利用基因线路构建神经网络实现神经拟态计算

杨珊[†], 刘芮存[†], 刘拓宇, 庄滢潭, 李金玉, 滕越^{*}

军事科学院军事医学研究院微生物流行病学研究所, 病原微生物生物安全国家重点实验室, 北京 100071

[†] 同等贡献

^{*} 联系人, E-mail: yueteng@me.com

2021-07-29 收稿, 2021-08-14 修回, 2021-08-16 接受, 2021-08-17 网络版发表

摘要 随着大数据时代的到来, 现有计算体系限制了人工智能与类脑计算等新技术的发展. 基于人工神经网络的神经拟态计算则提供了潜在解决方案, 而具备低能耗并行化等优势的生物计算对其研究至关重要, 其中基因线路将是构造人工神经网络的关键. 本研究利用生物元件构建设计了人工神经网络, 并实现了线性分类、非线性分类以及图案分类等神经拟态计算应用, 充分利用生物元件的特性, 模拟神经网络中神经元的连接关系, 根据生物元件在基因线路中产生不同响应作为元件选择标准, 构建基因线路实现不同功能的神经拟态计算. 本研究提出通过工程化基因线路构建人工神经网络实现神经拟态计算, 由此集成的分子计算系统有望应用于人工智能芯片的制造, 并进一步广泛应用于类脑计算、脑机接口及国防建设等领域.

关键词 神经拟态计算, 人工神经网络, DNA计算, 基因线路, 合成生物学

神经拟态计算(neuromorphic computing), 也称为神经形态工程(neuromorphic engineering), 是使用包含电子模拟电路的超大规模集成系统(very-large-scale integration, VLSI)来模仿人脑中神经网络的计算结构^[1~3]. 其核心思想是, 以数学模型来模拟人类神经元(neuron)及突触(synapse)的结构, 并结合多层次传导来模拟大量神经元联结^[4,5]. 这种仿生学方法创造了高度连接的合成神经元和突触结构, 从而可以执行复杂的计算任务^[6]. 利用人工神经网络实现神经拟态计算已在深度学习取得了空前的成功, 并广泛应用于人工智能的各种研究方向^[7~9]. 其如何定量、准确、简化地描述神经元和突触, 直接决定了计算系统的性能、功耗与设计复杂度.

与基于硅材料的传统计算系统不同, 以DNA(deoxyribonucleic acid)分子等生物材料为基础的生物计算, 因具有超低能耗并行处理信息的能力而备受关注^[10~13]. DNA分子具有互补的碱基配对特性, 可以特异性地实

现分子识别、自组装和大规模平行反应, 基于DNA链置换反应的生物分子计算向人们展示了分子系统如何表现出类脑计算的自主行为, 使计算形式得到了革命性的突破^[14~20]. 使用简单的DNA逻辑门结构, 可以系统地将人工神经网络转换为DNA置换链级联反应, 并可以执行类似大脑的学习和记忆等操作^[14]. 这展示了利用生物材料构造人工神经网络的巨大发展前景, 并将促进神经拟态计算的广泛部署. 然而, 通过DNA置换链级联反应构建神经网络也面临诸多挑战^[21]. 首先, DNA链置换反应使用为特定计算而设计的预编程一次性体系, 执行不同逻辑电路任务将需要不同的一组DNA分子; 其次, DNA或酶反应多为不可逆反应, 导致逻辑门不能反复地“打开”和“关闭”, 从根本上限制了DNA计算电路从单层扩展到多层结构^[22~26].

基因线路作为一种新型的DNA计算方式为解决上述问题提供了新的思路. 在利用基因线路构建的人工神经网络中, 基因电路元件称为动态单元, 这些动态单

引用格式: 杨珊, 刘芮存, 刘拓宇, 等. 利用基因线路构建神经网络实现神经拟态计算. 科学通报, 2021, 66: 3992–4002

Yang S, Liu R C, Liu T Y, et al. Constructing artificial neural networks using genetic circuits to realize neuromorphic computing (in Chinese). Chin Sci Bull, 2021, 66: 3992–4002, doi: 10.1360/TB-2021-0501

元可以随着输入或输出的不同而反复使用,并可动态地将上游电路的输出直接连接到下游电路的输入,从而实现更复杂的逻辑模式识别^[27~29]。本研究通过基因线路设计构建了人工神经网络,实现了线性分类、非线性分类以及图案分类等神经拟态计算应用。

1 方法

利用基因线路构建神经网络,首先需要构建人工神经网络的结构,其次通过希尔方程描述的生化反应为依据,制定生物元件的选择标准,最后完成从神经网络到基因线路的转化。

1.1 通过基因线路构建人工神经网络的结构设计

在利用基因线路构建的神经网络结构中,我们通过基因的表达与调控实现复杂的神经拟态计算。图1展示了神经网络结构与基因线路的对应关系,神经网络的结构包括神经元以及神经元间的连接关系。其中,神经网络的输入层、隐藏层和输出层的神经元,分别对应基因线路中的诱导剂(例如异丙基-β-D-硫代半乳糖苷、阿拉伯糖等)、阻遏蛋白(基因线路的中间产物)以及荧光蛋白(最终输出产物)。其中基因线路中的所有生物元件均根据其在对应启动子下测量荧光值,并以相对启动子单元(relative promoter unit, RPU)表示^[30]。本研究使用的生物元件库来源于Genetic Circuit Design Automation^[30](表S1)。

