^{p_u} + x^{p_u}}, \quad r_g(u) = m_g \cdot \frac{A_g^{p_g} u^{p_g}}{(A_g^{p_g} + u^{p_g})^2}. \quad (14)$$

Здесь параметр  $\alpha$  определяет «фоновую» экспрессию сенсорного промотора в отсутствие активации входным сигналом;  $A_u$ ,  $A_g$  — константы диссоциации;  $p_u$ ,  $p_g$  — коэффициенты Хилла; параметры  $m_u$ ,  $m_g$  определяют максимальную силу экспрессии (скорость синтеза) соответствующих белков. Значение переменной  $z$  в состоянии равновесия системы (13) как функция входного сигнала  $x$  определяется выражением

$$z^*(x; m_u) = \frac{r_g(r_u(x; m_u)/\mu_u)}{\mu_g}. \quad (15)$$

Функция  $z^*(x)$  имеет колоколообразную форму при  $m_u/\mu_u \in (A_g, A_g/\alpha)$ , а положение её максимума определяется значением отношения  $m_u/\mu_u$  [1]. Это обстоятельство может быть использовано для создания нужного количества типов клеток, различающихся положением максимума функции отклика. Для этого значение параметра  $m_u = m_u^{(i)}$  задаётся разным для разных типов клеток. Такое различие между типами клеток может быть обеспечено, например, за счёт использования разных сайтов связывания рибосомы  $\text{RBS}_U$  в генной сети клеток разных типов [1]. Максимальный разброс положений максимума функции отклика достигается при кратности разброса параметра  $m_u$  в  $1/\alpha$  раз, что составляет порядка

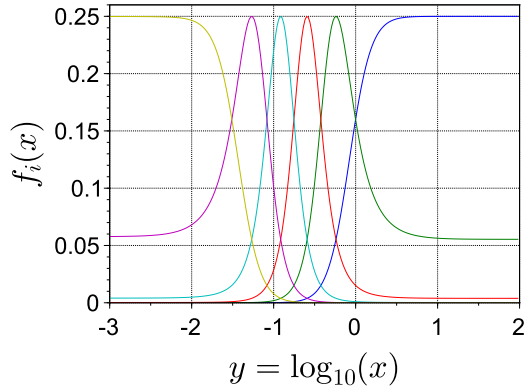


Рис. 2. Функции отклика клеток  $f_i(x)$  (16) для  $s = 6$  типов клеток. Каждая кривая соответствует одному типу клеток. Значения параметров указаны в разд. 2.1.

Fig. 2 Cell response functions  $f_i(x)$  (16) for  $s = 6$  cell types. Each curve corresponds to one cell type. Parameter values are indicated in Sec. 2.1.

$10^3$  при типичных значениях параметра  $\alpha \approx 10^{-3}$ . С учётом вышесказанного, функция отклика клеток  $i$ -го типа задаётся в виде

$$f_i(x) = z^*(x; m_u^{(i)}) = \frac{r_g(r_u(x; m_u^{(i)})/\mu_u)}{\mu_g}, \quad (16)$$

где значения параметров  $m_u^{(i)}$  должны выбираться таким образом, чтобы функции отклика покрывали требуемую область значений признака  $x$  без чрезмерных просветов и перекрытий, подобно тому, как это обеспечивается в случае треугольного базиса (12) (количественное исследование влияния набора функций отклика на результат классификации описывается ниже в разделе 2.2). Заметим, что ширина колоколообразных функций отклика клеток разных типов при варьировании параметра  $m_u$  получается приблизительно одинаковой в логарифмическом, а не в линейном масштабе признака  $x$  (см. рис. 2 или [1, Fig. 2]), поэтому при постановке задач классификации и задании вероятностных распределений классов дополнительно вводится логарифмический признак

$$y = \log_{10} x, \quad (17)$$

а значения параметров  $m_u^{(i)}$  задаются так, что максимумы функций отклика образуют приближённо эквидистантную сетку по  $y$  (см. рис. 2).

В отличие от треугольных функций отклика (12), функции (16) не удовлетворяют условию (7) строго, однако можно говорить о выполнении этого условия в приближённом смысле, чем и обосновывается применение результатов [4], изложенных в разделе 1.2, для приближённого описания классификатора с функциями отклика (16).

