

КЛАССИФИКАЦИЯ И МОДЕЛИРОВАНИЕ ГЕНЕТИЧЕСКОГО КОДА И ГЕННО-НЕЙРОННЫХ СЕТЕЙ

Адиль Тимофеев

Аннотация: Предлагаются методы автоматической классификации и моделирования генетического кода. Излагаются принципы проектирования и результаты использования гетерогенных генно-нейронных сетей.

Ключевые слова: модели генетического кода, генетическая информатика, гетерогенная генно-нейронная сеть.

ACM Classification Keywords: E.4 Coding and Information Theory

Conference: The paper is selected from International Conference "Classification, Forecasting, Data Mining" CFDM 2009, Varna, Bulgaria, June-July 2009

Введение

Основным "строительным материалом" живых организмов являются белки, включающие в себя 20 основных аминокислот. При биохимическом синтезе белков организма используется генетическая информация, закодированная в главном "наследственном материале" – дезоксирибонуклеиновой кислоте (ДНК) [1].

В 1953 г. Дж.Уотт и Ф.Крик описали структуру ДНК и высказали гипотезу о генетическом коде и механизме самовоспроизведения ДНК [2]. За это открытие авторы были удостоены Нобелевской премии.

ДНК является полимером и представляет собой цепочки мономеров определенных типов, образующих "двойную спираль" [1–3]. В состав ДНК может входить только четыре типа оснований: аденин (А), тимин (Т), гуанин (G), цитозин (С). Цепи оснований ДНК всегда соединены по принципу комплементарности (взаимодополнительности): А связано с Т, а G – с С. Таким образом, водородные связи между основаниями А и Т, G и С определяются "правилом комплементарности" [1–3].

Комплементарность оснований в двух цепях ДНК создает основу для репликации, т.е. самовоспроизведения ДНК. Действие этого механизма проявляется в раскручивании "двойной спирали" ДНК, после чего в точках разветвления пристраиваются соответствующие новые основания. В результате ДНК самоудваивается.

Наряду с процессом репликации ДНК происходит процесс транскрипции, т.е. перенос генетической информации с ДНК на рибонуклеиновую кислоту (РНК). Основное отличие "информационной" РНК от порождающей ее ДНК заключается в том, что вместо основания Т включается основание U. В процессе транскрипции происходит "перекодировка" информации с преобразованием оснований $T \rightarrow U$.

Таким образом, транскрипция обеспечивает перенос генетической информации с ДНК на РНК. Размеры "информационной" РНК невелики по сравнению с размерами "родительской" ДНК [1,3].

1. Генетический язык: алфавиты, слова и семантика

Три рядом стоящих основания в ДНК соответствует только одной аминокислоте. Последовательность из трех оснований называется триплетом или кодоном. Поэтому любая цепь ДНК является последовательностью кодонов, начинающейся с определенного "стартового" участка.

Линейное расположение оснований в ДНК позволяет ввести простой "генетический язык" для кодирования и описания "наследственного материала". Алфавитом этого языка является следующий набор букв

$$\alpha_D = \{A, C, G, T\}. \quad (1)$$

Будем называть этот набор букв (1) алфавитом оснований ДНК. Словами в этом алфавите являются «осмысленные» последовательности букв. Такими словами служат записи кодонов – упорядоченных троек (триплетов) из оснований ДНК, кодирующих некоторую аминокислоту. Очевидно, что в рассматриваемом алфавите (1) можно составить $4^3 = 64$ различных комбинаций слов из трех букв. Полезно также ввести «стартовый» участок цепей ДНК и «стоп-кордоны», обозначающие конец цепи.