神经网络中神经元的联结关系即基因线路中输入生物元件与输出生物元件的调控关系,即加权运算(权重)与激活运算(激活函数)。激活函数主要的作用是进行数据的非线性变换。在基因线路的表达调控中,其输

入与输出的响应关系符合希尔方程(式(1),方法3.1),即基因表达是可以完成输入和输出的非线性变换^[31]。权重值的设置为0或1,表示连接两个生物元件的启动子是否存在。因此,式(2)表示在基因线路中输入元件X激活或抑制启动子促进或阻止输出元件Y的生成的加权运算与激活运算过程。基于输入元件 X_1 、 X_2 通过式(2)得到的 Y_1 、 Y_2 在扩散反应中实现求和过程,式(3)进行了求和、加权以及激活运算过程:

$$y = y_{\min} + \frac{(y_{\max} - y_{\min})K^n}{x^n + K^n}, \quad (1)$$

$$f((X) \times W) \rightarrow Y, \quad (2)$$

$$f((Y_1 + Y_2) \times W) \rightarrow Y_3. \quad (3)$$

式(1)为希尔方程,其中 $y_{\max}$ 和 $y_{\min}$ 分别为观测RPU的最大值和最小值, K 是响应阈值即最大输出RPU的一半处对应的输入RPU的大小, n 是希尔方程的系数。因此,不同启动子所对应的希尔方程是不同的。式(2)中 X 为输入RPU, W 为权重,箭头右侧 Y 为输出RPU, $f()$ 为激活函数,即对应的希尔方程曲线。式(3)中 Y_1 和 Y_2 是式(2)在不同输入下的输出。

1.2 基于希尔方程的激活函数运算

希尔方程描述了高分子被配体饱和的分数是一个关于配体浓度的函数,该函数的曲线会呈现S形(式(1))^[31]。为了便于基因线路的设计,将希尔方程的输入和输出均定义为RPU。在本研究中,诱导剂通过作用于启动子区域,激活下游基因表达的生化反应过程符合希尔方程激活型曲线(图S1(a));阻遏蛋白(中间产物)抑制下游基因表达的生化反应过程符合希尔方程抑制型

图1 神经网络与基因线路对应关系示意图。利用基因线路实现神经网络的结构设计,可由人工神经网络得到基因线路

Figure 1 Mapping between neural networks and genetic circuits. The structural design of the neural network can be realized by using the genetic circuit, and the genetic circuit can be obtained from the artificial neural network

曲线(图S1(a)), 其中基因线路中的所有生物元件均根据其在对应启动子下测量荧光值并以RPU表示。

连接两个生物元件的启动子根据其存在与否, 可以分为开状态和关状态, 与神经网络的权值1和0对应。因此, 我们设定了开阈值和关阈值。开阈值的大小是输入元件RPU为输出元件RPU 0.5倍时的最小值(图S1(b)中的蓝色区域), 关阈值的大小是输入元件RPU为输出元件RPU 2倍时的最大值(图S1(b)中的红色区域)。当输入启动子的活性(RPU)在0到开阈值时, 输出启动子的活性(RPU)为开状态; 当输入启动子的活性(RPU)在关阈值到无穷时, 输出启动子的活性(RPU)为关状态。

1.3 实现信号传输的生物元件的选择标准

表S1列举了18种生物元件作为本研究的生物元件库^[30]。为了保证基因线路的准确性, 在选择合适的实现信号传输的生物元件时, 需要确保前一个元件的输出在后一个元件输入的有效范围内(图S1(b)的红色或蓝色区域)。以一个简单的线路图(图S2)为例, 该电子线路中当输入为0时, 输出仍是0; 输入为1时, 输出仍为1。实现该功能的基因线路由两个启动子构成, 且输入与第1个启动子结合后转录翻译产生的小分子, 可作用于第2个启动子。图S2(a)中的第1个神经元两种状态下的输出均在第2个神经元的有效范围内, 且第2个神经元的输出符合该线路的功能。图S2(b)中的第1个神经元两种状态下的输出均在第2个神经元的有效范围内, 但是第2个神经元的输出不符合该线路功能, 所以需要重新进行选择。

1.4 神经网络到基因线路的转化过程

在基因线路中, 将神经网络中的神经元视为生物元件, 根据神经网络的层次与连接关系设计基因线路的结构与调控关系。首先, 每层基因线路的结构与神经网络每个隐藏层相对应, 隐藏层中神经元的个数与每层基因线路中生物元件的个数相同。其次, 隐藏层某一神经元与上一层神经元的权重为1时, 则与之对应的生物元件由上一层生物元件调控产生。调控是启动子受到上一层神经元刺激或抑制的生化过程。最后, 神经网络的输出层对应的是生物元件荧光蛋白, 因此输出层可以视为一个感受器。该感受器的输入为上一隐藏层的生物元件, 感受器的输出为荧光蛋白。此外, 神经网络的输入是激活第一个隐藏层中启动子的小分子。通过上述步骤, 将得到实现神经网络的基因线路。