**1.4. Задачи классификации, обучение и тестирование классификатора.** Рассмотрим два примера задач классификации, полностью аналогичные рассмотренным в [1]. В первом примере (называемом далее «униmodalной задачей») оба класса описываются логнормальными распределениями признака  $x$ , что соответствует нормальному (гауссову) распределению логарифмического признака  $y$ . Во втором (называемом далее «бимodalной задачей») положительный класс описывается бимodalным распределением, получаемым

суперпозицией двух нормальных распределений признака  $y$ , а отрицательный класс описывается нормальным распределением признака  $y$ , максимум которого расположен между максимумами бимодального распределения положительного класса. Функции условной плотности вероятности логарифмического признака для обеих задач описываются выражениями

$$w_+(y) = \frac{1}{2\sqrt{2\pi}} \left( \frac{1}{\sigma_1} e^{-\frac{(y-\mu_1)^2}{2\sigma_1^2}} + \frac{1}{\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}} \right), \quad w_-(y) = \frac{1}{\sigma_3\sqrt{2\pi}} e^{-\frac{(y-\mu_3)^2}{2\sigma_3^2}}, \quad (18)$$

где в случае унимодальной задачи полагается  $\mu_1 = \mu_2 = -0.55$ ,  $\mu_3 = -2.05$ ,  $\sigma_1 = \sigma_2 = \sigma_3 = 0.5$ , что сводит  $w_+(y)$  к унимодальному нормальному распределению с центром, отличным от центра  $w_-(y)$ ; в случае бимодальной задачи  $\mu_1 = -1.3$ ,  $\mu_2 = -0.3$ ,  $\mu_3 = -0.8$ ,  $\sigma_1 = \sigma_2 = \sigma_3 = 0.14$ .

Обучение классификатора моделируется, как описано в [4]. Для генерации обучающего примера  $\pi_m = (x_m, c_m)$  случайно и равновероятно выбирается один из классов  $c_m$ ; в соответствии с выбранным классом, случайно выбирается значение  $y$  из распределения (18) и пересчитывается согласно (17) в значение признака  $x_m$ . Уравнения системы (2a,b) интегрируются численно. Начальные условия до обучения  $n_i(0) = 1/s$ ,  $k_i(0) = 1$ .

Вероятность правильного ответа  $P$  оценивается с использованием ансамбля тестовых примеров, сгенерированных заранее по тому же алгоритму, что и обучающие примеры. Количество тестовых примеров для каждого класса составляет  $M_{\text{perf}} = 1000$ . Для этого по формуле (11) определяется значение порога классификации  $\Theta$ ; для каждого тестового примера вычисляется ответ классификатора (1); величина  $P$  определяется как доля примеров, на которые получен правильный ответ, от общего числа примеров. Кроме того, решается задача максимизации  $P$  путём подбора оптимального значения порога  $\Theta$ ; найденное максимальное значение  $P$  обозначается  $\hat{P}$ .

Для сравнения, моделируется стратегия обучения, предложенная в [1]. Эта стратегия была представлена в виде алгоритма селекции клеток [1, Fig. 4], однако было также показано (в предположении достаточной «мягкости» обучения, то есть малой вероятности удаления клетки на каждом шаге<sup>6</sup>), что она может быть с хорошей точностью описана уравнениями популяционной динамики

$$\dot{n}_i = n_i \left( \lambda_i - \frac{1}{\sum_{l=1}^s n_l} \sum_{j=1}^s \lambda_j n_j \right), \quad i = 1 \dots s, \quad (19)$$

где параметры конкурентоспособности  $\lambda_i$  выражаются согласно [1, Supporting Information Section 1], причём при выводе уравнений (19) подразумевается, что продолжительность предъявления одного обучающего примера составляет  $T_s = 1/2$ . Мы не приводим здесь полное громоздкое описание стратегии обучения и метода расчёта параметров  $\lambda_i$ , поскольку для настоящего исследования принципиально лишь то обстоятельство, что при наличии единственного наибольшего среди параметров  $\lambda_i$ , система (19) описывает конкурентную динамику с единственным победителем. В этом случае асимптотический результат обучения всегда представляет собой ансамбль, состоящий из единственного типа клеток, а его функция отклика  $F(x)$  (1) пропорциональна функции отклика одной клетки  $f_i(x)$  единственного оставшегося типа, что принципиально ограничивает возможности обучения классификатора (то есть его подстройки под конкретную задачу классификации). Для получения функ-

<sup>6</sup>Стратегия обучения [1] имеет «параметр мягкости»  $\gamma$ . Используется значение  $\gamma = 1$ , при котором непрерывная модель (19) хорошо описывает популяционную динамику, порождаемую селекцией клеток.

ций отклика, имеющих другой вид (например, при наличии бимодального класса), необходимо прерывание процесса обучения в некоторый не известный заранее момент, предшествующий достижению этого асимптотического состояния [1]. Кроме того, эта стратегия обучения не имеет аналитической оценки порога классификации, что подразумевает необходимость оптимизации величины порога после обучения и дополнительно ограничивает применимость этой стратегии.