Процесс записи последовательности слов, соответствующих рассматриваемой цепи ДНК, целесообразно начать с «пустого слова» (не пишется ничего), обозначающего «начало отсчета», затем к нему справа приписывается первая буква, к ней приписывается вторая буква и т.д. до конца цепи, обозначенного одним из «стоп-кодонов». При этом не используются никакие «знаки препинания». В результате получается предложение вида

$$AGTCCATGGTAC \quad (2)$$

Каждому предложению, описывающему одну цепь ДНК, однозначно соответствует комплементарное (взаимодополняющее) предложение, описывающее другую цепь ДНК. Например, для предложения (2) оно имеет вид

$$TCAGGTACCATG \quad (3)$$

Генетическая информация, содержащаяся в кодонах ДНК, сначала «переписывается» в соответствующие кодоны «информационной» РНК. Эта РНК синтезируется в процессе транскрипции. В результате этого кодоны ДНК преобразуются в кодоны РНК. Алфавит оснований ДНК (2) порождает алфавит оснований РНК вида

$$\alpha_R = \{A, C, G, U\}. \quad (4)$$

Фрагменту цепи ДНК (2), записанной в алфавите (1), соответствует следующее описание синтезированной РНК

$$AGUCCUGGUAC \dots \quad (5)$$

записанная в новом алфавите (4).

Слова в алфавитах (1) или (4), т.е. кодоны ДНК и РНК, могут быть графически равными (если они составлены из одинаковых букв, расположенных одинаковым образом) или графически различными (в противном случае). В каждом слове содержится информация о соответствующей аминокислоте, а в каждом предложении - информация о типе и последовательности аминокислот, синтезированных с помощью РНК.

Таким образом, каждое предложение в алфавитах оснований (1) или (4) содержит генетическую информацию, определяющую специфику организма с данной ДНК, т.е. его «унаследованную индивидуальность».

Белки организмов обычно состоят из 20 типов аминокислот. Поэтому любой белок можно закодировать словами, состоящими из последовательности букв алфавита аминокислот вида

$$\alpha_a = \{a_1, a_2, \dots, a_{20}\}. \quad (6)$$

Буквы этого алфавита обозначают следующие аминокислоты:

a_1 – фенилаланин (Phe), a_2 – лейцин (Leu), a_3 – изолейцин (Ile), a_4 – метионин (Met), a_5 – валин (Val), a_6 – серин (Ser), a_7 – пролин (Pro), a_8 – треонин (Thr), a_9 – аланин (Ala), a_{10} – тирозин (Tyr), a_{11} – гистин (His), a_{12} – глютамин (Gln), a_{13} – аспарагин (Asn), a_{14} – лицин (Lys), a_{15} – аспарагиновая

кислота (Asp), ^{a16} – глутаминовая кислота (Glu), ^{a17} – цистеин (Cys), ^{a18} – триптофан (Trp), ^{a19} – аргинин (Arg), ^{a20} – глицин (Gly).

2. 3D-геометрическая и 2D-графовая модели генетического кода

В результате сложных биохимических исследований был установлен генетический код, т.е. соответствие между алфавитом аминокислот (6) и алфавитом оснований "информационной" РНК (4) [2,3]. Он состоит из 61 кодона, соответствующих 20 аминокислотам. Генетический код является вырожденным в том смысле, что одному типу аминокислоты может соответствовать несколько слов - синонимов (кодонов) в алфавите оснований РНК (4).

Наряду с классической табличной моделью генетического кода, полезна его трехмерная геометрическая модель типа «гиперкуб», предложенная автором в [4]. Каждому узлу этой 3D-модели соответствует аминокислота с соответствующим номером, а его проекции определяют кодон генетического кода.

Весьма удобной и полезной является также предложенная автором в [4] графовая модель представления генетического кода, Это новое 2D-представление генетического кода в виде графа (кодирующего дерева) имеет ряд общих черт с известной генетике "круговой диаграммой", описанной в [3].

Табличную, 3D-геометрическую и 2D-графовую модели генетического кода можно одинаково успешно применять для расшифровки ДНК и РНК растений, животных и человека.