2 结果

基因线路构建神经网络实现神经拟态, 通过神经网络的结构可以得到基因线路中每个生物元件的连接关系, 利用迭代的方式在生物元件库中寻求最合适的元件, 判断元件的标准是模拟基因线路的输出与神经网络的需求是否一致。基于上述过程, 我们在线性分类、非线性分类和图案分类中进行了模拟仿真, 其工程化结果与预期一致。

2.1 基因线路构建神经网络实现线性分类

我们以线性分类中包含的与非门(NAND gate)作为研究对象, 展示了利用基因线路构建神经拟态计算的方法。图2(a)为一个NAND的线性分类, 图2(b)为实现该分类的神经网络, 图2(c)为实现该神经网络的基因线路。首先输入元件 X_1 作用于启动子 P_1 , 从而激活下游基因进行转录翻译生成隐藏层元件蛋白质 Y_1 。同理, 输入元件 X_2 作用于启动子 P_2 , 从而激活下游基因进行转录翻译生成隐藏层元件蛋白质 Y_2 。在扩散作用下, Y_1 、 Y_2 分别作用于启动子 P_3 、 P_4 , 抑制下游基因转录翻译生成最终输出元件 Y_3 。为了便于定性与定量的检测, Y_3 一般为荧光蛋白。

图3展示了上述基因线路构建神经网络实现线性分类的仿真结果。在图3(a)中输入为“00”, 在图3(b)中输入为“01”, 在图3(c)中输入为“10”, 此时输出元件 Y_3 的RPU较高, 表示输出信号为“1”。在图3(d)中输入为“11”, 此时输出元件 Y_3 的RPU较低, 表示输出信号为“0”。即图3(a)~(c)中两个纵坐标之和为其输出对应的开状态, 而图3(d)中两个纵坐标之和为其输出对应的关状态。在基因线路构建神经网络实现线性分类的仿真中, 各个生物元件(输入元件、输出元件及隐藏层元件)使用的具体RPU数值, 如表1所示。因此, 该基因线路通过神经网络的方式并以输出荧光值作为检测标准实现了NAND的功能。

同时, 我们还以线性分类中包含的与门(AND)作为研究对象, 展示了基因线路构建神经拟态计算的过程与仿真结果(图S3、S4以及表S2)。

2.2 基因线路构建神经网络实现非线性分类

我们设计的单层神经网络除了可以实现上述NAND和AND的线性可分的分类之外, 还可处理线性不可分问题。因此, 在神经元实现逻辑运算NAND、

图 2 基因线路构建神经网络实现线性分类的过程. (a) NAND的线性分类, 通过超平面将空间中的4个数据分为两类, 其中(0,0)、(0,1)和(1,0)属于一类, (1,1)属于另一类, 符合NAND的分类结果. (b) 实现线性分类的神经网络, 该神经网络有3层, 依次含有2、2和1个神经元. 本图仅展示了神经元间的部分连接关系. (c) 实现线性分类神经网络的基因线路, 其包括5个生物元件, 不同层间的生物元件通过基因调控进行连接

Figure 2 The process of constructing neural network to realize linear classification. (a) The linear classification of NAND. The four datasets in the space are divided into two categories by hyperplane, in which (0,0), (0,1) and (1,0) belong to one category, and (1,1) belongs to another category, conforming to the classification results of NAND. (b) The neural network for linear classification. The neural network has three layers, containing 2, 2 and 1 neurons in sequence. (b) shows only partial connections between neurons. (c) The genetic circuit of the linear classification neural network, which includes five biological elements, and the biological elements between different layers are connected by gene regulation

图 3 基因线路构建神经网络实现线性分类的仿真结果. 在实现线性分类基因线路的仿真结果中, 输入和输出均为生物元件的RPU. 蓝色曲线代表的是希尔方程. 橙色线段与横坐标的交点代表的是输入生物元件的RPU. 绿色线段与纵坐标的交点代表输出的RPU. NAND基因线路的最终输出, 为图3两个仿真图输出RPU之和. 第一列为基因线路的布尔输入, 第二列为基因线路的布尔输出, 第三列为基因线路的模拟输出

Figure 3 Simulation results of linear classification based on a neural network constructed from a genetic circuit. In the simulation result of realizing a linear classification genetic circuit, the input and output are the RPU of the biological component. The blue curve is Hill's equation. The intersection of the orange segment and the abscissa represents the RPU of the input biological component. The intersection of the green line segment and the ordinate represents the output RPU. The final output of the NAND gene circuit is the sum of the output RPU of the two simulation graphs in Figure 3. The first column is the Boolean input of the genetic circuit, the second column is the Boolean output of the genetic circuit, and the third column is the analog output of the genetic circuit

NOT等的基础上,可以级联形成多层神经网络,来解决复杂的线性不可分问题.我们以异或门(XOR)作为非线性分类的研究对象,通过基因线路构建神经拟态计算.通过神经网络将两个相同的输入分为一个类,将另外两个不同的输入分为另一个类,因此应用多层神经网络可以解决线性不可分问题.图4(a)为一个XOR的非线性分类,图4(b)为实现该非线性分类的神经网络,图4(c)为实现该神经网络的基因线路.其中, X_1 、 X_2 为输入元件, Y_1 、 Y_2 和 Y_3 为两层隐藏层中的元件, Y_4 为最终输出元件,并按如图4所示方式进行基因的表达和调控.