## 2. Результаты

**2.1. Временная эволюция процесса обучения.** Выясним, как меняется вероятность правильного ответа классификатора в процессе обучения в рамках постановки задачи, максимально приближенной к исходной [1]. Для этого используем значения параметров генной сети, идентичные [1]:  $A_g = 20$ ,  $\mu_u = \mu_g = m_g = A_u = 1$ ,  $p_g = p_u = 2$ ,  $\alpha = 10^{-3}$ , и рассмотрим коллективный классификатор, состоящий из  $s = 6$  типов клеток со значениями параметров  $m_u^{(i)}$ , указанными в [1, Fig. 2]:

$$m_u^{(i)} = \mu_u A_g \cdot 2^{2(i-1)}, \quad i = 1, \dots, 6. \quad (20)$$

Графики функций отклика клеток с такими параметрами представлены на рис. 2. Продолжительность предъявления одного обучающего примера составляет  $T_s = 1/2$ . Полное количество обучающих примеров  $N_{\text{learn}} = 280$  для унимодальной задачи,  $N_{\text{learn}} = 560$  для бимодальной. Продолжительность свободной эволюции системы (в режиме вырожденной конкуренции) после предъявления всех обучающих примеров составляет  $T_{\text{free}} = 20$ . Полная длительность интервала времени, на котором моделируется система, таким образом, составляет  $N_{\text{learn}} T_s + T_{\text{free}} = 160$  для унимодальной задачи и 300 для бимодальной. Параметры подсистемы обучения  $k^+ = 0.5$ ,  $k^- = 5 \gg k^+$ , постоянная времени  $\tau_l = 10 \gg T_s$ ,  $\tau_{\text{off}} = 0.1 \ll 1$ .

Заметим, что основные результаты моделирования в [1] были получены для случая  $s = 60$ . Реализация внутривидовой конкуренции (которая принципиально необходима для получения режима сосуществования в конкурентной динамике (2a)) для такого количества типов клеток недостижима на современном уровне синтетической биологии (см. раздел Обсуждение результатов), поэтому в настоящем исследовании используется меньшее количество типов клеток.

Результаты моделирования для унимодальной задачи представлены на рис. 3, для бимодальной — на рис. 4. Для унимодальной задачи вероятность правильного ответа классификатора во всех рассмотренных вариантах стратегий обучения стремится к значению, близкому к байесовскому (наблюдаемое превышение над байесовским значением связано с конечной точностью оценки вероятности по тестовым примерам). Для бимодальной задачи обе стратегии (конкуренция с единственным победителем [1] и с сосуществованием [4]) при наличии оптимизации порога классификации в процессе обучения достигают значения вероятности правильного ответа, близкого к байесовскому, однако стратегия с единственным победителем [1] при дальнейшем обучении, как и ожидалось, приводит к падению вероятности правильного ответа (что обусловлено единственностью типа клеток, выживающего в конкуренции). Стратегия конкуренции с сосуществованием свободна от этого недостатка: вероятность правильного ответа в процессе обучения не снижается (за исключением флуктуаций, связанных с вероятностным характером процесса обучения). Более того, эта стратегия работоспособна и без процедуры оптимизации порога классификации: значение порога, оцененное по формуле (11), позволяет достичь вероятности правильного ответа по-