3. Передача генетических сообщений

Рассмотрим алфавит $\alpha_X = \{x_1, \dots, x_4\}$, буквы которого совпадают с буквами алфавита оснований ДНК (1). Словом в этом алфавите будем называть последовательность из трех букв вида

$$X = x_1x_2x_3, \quad x_{ij} \in \alpha_X. \quad (7)$$

Обозначим через $S(\alpha_X)$ множество всех триплетных слов вида (7), а через $S'(\alpha_X)$ - подмножество слов из $S(\alpha_X)$, имеющих "генетический смысл", т.е. обозначающих соответствующие аминокислоты.

Объект, порождающий "осмысленные" слова из S' , называется в теории кодирования источником сообщения, а слова из S' - сообщениями. В роли источника сообщения в генетике выступает ДНК. Описание дополнительной информации о ДНК как источнике сообщений может задаваться различными способами:

1) теоретико-множественное описание мощности, т.е. числа элементов, и других характеристик множеств α_X , $S(\alpha_X)$, $S'(\alpha_X)$. Для ДНК мощность этих множеств определяется соотношениями

$$|\alpha_X| = 4, \quad |S| = 4^3 = 64, \quad |S'| = 61;$$

2) статистическое (частотное) описание осуществляется заданием вероятностей (частот) характеристик α_X , $S(\alpha_X)$, $S'(\alpha_X)$. Для ДНК могут быть известны, например, вероятности (частоты) появления букв $p_1 = p(A)$, $p_2 = p(C)$, $p_3 = p(G)$, $p_4 = p(T)$, соответствующих основаниям из алфавита (1);

3) логическое описание множеств с помощью языка исчисления двужначных или многозначных предикатов.

Пусть задан также алфавит $\alpha_Y = \{y_1, \dots, y_4\}$, буквы которого совпадают с буквами алфавита оснований РНК (4). Через Y обозначим триплетное слово в алфавите α_Y , а через $S(\alpha_Y)$ - множество всех слов в этом алфавите.

Генетическое преобразование (1) при транскрипции задает отображение F , которое каждому слову $X \in S'(\alpha_X)$, т.е. кодону ДНК, однозначно ставит в соответствие слово

$$Y = F(X) = y_1y_2y_3, \quad Y \in S(\alpha_Y), \quad (8)$$

являющееся кодоном синтезированной “информационной” РНК. Слово (8) будем называть кодом сообщения X при синтезе РНК, а переход от слова X к слову Y кодированием ДНК в структуре РНК. Этот переход, происходящий в процессе транскрипции, можно интерпретировать, как передачу наследственной информации из “постоянной” памяти ДНК в “оперативную” память РНК.

Код РНК- сообщения Y подается в “рибосомный” канал связи и синтеза белков. Однако код сообщения Y' на выходе канала связи может отличаться от входного кода Y . Источником искажения могут быть мутации генов, ошибки “считывания” кода и т.п.

В случае идеального канала связи передача генетической информации не искажается, т.е. $Y' = F(X) = Y$.

Поэтому возможно точное декодирование “генетического сообщения”, если существует обратное отображение F^{-1} для (8). В случае, когда генетическая информация искажается в канале связи и при синтезе белков, включается система “репарации” (коррекции), обеспечивающая обнаружение и исправление ошибок.

Различные слова (8), являющиеся кодонами РНК, можно закодировать различными буквами a_1, a_2, \dots, a_{20} алфавита аминокислот (6). Этот натуральный генетический код является вырожденным, поскольку он не удовлетворяет требованию взаимной однозначности. Однако его избыточность значительно повышает надежность передачи генетической информации.

4. Мера и оценка генетической информации

В генетике важную роль играют дискретные формы кодирования, хранения и передачи “наследственной информации”. Поэтому естественно определить “количество” генетической информации в терминах двоичных знаков, т.е. в битах. При этом целесообразно исходить из комбинаторного подхода к теории информации, предложенного А.Н. Колмогоровым [5]. Этот подход обобщает вероятностный подход, развитый К.Шенноном.