图5展示了基因线路构建神经网络实现非线性分类运算XOR的仿真结果.在图5(a)中输入为“00”,在图5(d)中输入为“11”,此时输出元件 Y_4 的RPU较低,表示输出信号为“0”.在图5(b)中输入为“01”,在图5(c)中输入

为“10”,此时输出元件 Y_4 的RPU较高,表示输出信号为“1”.即图5(a), (d)中输出两个纵坐标之和为其输出对应的关状态,而图5(b), (c)中输出两个纵坐标之和为其输出对应的开状态.在基因线路构建神经网络实现非线性分类的仿真中,各个生物元件(输入元件、输出元件及隐藏层元件)使用的具体RPU数值,如表2所示.因此,该基因线路通过神经网络的方式,并以输出荧光值作为检测标准,实现了XOR的功能.

2.3 基因线路构建神经网络实现图案分类

为了验证神经网络的计算能力,本文实现了一个简单的图案分类.在图案分类中,我们对图6(a)中的下层 3×3 网格形成的图案进行分类,这些图案是由3个输入控制(图6(a)中的上层 1×3 网格). 1×3 网格中每个格子

表1 线性分类NAND的输入元件、输出元件以及隐藏层元件的RPU值

Table 1 RPU values of input element, output element and hidden layer element of linear classification NAND

Input(X_1)	Input(X_2)	Hidden layer(P_1 - Y_1 - P_3)	Hidden layer(P_2 - Y_2 - P_4)	Output(Y_3)
0.0082	0.0013	3.7996	3.7936	7.5932
0.0082	4.4	3.7996	0.0649	3.8645
2.5	0.0013	0.1180	3.7936	3.9116
2.5	4.4	0.1180	0.0649	0.1829

图4 基因线路构建神经网络实现非线性分类的过程. (a) XOR的非线性分类,通过超平面不能将空间中的4个数据按XOR进行分类,其中(0,0)和(1,1)属于一类,(0,1)和(1,0)属于另一类. (b)实现非线性分类的神经网络,该神经网络有4层,依次含有2、1、2和1个神经元. (c)仅展示了神经元间的部分连接关系. (c)实现非线性分类神经网络的基因线路,其包括6个生物元件,不同层间的生物元件通过基因调控进行连接

Figure 4 The process of constructing a neural network of a gene circuit to realize nonlinear classification. (a) The nonlinear classification of XOR. It is not possible to classify the four datasets in space according to XOR by hyperplane, where (0,0) and (1,1) belong to one class, and (0,1) and (1,0) belong to another class. (b) The neural network for nonlinear classification. The neural network has four layers, including 2, 1, 2 and 1 neurons in sequence. (b) Only shows partial connections between neurons. (c) The genetic circuit of the nonlinear classification neural network, which includes six biological elements, and the biological elements between different layers are connected by gene regulation

图 5 基因线路构建神经网络实现非线性分类的仿真结果. 在实现非线性分类的基因线路的仿真结果中, 输入和输出均为生物元件的RPU. 蓝色曲线代表的是希尔方程. 橙色线段与横坐标的交点代表的是输入生物元件的RPU. 绿色线段与纵坐标的交点代表输出的RPU. 第一列为基因线路的布尔输入, 第二列为基因线路的布尔输出, 第三列为基因线路的模拟输出. XOR基因线路的最终输出, 为图5两个仿真图输出RPU之和

Figure 5 Simulation results of a neural network constructed from a gene circuit to realize nonlinear classification. In the simulation results of the gene circuit to achieve nonlinear classification, the input and output are the RPU of the biological element. The blue curve is Hill's equation. The intersection of the orange segment and the abscissa represents the RPU of the input biological element. The intersection of the green line segment and the ordinate represents the output RPU. The first column is the Boolean input of the genetic circuit, the second column is the Boolean output of the genetic circuit, and the third column is the analog output of the genetic circuit. The final output of the XOR gene circuit is the sum of the output RPU values of the two simulation graphs in Figure 5

表 2 非线性分类XOR的输入元件、输出元件以及隐藏层元件的RPU值

Table 2 RPU values of the input element, output element and hidden layer element of nonlinear classification XOR