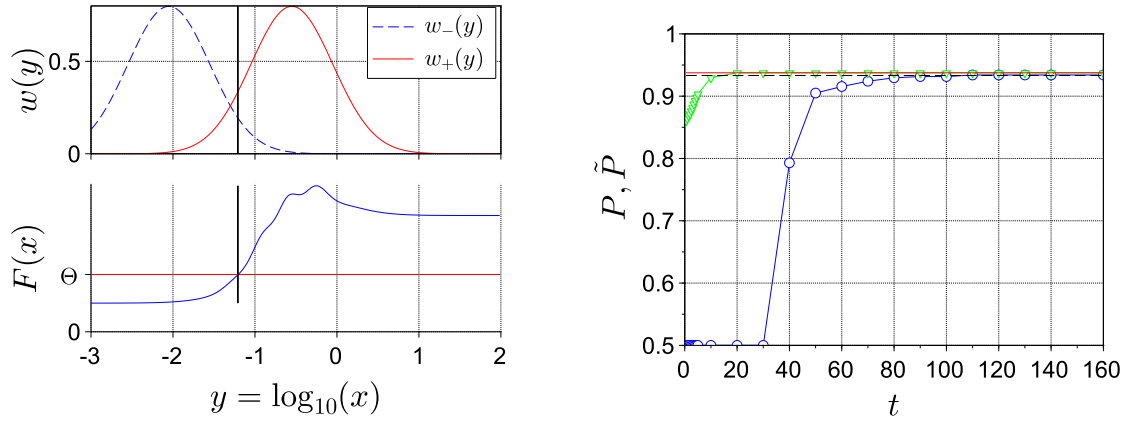


Рис. 3. Обучение коллективного классификатора для унимодальной задачи. Слева: условные плотности вероятности признака для обоих классов  $w_+(y)$  и  $w_-(y)$  (18); суммарная функция отклика  $F(x)$  (1) в конце процесса обучения; оценка порога  $\Theta$  (11); вертикальная линия — граница  $F(x) = \Theta$ , разделяющая области решений классификатора в пространстве признака. Справа: изменение вероятностей правильного ответа классификатора в процессе обучения. Синий график с круглыми маркерами — модель конкуренции с сосуществованием (2a,b) с порогом  $\Theta$  согласно (11); зелёный график с треугольными маркерами — та же модель с оптимизированным значением порога; красный график без маркеров — модель конкуренции с единственным победителем (19) с оптимизированным значением порога; горизонтальная штриховая линия — вероятность правильного ответа байесовского классификатора.

Fig. 3. Collective classifier training for the unimodal problem. Left: conditional probability densities of the feature for the both classes  $w_+(y)$  and  $w_-(y)$  (18); summary response function  $F(x)$  (1) at the end of the learning process; estimate of the threshold  $\Theta$  (11); vertical line is the boundary  $F(x) = \Theta$ , which separates the classifier answers in the feature space. Right: evolution of the classifier's correct answer probability in the course of learning. Blue line with circle markers — competition model with coexistence (2a,b) with threshold  $\Theta$  according to (11); green line with triangle markers — same model with optimized threshold value; red line without markers — winner-take-all competition model (19) with optimized threshold value; horizontal dashed line — correct answer probability for the Bayesian classifier.

рядка 0.85, что заметно хуже байесовской вероятности, однако значительно лучше асимптотического результата стратегии с единственным победителем<sup>7</sup>, который получается, если своевременно не прервать обучение.

**2.2. Влияние количества типов клеток и ширины функций отклика.** Исследуем теперь зависимость результата обучения коллективного классификатора от параметров набора функций отклика клеток  $\{f_i(x)\}$ . Для этого выполним моделирование при различных значениях коэффициента Хилла репортерного промотора  $p_g \in [1; 6]$ , от которого зависит ширина колоколообразных функций отклика (16), и при двух значениях количества типов клеток  $s = 6$  и  $s = 8$ . Значения параметров  $m_u^{(i)}$  задаются в виде геометрической прогрессии (образуют логарифмическую сетку)

$$m_u^{(i)} = m_u^{(1)} \cdot b^{i-1}, \quad b = \left( \frac{m_u^{(s)}}{m_u^{(1)}} \right)^{\frac{1}{s-1}}, \quad (21)$$

где фиксируются граничные значения  $m_u^{(1)} = \mu_u A_g \cdot 2^0$  и  $m_u^{(s)} = \mu_u A_g \cdot 2^{10}$  — таким образом, что сетка (21) совпадает с (20) при  $s = 6$ , а увеличение количества типов клеток  $s$  приводит к более плотному расположению максимумов функций отклика  $f_i(x)$ . Вероятность

<sup>7</sup>в которой значение порога всегда оптимизируется ввиду отсутствия аналитической оценки!