Обозначим основание ДНК или РНК переменной x . Эта переменная может принимать значения, принадлежащие конечным алфавитам оснований ДНК (1) или РНК (4), которые состоят из четырех элементов. Поэтому “энтропия” H основания x в ДНК или РНК равна

$$H(x) = \log 4 = 2. \quad (9)$$

В каждой ДНК или РНК основание x имеет определенное значение (например, $x = A$). Это означает, что каждое основание в цепи ДНК или РНК сообщает генетическую информацию, равную

$$I_x = H(x) = \log 4, \quad (10)$$

и требует для своего описания 2 двоичных знака. При этом “снимается” энтропия, т.е. априорная неопределенность этих знаков.

Аналогично обозначим через Y произвольный кодон ДНК или РНК. Число различных кодонов, которые можно формально образовать в алфавите оснований ДНК или РНК, равно $M = 4^3 = 64$. Однако в природном генетическом коде обычно содержится $M_\Gamma = 61$ кодонов. Поэтому “количество генетической информации”, содержащейся в определенном кодоне Y ДНК или РНК, равно

$$I_Y = \log 61 < \log 64. \quad (11)$$

Для записи любого кодона Y в двоичной системе требуется 6 двоичных знаков.

ДНК и РНК конкретных организмов имеет определенную длину L , равную числу оснований в цепи. Различные основания встречаются в этой цепи с различной частотой. Обозначим через n_1, n_2, n_3, n_4 число вхождений соответствующих оснований (например, U, C, A, G в алфавите оснований РНК) в цепь

длины L . Тогда, учитывая, что $L = n_1 + n_2 + n_3 + n_4$, легко подсчитать общее число возможных цепей длины L по формуле

$$R = \frac{L!}{n_1!n_2!n_3!n_4!} . \quad (12)$$

Количество генетической информации в цепи ДНК или РНК длины L , закодированной в соответствующем алфавите оснований, будет равно

$$I_L = \log R . \quad (13)$$

При больших длинах L , характерных для ДНК и РНК, при вычислении (14) можно воспользоваться формулой Стирлинга:

$$\log(L!) \sim L \log L .$$

Тогда получим следующую приближенную формулу

$$I_L \sim -L \sum_{i=1}^4 p_i \log p_i , \quad (p_i = \frac{n_i}{L}) \quad (14)$$

Отсюда следует, что, если в цепи ДНК или РНК основания встречаются с частотами p_i , то количество генетической информации, приходящейся на одно основание, равно

$$H = - \sum_{i=1}^4 p_i \log p_i \quad (15)$$

В случае равных частот $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$ из (15) вновь получим формулу (9). При любых других соотношениях частот встречаемости оснований в цепи ДНК или РНК справедливо неравенство

$$H < \log 4 .$$

Следовательно, для передачи "генетического сообщения" длины L достаточно употребить примерно $L H$ двоичных знаков, не превышающее $2L$.

Интересно также оценить количество "генетической информации", содержащейся в переменной a из алфавита аминокислот α_a относительно связанных с ней кодонов Y генетического кода. Связь между переменными a и Y заключается в том, что генетический код допускает не все формально возможные пары (a, Y) . Однако для любой аминокислоты $a \in \alpha_a$ можно найти все кодоны, допускаемые генетическим кодом.

Тогда генетическую информацию в a относительно Y можно определить по формуле

$$I_a = H(Y) - H(Y/a) , \quad (16)$$

где $H(Y/a) = \log M_a$, M_a - число кодонов генетического кода для a . Например, если $a = a_2$, то $M_a = 6$, и, следовательно, $I_a = \log 61 - \log 6$. Если же $a = a_4$, то $M_a = 1$ и $I_a = \log 61$.

5. Информационная сложность РНК

Генетическая информация тесно связана со "сложностью" ее носителя. Если этот носитель устроен "просто" (например, кодон), то для его описания достаточно небольшого количества информации. Для сложных "носителей" (например, для РНК) требуется много информации для его описания и передачи.

Стандартным способом описания "информационной" РНК является последовательность оснований Z в четырехбуквенном алфавите оснований (5). Поставим в соответствие рассматриваемой РНК некоторое

число $n = f(Z)$. Например, это может быть десятичное число, определенное по двоичному представлению Z . Обозначим через $l(Z)$ наименьшую длину цепочки оснований, определяющей данную (неизбыточную) РНК.