Input(X_1)	Input(X_2)	Hidden layer($(P_1+P_2)-Y_1-P_3$)	Hidden layer($(P_1+P_3)-Y_2-P_4$)	Hidden layer($(P_2+P_3)-Y_3-P_5$)	Output(Y_4)
0.0082	0.0013	3.7995	0.0710	0.0200	0.091
0.0082	4.4	0.0824	2.1904	0.0200	2.2104
2.5	0.0013	0.1179	0.0726	2.3987	2.4713
2.5	4.4	0.0742	0.0727	0.0200	0.0927

代表一个输入, 当输入存在时, 该格子为绿色, 对应的3×3网格形成具有黄色形状的模式. 不同输入构成的图案可以进行叠加, 上述这些3×3网格形成的图案也被称为“内存”. 目标模式(图6(b))作为权重, 示例网络具有两

个目标模式, 即网络可以记住两个模式“L”和“T”. 网络中存在3个输入, 通过不同的排列组合, 可以形成8种不同的“内存”(图6(c)). 这8种“内存”是含有噪声的“L”和“T”. 网络通过将“内存”与目标模式进行比较, 用于判

图6 基因线路构建神经网络实现图案分类的过程。(a) 输入与图案的对应关系。3个输入由上层1×3网格表示, 当输入存在时, 在1×3网格中用绿色格子表示。由输入控制的“内存”用下层3×3网格中黄色格子构成的图案表示。(b) 图案分类中的两个目标模式“L”和“T”。(c) 不同输入对应的“内存”, 以及这些“内存”的所属模式

Figure 6 The process of constructing a neural network by a genetic circuit to realize pattern classification. (a) The correspondence between input and pattern. The three inputs are represented by the upper 1×3 grid, and when the input is present, it is represented by a green grid in the 1×3 grid. The input-controlled “memory” is represented by a pattern of yellow cells in the underlying 3-by-3 grid. (b) Two target patterns “L” and “T” in the pattern classification. (c) The “memories” corresponding to the different inputs and the mode in which these “memories” belong to

断“内存”与目标模式中的“L”和“T”中的哪个模式最为相似。判断的依据是“内存”分别与目标模式进行权重和运算, 计算结果越大, 则“内存”与该目标模式越相近。若计算结果相同, 则认定“内存”为目标模式“L”。图6(c)所包含8种不同的“内存”, 3种与目标模式“T”情况相似, 5种与目标模式“L”情况相似。

图7展示了基因线路构建神经网络实现图案分类的仿真结果。图7(a)为实现图案分类的神经网络, 图7(b)为实现图案分类神经网络的基因线路。每个“内存”通过对应3个输入元件的有无形成了由二进制数(0/1)构成的属于自己的编号。该编号类似于基因线路中输入元件 X_1 、 X_2 、 X_3 的存在与否。在基因线路中 X_1 、 X_2 、 X_3 为输入元件, Y_1 、 Y_2 、 Y_3 和 Y_4 为3层隐藏层中的元件, Y_5 为最终输出元件。通过对输入元件 X_1 、 X_2 、 X_3 的不同状态进行响应, 按如图7(b)所示方式进行基因的表达和调控, 通过荧光蛋白的荧光值作为检测标准判断“内存”的分类情况。图7(c)仿真图纵坐标的RPU代表“内存”的分类情况。在图7(c)中当输入为“000”、“010”、“100”、“101”、“110”时, 输出元件 Y_5 的RPU较低, 表示输出信号为“0”, 即在这5种输入情况下构成的“内存”与目标模式“L”更接近。在图7(c)中当输入为“001”、“011”、“111”, 此时输出元件 Y_5 的RPU较高, 表示输出信号为“1”, 即在这3种输入情况下构成的“内存”与目标模式“T”更接近。在基因线路构建神经网络实现图案分类的仿真中, 各个生物元件(输入元件、输出元件及隐藏层元件)使用的具体RPU数值, 如表3所示。因此, 该基因线路通过神经网络的方式, 并以输出荧光值作为检测标准, 实现了图案分类。

3 讨论

神经拟态是计算机实现对人类大脑的模拟, 并利用大脑的工作机制, 最终使计算机或机器具备如人脑一样的近自然智能特性, 从而使计算机或机器具备更低能耗、更加高效和容错率更高的能力。由我们第一次提出的利用基因线路设计人工神经网络, 其可以实现线性分类、非线性分类和图案分类, 并通过仿真实验进行验证。仿真实验结果证明了神经元可以进行级联, 通过级联构建的神经网络具有计算能力, 从而基因线路可以实现神经网络的简单分类功能。

目前已有将生物学与神经网络相结合的工作。Cherry和Qian^[32]采用DNA置换链构建竞争神经网络, 以完成图像识别的任务。Li等人^[33]基于细胞的神经网络, 接受细菌通过群体感应与发送细菌共同作用以进行决策, 完成图案识别的任务。与依赖DNA置换链反应和细胞群体感应的神经网络不同, 我们利用基因线路构建神经网络, 采用基因表达中的转录翻译调控过程进行信号的传输, 这种方式具有独有的优势^[34~37]。在基因线路构建的神经网络中, 作为动态单元的生物元件可随输入或输出的不同而反复使用, 并可动态地将神经网络隐藏层直接连接到下一隐藏层的输入, 从而实现更复杂的逻辑模式识别。这在一定程度上解决了基于DNA置换链反应的神经网络仅为预编程的一次性体系的缺陷, 也实现了计算电路由单层到多层结构的扩展。