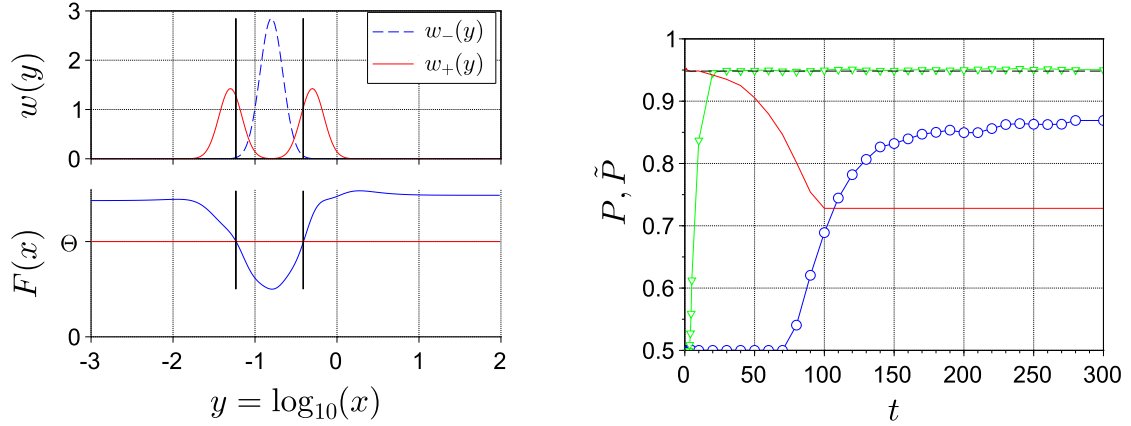


Рис. 4. Обучение коллективного классификатора для бимодальной задачи (обозначения те же, что на рис. 3).  
 Fig. 4. Collective classifier training for the bimodal problem (same legend as in Fig. 3).

правильного ответа классификатора оценивается по окончании обучения, состоящего из 600 обучающих примеров и свободной эволюции  $T_{\text{free}} = 20$ , тем же способом, как в разделе 2.1. Все остальные параметры задаются так же, как в разделе 2.1.

Результаты моделирования представлены на рис. 5 и свидетельствуют о том, что, во-первых, для фиксированного количества типов клеток (то есть для фиксированной плотности сетки максимумов функций отклика  $f_i(x)$ ) существует диапазон оптимальных значений параметра  $p_g$  (то есть оптимальной ширины колоколообразных функций отклика), в котором для рассмотренных задач классификации достигается приближение вероятности правильного ответа к байесовскому значению как в случае оптимизации по порогу классификации, так и без такой оптимизации (при расчёте порога через выражение (11)); во-вторых, с увеличением количества типов клеток ширина этого диапазона увеличивается, а вероятность правильного ответа повышается. Наличие оптимального диапазона значений  $p_g$  объясняется тем, что с увеличением  $p_g$  колоколообразные функции отклика  $f_i(x)$  сужаются, из-за чего суммарная функция отклика классификатора  $F(x)$  при больших  $p_g$  разваливается на отдельные узкие пики и больше не представляет собой адекватную интерполяцию сеточных значений  $F(x_i)$ . Увеличение количества типов клеток приводит к сокращению расстояния между максимумами — а значит, к уменьшению оптимальной ширины пика и к увеличению оптимального значения  $p_g$  и, как следствие, к более точной подстройке отклика обученного классификатора к условным распределениям классов.

### 3. Обсуждение результатов

Общий вид решающего правила (1) накладывает принципиальные ограничения на применимость коллективного классификатора для конкретных задач классификации (в смысле способности аппроксимировать оптимальное решающее правило), независимо от выбора стратегии обучения. Если функции отклика клеток образуют базис линейной интерполяции, то стратегией обучения определяются только значения функции отклика  $F(x_i)$  в узлах дискретной сетки значений признака  $\{x_i\}$ , а промежуточные значения  $F(x)$  получаются линейной интерполяцией. Соответственно, точность аппроксимации оптимального решающего правила всегда ограничена точностью этой интерполяции. Переход к колоколообразным клеточным функциям отклика общего вида (например, к модели синтетической

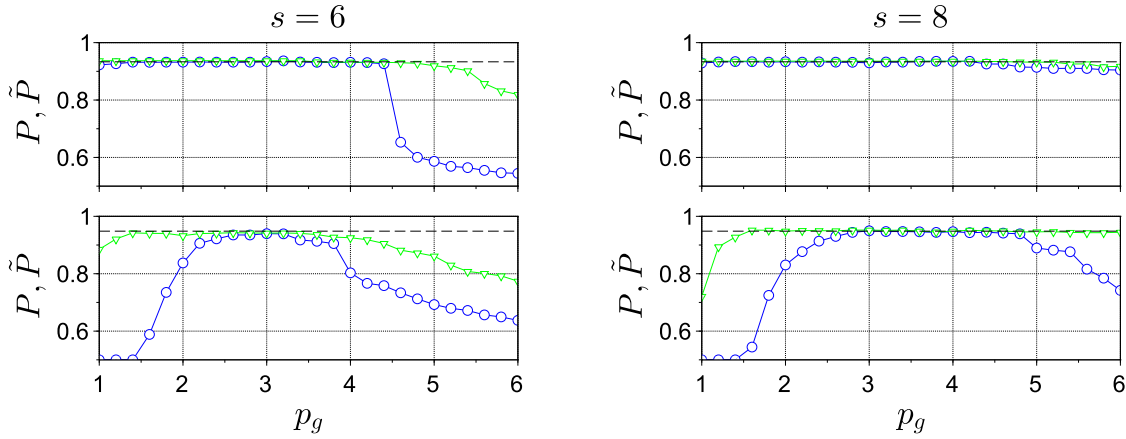


Рис. 5. Зависимость вероятности правильного ответа обученного коллективного классификатора (стратегия обучения на основе конкуренции с сосуществованием (2a,b)) от параметра  $p_g$ , определяющего ширину колоколообразных функций отклика клеток, для количества типов клеток  $s = 6$  и  $s = 8$ . Верхний ряд — унимодальная задача, нижний ряд — бимодальная задача. Синие графики с круглыми маркерами — для значения порога классификации  $\Theta$  согласно (11); зелёные графики с треугольными маркерами — с оптимизацией путём подстройки  $\Theta$ ; горизонтальная штриховая линия — вероятность правильного ответа байесовского классификатора.

Fig. 5. Dependence of the correct answer probability for the trained collective classifier (learning strategy based on competition with coexistence (2a,b)) upon the parameter  $p_g$ , which determines the width of the bell-shaped cell response functions, for cell type quantities  $s = 6$  and  $s = 8$ . Top row — unimodal problem, bottom row — bimodal problem. Blue lines with circle markers — for the threshold value  $\Theta$  as given by (11); green lines with triangle markers — with optimization by tuning  $\Theta$ ; horizontal dashed line — correct answer probability for the Bayesian classifier.

генной цепи (16)) изменяет вид суммарной функции отклика в сравнении с линейной интерполяцией, однако это изменение можно считать относительно небольшим, если характерная ширина функций отклика соответствует шагу сетки максимумов.

Дополнительные ограничения, обусловленные стратегией селекции, предложенной в [1, 2], обсуждавшиеся во Введении, преодолеваются в модифицированной стратегии обучения [3, 4]: во-первых, обученное состояние классификатора при обучении достигается в асимптотическом состоянии, а не в переходном процессе; во-вторых, конкуренция обеспечивается не внешним сортировщиком, а зависимой от обучающих примеров динамикой собственной «подсистемы обучения» элемента ансамбля; в-третьих, при определённых предположениях можно использовать простую оценку порога классификации как удвоенного минимального значения суммарной функции отклика ансамбля. Заметим, что вне процесса обучения, как отмечалось в разделе 1.2, в силу вырожденного характера конкурентной динамики, состояние равновесия больше не является асимптотически устойчивым, то есть состояние может дрейфовать под действием флуктуаций. Более того, система при этом является негрубой в силу предположения о равной конкурентоспособности типов клеток в конкуренции за общий ресурс, и неизбежное в реальности наличие разброса этих конкурентоспособностей переводит конкуренцию в режим с единственным победителем, что рано или поздно разрушает обученное состояние классификатора. Время жизни обученного состояния тем больше, чем меньше вышеупомянутые флуктуации и разброс параметров.

В настоящем исследовании оценена результативность этой модифицированной стратегии обучения на частных примерах задач классификации при использовании биологически обоснованных функций отклика клеток (16), что позволило провести прямое сопоставление

со стратегией селекции [1, 2] по вероятности правильного ответа обученного классификатора. В результате в целом подтверждены вышеперечисленные достоинства новой стратегии.

Во-первых, как и ожидалось в силу вышесказанного, новая стратегия стабильно выигрывает у стратегии селекции в задаче с бимодальным классом (и не уступает в задаче с унимодальными классами), если речь идёт об *асимптотическом* результате процесса обучения.

Во-вторых, если для стратегии селекции рассматривается не асимптотический, а *наилучший* результат, достигаемый в переходном процессе при обучении, то результат оказывается примерно равным в равных условиях, когда для обеих стратегий используется *оптимизация порога* классификации.

В-третьих, результат новой стратегии *без оптимизации порога* (с использованием только аналитической оценки), вообще говоря, может оказаться заметно хуже, чем результат с оптимизацией, однако этот результат остаётся не только лучше случайного угадывания, но и лучше *асимптотического* результата стратегии селекции (в то время как последняя вообще не предусматривает какой-либо оценки значения порога без его подбора).