但我们设计的神经网络还是存在不足。首先, 网络中的权重并非任意取值, 而与生物元件的特性息息相关。但目前的生物元件库中的元件个数不足, 致使神经

图 7 基因线路构建神经网络实现图案分类的仿真结果. (a) 实现图案分类的神经网络, 该神经网络有5层, 依次含有3、2、1、1和1个神经元. 本图仅展示了神经元间的部分连接关系. (b) 实现图案分类神经网络的基因线路, 其包括8个生物元件, 不同层间的生物元件通过基因调控进行连接. (c) 不同输入下产生的“内存”, 以及基因线路判断的模式. 在实现图案分类的基因线路的仿真结果中, 输入和输出均为生物元件的RPU. 蓝色曲线代表的是希尔方程. 橙色线段与横坐标的交点代表的是输入生物元件的RPU. 绿色线段与纵坐标的交点代表输出的RPU. 第1列根据不同输入情况形成的图案, 第2列为基因线路的布尔输出, 第3列为基因线路的模拟输出, 第4列为基因线路判断图案的所属模式. 在图案分类基因线路的最终输出为最右列仿真图输出的RPU

Figure 7 Simulation results of a neural network constructed by a genetic circuit to realize pattern classification. (a) A neural network for pattern classification. The neural network has five layers, containing 3, 2, 1, 1 and 1 neurons in sequence. (a) shows only partial connections between neurons. (b) The genetic circuit of the neural network for pattern classification. It includes eight biological elements, and the biological elements between different layers are connected by gene regulation. (c) The “memory” generated under different inputs and the pattern of genetic circuit judgment. In the simulation results of the genetic circuit to realize pattern classification, the input and output are the RPU of the biological element. The blue curve is Hill’s equation. The intersection of the orange segment and the abscissa represents the RPU of the input biological element. The intersection of the green line segment and the ordinate represents the output RPU. The first column is the pattern formed according to different input conditions, the second column is the Boolean output of the gene circuit, the third column is the analog output of the gene circuit, and the fourth column is the pattern of the gene circuit. The final output of the pattern classification gene circuit is the RPU output of the simulation diagram in the far right column

表 3 图案分类的输入元件、输出元件以及隐藏层元件的RPU值

Table 3 RPU values of input element, output element and hidden layer element for pattern classification

Input (X_1)	Input (X_2)	Input (X_3)	Hidden layer ($P_1-Y_1-P_4$)	Hidden layer ($P_2-Y_3-P_6$)	Hidden layer ($P_3-Y_2-P_5$)	Output ($(P_5+P_6)-Y_4-P_7$)
0.0034	0.0013	0.0082	3.7999	3.6827	0.0710	0.0200
0.0034	0.0013	2.5	3.7999	0.0722	0.0710	1.6792
0.0034	4.4	0.0082	3.7999	3.6827	0.0701	0.0200
0.0034	4.4	2.5	3.7999	0.0722	0.0701	1.7026
2.8	0.0013	0.0082	0.1067	3.6827	2.0450	0.0200
2.8	0.0013	2.5	0.1067	0.0722	2.0450	0.0200
2.8	4.4	0.0082	0.1067	3.6827	0.0706	0.0200
2.8	4.4	2.5	0.1067	0.0722	0.0706	1.6888

网络的大小受到限制。其次,我们设计的网络不能在生物系统中自主进行学习,需要借助计算机选择合适的权重。最后,当前由基因线路构建的神经网络还需要进一步在生物实验中进行应用,以发挥它最大的潜力。值得注意的是,基因线路中存在天然噪声,且在实验中对输入的测量也可能因实验而产生差异。因此,不可能精确匹配使用实验中所使用的输入浓度。在此基础上,在仿真结果中对输入元件的RPU进行了一定范围的浮动,来模拟湿实验中不同的输入浓度。实验结果显示,我们设计的基因线路具有鲁棒性。

神经拟态的提出是使计算机实现对人类大脑的模拟,了解大脑的工作机制,最终使计算机或机器具备如人脑一样的近自然智能特性,从而使计算机或机器具备更低能耗、更加高效、容错率更高的能力。因此,我们使用基因线路实现神经网络,在生物学领域内实现了神经拟态计算。基因线路与神经网络结合的方式构造基于生物材料的神经拟态计算将会是迈向具有更高适应性、鲁棒性和扩展性的人工智能芯片的关键一步^[38]。总之,分子神经拟态计算的全面发展还有很长的路要走,这条路充满希望也充满挑战^[39]。