Вышеперечисленные достоинства достигаются ценой усложнения реализации стратегии обучения, что может оказаться существенным для конкретных задач. Так, для живых клеток требуется реализация механизма «внутривидовой» конкуренции (обеспечивающей равновесное сосуществование типов клеток). В экологии такая конкуренция обеспечивается обычно за счёт «экологических ниш» — специфических условий среды (включая потребляемые ресурсы), благоприятных для размножения конкретного биологического вида и различающихся для разных видов. Бактерии обладают специальным механизмом «кворум-сенсинга», который организует внутривидовую конкуренцию путём обмена через среду специфическим для конкретного вида химическим сигналом (различным у разных видов). Этим механизмом обеспечивается экологически выигрышное ограничение численности популяции на уровне, не достигающем предела, обусловленного средой. Именно механизм кворум-сенсинга (а также синтетические механизмы, получаемые целенаправленной модификацией природных сигнальных систем) представляется наиболее перспективным способом искусственной реализации конкуренции с сосуществованием в ансамблях живых клеток. Количество конкурирующих типов клеток (и, как следствие, — универсальность классификатора) при этом будут ограничены количеством доступных ортогональных (то есть воспринимаемых только или преимущественно клетками своего типа) сигналов межклеточного взаимодействия. Построение таких ортогональных сигнальных систем (в частности, наращивание количества сигналов) является актуальной проблемой синтетической биологии; так, в работе [8] было достигнуто количество типов клеток, обменивающихся взаимно ортогональными сигналами, равное 8, чем и обусловлен выбор количества типов клеток в настоящем исследовании. Стратегия селекции [1, 2] не имеет такого ограничения.

Кроме того, в новой стратегии появляется «подсистема обучения» как собственная динамика элемента (клетки), которая отсутствует в стратегиях, основанных на внешнем сортировщике. Эта динамика также должна быть реализована в клетке, что повышает общую сложность синтетической генной сети клетки. Подсистема обучения должна обеспечивать эффективное усреднение влияния отдельных обучающих примеров для снижения дисперсии флуктуаций переменных в процессе обучения. Это требует большого характерного времени динамики подсистемы обучения в сравнении с длительностью предъявления одного обучающего примера  $\tau \gg T_s$  и, как следствие, замедляет процесс обучения в сравнении с селекцией клеток.

Надо, однако, учитывать, что относительная простота биологической части в реализации стратегии обучения, основанной на селекции клеток, обусловлена тем, что ряд

функций в процессе обучения (собственно сортировка клеток, оптимизация порога классификации, прерывание обучения в оптимальный момент) реализуются сложными внешними средствами, а не самой системой, тогда как в рамках новой стратегии, основанной на конкуренции с сосуществованием, не требуются ни сортировка клеток, ни прерывание обучения, а оптимизация порога не необходима, хотя и повышает точность. Таким образом, с точки зрения экспериментальной реализации коллективного классификатора, стратегии с простейшей биологической составляющей пока остаются предпочтительными; в то же время, перспективная цель создания коллективных классификаторов как самообучающихся систем, функционирующих на принципах синтетических (искусственных) экологических сообществ, неизбежно связана с минимизацией участия внешних средств в обучении системы и с соответствующим усложнением функциональности, реализуемой в самой системе средствами синтетической биологии.

Совокупность полученных результатов подтверждает принципиальную совместимость механизма обучения [3, 4] с биологически реализуемыми функциями отклика и открывает путь к экспериментальной реализации коллективного классификатора в виде ансамбля живых клеток с синтетическими генными структурами [1] при условии надлежащего согласования параметров генной схемы с характеристиками решаемой задачи.

### Список литературы

1. *Didovyk A., Kanakov O. I., Ivanchenko M. V., Hasty J., Huerta R., Tsimring L.* Distributed classifier based on genetically engineered bacterial cell cultures // *ACS Synthetic Biology*. 2015. Vol. 4, no. 1. P. 72–82. DOI: 10.1021/sb500235p.
2. *Kanakov O., Kotelnikov R., Alsaedi A., Tsimring L., Huerta R., Zaikin A., Ivanchenko M.* Multi-input distributed classifiers for synthetic genetic circuits // *PLoS ONE*. 2015. Vol. 10, no. 5. P. e0125144. DOI: 10.1371/journal.pone.0125144.
3. *Сутягин А. А., Канаков О. И.* Метод обучения коллективного классификатора на основе конкуренции в режиме сосуществования // *Известия вузов. ПНД*. 2021. Т. 29, № 2. С. 220–239. DOI: 10.18500/0869-6632-2021-29-2-220-239.
4. *Сутягин А. А., Канаков О. И.* Механизм обучения коллективного классификатора на основе конкуренции, управляемой обучающими примерами // *Известия вузов. ПНД*. 2024. Т. 32, № 2. С. 160–179. DOI: 10.18500/0869-6632-003089.
5. *Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д.* Прикладная статистика: Классификация и снижение размерности. М.: Финансы и статистика, 1989. 608 с.
6. *Alpaydin E.* Introduction to Machine Learning. Fourth Edition. Cambridge, Massachusetts: MIT Press, 2020. 683 p.
7. *Nowicka M., Siebert H.* Designing Distributed Cell Classifier Circuits Using a Genetic Algorithm // *Computational Methods in Systems Biology. CMSB 2019. Lecture Notes in Computer Science*. Cham: Springer, 2019. P. 96–119. DOI: 10.1007/978-3-030-31304-3\_6.
8. *Du P., Zhao H., Zhang H., Wang R., Huang J., Tian Y., Luo X., Luo X., Wang M., Xiang Y., Qian L., Chen Y., Tao Y., Lou C.* De novo design of an intercellular signaling toolbox for multi-channel cell–cell communication and biological computation // *Nature communications*. 2020. Vol. 11, no. 1. P. 4226. DOI: 10.1038/s41467-020-17993-w.