参考文献

- 1 Mead C. How we created neuromorphic engineering. *Nat Electron*, 2020, 3: 434–435
- 2 Kim Y, Zhang Y, Li P. Energy efficient approximate arithmetic for error resilient neuromorphic computing. *IEEE Trans VLSI Syst*, 2014, 23: 2733–2737
- 3 Li J, Liu R, Chen Y, et al. Computer-aided rational engineering of signal sensitivity of quorum sensing protein LuxR in a whole-cell biosensor. *Front Mol Biosci*, 2021, <https://doi.org/10.3389/fmolb.2021.729350>
- 4 Roy K, Jaiswal A, Panda P. Towards spike-based machine intelligence with neuromorphic computing. *Nature*, 2019, 575: 607–617
- 5 Pfeil T, Grübl A, Jeltsch S, et al. Six networks on a universal neuromorphic computing substrate. *Front Neurosci*, 2013, 7: 11
- 6 Esser S K, Merolla P A, Arthur J V, et al. Convolutional networks for fast, energy-efficient neuromorphic computing. *Proc Natl Acad Sci USA*, 2016, 113: 11441–11446
- 7 Shastri B J, Tait A N, Ferreira de Lima T, et al. Photonics for artificial intelligence and neuromorphic computing. *Nat Photonics*, 2021, 15: 102–114
- 8 Gupta M, Varma V, Damani S, et al. Compression of deep learning models for NLP. In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2020. New York: Association for Computing Machinery, 2020. 3507–3508
- 9 Zhang C J, Luo Q. Tropical cyclones objection detection based on faster R-CNN and infrared satellite cloud images. In: *Proceedings of the 2018 International Conference on Image and Video Processing, and Artificial Intelligence*. Shanghai: International Society for Optics and Photonics, 2018. 108360G
- 10 Marchisio M A. Parts & pools: A framework for modular design of synthetic gene circuits. *Front Bioeng Biotechnol*, 2014, 2: 42
- 11 George A K, Singh H. Enzyme-free scalable DNA digital design techniques: A review. *IEEE Transon Nanobiosci*, 2016, 15: 928–938
- 12 Green A A, Kim J, Ma D, et al. Complex cellular logic computation using ribocomputing devices. *Nature*, 2017, 548: 117–121
- 13 Siuti P, Yazbek J, Lu T K. Synthetic circuits integrating logic and memory in living cells. *Nat Biotechnol*, 2013, 31: 448–452
- 14 Qian L, Winfree E. Scaling up digital circuit computation with DNA strand displacement cascades. *Science*, 2011, 332: 1196–1201
- 15 Halvorsen K, Wong W P. Binary DNA nanostructures for data encryption. *PLoS One*, 2012, 7: e44212
- 16 Yazdi S H T, Gabrys R, Milenkovic O. Portable and error-free DNA-based data storage. *Sci Rep*, 2017, 7: 5011
- 17 Yazdi S H T, Yuan Y, Ma J, et al. A rewritable, random-access DNA-based storage system. *Sci Rep*, 2015, 5: 14138
- 18 Dean M E, Daffron C. A VLSI design for neuromorphic computing. In: *Proceedings of the 2016 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. New York: IEEE, 2016. 87–92
- 19 Organick L, Ang S D, Chen Y J, et al. Random access in large-scale DNA data storage. *Nat Biotechnol*, 2018, 36: 242–248
- 20 Li J Y, Yang S, Cui Y J, et al. Research progress of bacterial minimal genome (in Chinese). *Hereditas*, 2021, 43: 142–159 [李金玉, 杨珊, 崔玉军, 等. 细菌最小基因组研究进展. *遗传*, 2021, 43: 142–159]
- 21 Maher M A C, Deweerth S P, Mahowald M A, et al. Implementing neural architectures using analog VLSI circuits. *IEEE Trans Circuits Syst*, 1989, 36: 643–652
- 22 Stojanovic M N, Mitchell T E, Stefanovic D. Deoxyribozyme-based logic gates. *J Am Chem Soc*, 2002, 124: 3555–3561
- 23 Seelig G, Soloveichik D, Zhang D Y, et al. Enzyme-free nucleic acid logic circuits. *Science*, 2006, 314: 1585–1588
- 24 Komiya K, Sakamoto K, Kameda A, et al. DNA polymerase programmed with a hairpin DNA incorporates a multiple-instruction architecture into molecular computing. *Biosystems*, 2006, 83: 18–25
- 25 Yang J, Wu R, Li Y, et al. Entropy-driven DNA logic circuits regulated by DNAzyme. *Nucleic Acids Res*, 2018, 46: 8532–8541

- 26 Yang S, Li J Y, Cui Y J, et al. The current status and future prospects of DNA computing (in Chinese). *Chin J Biotechnol*, 2021, 37: 1120–1130 [杨 珊, 李金玉, 崔玉军, 等. DNA计算的发展现状及未来展望. *生物工程学报*, 2021, 37: 1120–1130]
- 27 Sarpeshkar R. Analog synthetic biology. *Philos Trans R Soc A-Math Phys Eng Sci*, 2014, 372: 20130110
- 28 Daniel R, Rubens J R, Sarpeshkar R, et al. Synthetic analog computation in living cells. *Nature*, 2013, 497: 619–623
- 29 Friedland A E, Lu T K, Wang X, et al. Synthetic gene networks that count. *Science*, 2009, 324: 1199–1202
- 30 Nielsen A A K, Der B S, Shin J, et al. Genetic circuit design automation. *Science*, 2016, 352: aac7341
- 31 Weiss J N. The Hill equation revisited: Uses and misuses. *FASEB J*, 1997, 11: 835–841
- 32 Cherry K M, Qian L. Scaling up molecular pattern recognition with DNA-based winner-take-all neural networks. *Nature*, 2018, 559: 370–376
- 33 Li X, Rizik L, Kravchik V, et al. Synthetic neural-like computing in microbial consortia for pattern recognition. *Nat Commun*, 2021, 12: 3139
- 34 Teng Y, Yang S, Liu R. Progress on neuromorphic computing based on biomolecules (in Chinese). *Chin Sci Bull*, 2021, 66: 3944–3951 [滕越, 杨 珊, 刘芮存. 基于生物分子的神经拟态计算研究进展. *科学通报*, 2021, 66: 3944–3951]
- 35 Didovyk A, Kanakov O I, Ivanchenko M V, et al. Distributed classifier based on genetically engineered bacterial cell cultures. *ACS Synth Biol*, 2015, 4: 72–82
- 36 Alnahhas R N, Sadeghpour M, Chen Y, et al. Majority sensing in synthetic microbial consortia. *Nat Commun*, 2020, 11: 3659
- 37 Miano A, Liao M J, Hasty J. Inducible cell-to-cell signaling for tunable dynamics in microbial communities. *Nat Commun*, 2020, 11: 1193
- 38 Teng Y, Yang S, Li J Y, et al. Principle and progress of DNA data storage (in Chinese). *Prog Biochem Biophys*, 2021, 48: 494–504 [滕越, 杨珊, 李 金玉, 等. DNA数据存储技术原理及其研究进展. *生物化学与生物物理进展*, 2021, 48: 494–504]
- 39 Marković D, Mizrahi A, Querlioz D, et al. Physics for neuromorphic computing. *Nat Rev Phys*, 2020, 2: 499–510

补充材料

图S1 基因线路的激活函数

图S2 一个简单的线路及其对应的神经元的计算过程

图S3 利用基因线路构建神经网络实现AND线性分类的过程

图S4 利用基因线路构建神经网络实现AND线性分类的仿真结果

表S1 构建基因线路的生物元件库

表S2 线性分类AND的输入元件、输出元件以及隐藏层元件的RPU值

表S3 在实现神经网络的基因线路中所使用的生物元件

本文以上补充材料见网络版csb.scichina.com. 补充材料为作者提供的原始数据, 作者对其学术质量和内容负责.

Summary for “利用基因线路构建神经网络实现神经拟态计算”

Constructing artificial neural networks using genetic circuits to realize neuromorphic computing

Shan Yang[†], Ruicun Liu[†], Tuoyu Liu, Yingtan Zhuang, Jinyu Li & Yue Teng^{*}

State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Academy of Military Medical Sciences, Beijing 100071, China

[†] Equally contributed to this work

^{*} Corresponding author, E-mail: yueteng@me.com

With the advent of the era of big data, existing computing systems limit the development of new technologies such as cloud computing and artificial intelligence. As the demand for high-performance computing continues to grow, traditional computing models are facing unprecedented challenges. Neural mimicry computing based on artificial neural networks provides a potential solution, and biological computing with advantages such as low energy consumption parallelization is very important for its research, in which gene circuitry will be the key to the construction of artificial neural networks. In this study, we used biological elements to construct gene circuits to form neural network structures to realize neural mimicry calculations. Based on the similarity between the structure of gene circuits and neural networks, we describe a neural network design based on the logic gate calculation of gene circuits, and realize the applications of linear classification, nonlinear classification and pattern classification. In the linear classification of NAND, the four datasets in the space are divided into two categories by hyperplane. The structural design of the neural network for linear classification can be realized by using the genetic circuit, which includes five biological elements and the biological elements between different layers are connected by gene regulation. The simulation results of linear classification showed that the gene circuit realizes the function of NAND by means of the neural network and the output fluorescence value as the detection standard. In the nonlinear classification of XOR, the four datasets in the space are divided into two categories. The structural design of the neural network for nonlinear classification can be realized by using the genetic circuit, which includes six biological elements. The simulation results of nonlinear classification showed that the gene circuit realizes the function of XOR by means of the neural network and the output fluorescence value as the detection standard. In the pattern classification, the neural network with five layers and eight neurons was constructed by a genetic circuit with eight biological elements connected by gene regulation. The simulation results of pattern classification showed that the gene circuit realizes the function of pattern classification by means of the neural network and the output fluorescence value as the detection standard. This study realizes neural mimicry computing by constructing neural network through engineering gene circuits. The integrated molecular computing system is expected to be applied to the manufacturing of artificial intelligence chips, and further applied to aerospace, information security and national defense construction, along with other fields.

neuromorphic computing, artificial neural network, DNA computing, genetic circuit, synthetic biology

doi: [10.1360/TB-2021-0501](https://doi.org/10.1360/TB-2021-0501)