### References

1. *Didovyk A, Kanakov OI, Ivanchenko MV, Hasty J, Huerta R, Tsimring L.* Distributed classifier based on genetically engineered bacterial cell cultures. *ACS Synthetic Biology*. 2015;4(1):72–82. DOI: 10.1021/sb500235p.

2. Kanakov O, Kotelnikov R, Alsaedi A, Tsimring L, Huerta R, Zaikin A, Ivanchenko M. Multi-input distributed classifiers for synthetic genetic circuits. PLoS ONE. 2015;10(5):e0125144. DOI: 10.1371/journal.pone.0125144.
3. Sutyagin AA, Kanakov OI. Collective classifier learning strategy based upon competition in the coexistence regime. Izvestiya VUZ. Applied Nonlinear Dynamics. 2021;29(2):220–239 (in Russian). DOI: 10.18500/0869-6632-2021-29-2-220-239.
4. Sutyagin AA, Kanakov OI. Learning mechanism for a collective classifier based on competition driven by training examples. Izvestiya VUZ. Applied Nonlinear Dynamics. 2024;32(2):160–179. DOI: 10.18500/0869-6632-003089.
5. Aivazyan SA, Buchstaber VM, Yenyukov IS, Meshalkin LD. Applied Statistics: Classification and Reduction of Dimensionality. Moscow: Finansy i Statistika; 1989. 608 p. (in Russian).
6. Alpaydin E. Introduction to Machine Learning. Fourth Edition. Cambridge, Massachusetts: MIT Press; 2020. 683 p.
7. Nowicka M, Siebert H. Designing Distributed Cell Classifier Circuits Using a Genetic Algorithm. In: Computational Methods in Systems Biology. CMSB 2019. Lecture Notes in Computer Science. Cham: Springer, 2019. P. 96–119. DOI: 10.1007/978-3-030-31304-3\_6.
8. Du P, Zhao H, Zhang H, Wang R, Huang J, Tian Y, Luo X, Luo X, Wang M, Xiang Y, Qian L, Chen Y, Tao Y, Lou C. De novo design of an intercellular signaling toolbox for multi-channel cell–cell communication and biological computation. Nature communications. 2020;11(1):4226. DOI: 10.1038/s41467-020-17993-w.



*Сутягин Алексей Алексеевич* – родился в Нижнем Новгороде (1992). Окончил с отличием радиофизический факультет Нижегородского государственного университета им. Н.И. Лобачевского (ННГУ) по направлению «Радиофизика» (2016). Неоднократный победитель и призёр творческих и научных конкурсов. С 2019 преподаёт на кафедре теории колебаний и автоматического регулирования ННГУ в должности ассистента. Научные интересы – конкурентная динамика, коллективные классификаторы. Зарегистрировал РИД в рамках научной работы.

Россия, 603950 Нижний Новгород, проспект Гагарина, 23

Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского

E-mail: alealesutyagin@gmail.com



*КанакOV Олег Игоревич* – родился в Нижнем Новгороде (1981). Окончил с отличием радиофизический факультет Нижегородского государственного университета им. Н.И. Лобачевского (ННГУ) по направлению «Радиофизика» (2004). Защитил диссертации на соискание учёной степени кандидата физико-математических наук на тему «Структуры и локализация энергии в нелинейных решёточных системах» (2007, ННГУ) и доктора физико-математических наук на тему «Структурообразование в ансамблях нелинейных элементов (неустойчивости, синхронизация, конкуренция, механизмы обучения)» (2017, ННГУ) по специальности «Радиофизика». С 2007 года преподаёт на кафедре теории колебаний и автоматического регулирования ННГУ (с 2019 – в должности профессора). Научные интересы – нелинейная динамика биологических моделей, коллективные классификаторы, локализованные решения и взаимодействие мод в нелинейных гамильтоновских решётках. Опубликовал 30 научных статей по указанным направлениям.

Россия, 603950 Нижний Новгород, проспект Гагарина, 23

Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского

E-mail: okanakov@rf.unn.ru