“Сложностью” РНК при способе ее задания с помощью f будем называть величину

$$K_f(\omega) = \min l(Z) \text{ при } f(Z) = n(\omega). \quad (17)$$

На генетическом языке это определение “сложности” РНК можно проинтерпретировать следующим образом. Конкретная цепочка оснований Z определяет “генетическую программу” синтеза белков, а оператор f – способ задания РНК. Тогда естественно считать, что $K_f(\omega)$ есть наименьшая длина “генетической программы”, с помощью которой можно синтезировать объект ω при способе задания f .

Задание какого-либо носителя “генетической информации” (например, РНК) можно упростить, если уже задан какой-то другой объект (например, кодон). Для этого введем показатель “условной сложности” объекта ω при заданном объекте Z . Следуя А.Н. Колмогорову [5], определим этот показатель в виде

$$K_f(\omega/Z) = \min l(Z) \text{ при } f(n(\omega), Z) = n(\omega). \quad (18)$$

Здесь способ задания f является функцией от номера объекта Z и номера “генетической программы” вычисления $n(\omega)$ при заданном объекте Z .

Если “условная сложность” значительно меньше, чем безусловная, т.е.

$$K_f(\omega/Z) \ll K_f(\omega),$$

то естественно считать, что в объекте Z содержится значительная “генетическая информация” об объекте ω . Количество этой условной информации зададим формулой

$$I_f(\omega/Z) = K_f(\omega) - K_f(\omega/Z). \quad (19)$$

В частном случае, когда $K_f(\omega/\omega) = 0$, получаем $I_f(\omega/\omega) = K_f(\omega)$.

В этом случае “информационная сложность” объекта ω совпадает с его “генетической информацией” о себе самом.

Важными достоинствами предложенных определений меры “генетической информации” и “информационной сложности”, является то, что они относятся к индивидуальным объектам, т.е. к конкретным кодонам, хромосомам, РНК и т.п. Однако их можно с одинаковым успехом использовать и в тех случаях, когда заданы вероятностные или частотные характеристики рассматриваемых объектов.

6. Генетические базы данных и знаний

Каждый ген, управляя синтезом белка, определяет некоторый элементарный признак организма. Множество признаков, характеризующих различные виды организмов, удобно представить в виде реляционной базы данных (БД) табличного типа. При формировании генетической БД каждому признаку ставится в соответствие “домен”, т.е. множество дискретных значений признака. Ген, порождающий признак, может находиться в одном из возможных альтернативных состояний, определяемых аллелями. Например, у каждого кролика имеется ген, определяющий признак окраса его меха. Принято подразделять окрасы на “шиншиловый”, “дикий тип”, “альбинос” и “гималайский”, что соответствует четырем аллелям.

Сложные признаки определяются хромосомой, состоящей из набора генов x_1, \dots, x_n . Число хромосом у каждого вида организмов фиксировано и равно $2n$, где n - гаплоидное число, являющееся инвариантом данного вида. Например, у человека $n=23$, а у краба $n=127$. Поэтому в генетическую БД человека включается 23 отношения, а в БД краба - 127 отношений.

Процессу мейоза в генетической БД соответствует процесс соединения всех отношений, т.е. образуется их прямое произведение.

Генетическая БД является хранилищем индивидуальной информации, передаваемой от родителей к потомкам. Однако эта информация допускает обобщенное представление в виде "генетических знаний".

В отличие от "индивидуальных данных", закодированных в ДНК и РНК в алфавите оснований или аминокислот в форме длинных последовательностей слов (предложений), "обобщенные" знания представляют собой "высказывания" в терминах многозначных предикатов, которые являются истинными по крайней мере на всех "предложениях" БД. Совокупность этих "высказываний" образует генетическую базу знаний (БЗ).

Для автоматического синтеза БЗ по заданной генетической БД и минимизации ее сложности (без потери "генетической информации") можно использовать логико-аксиоматический и логико-вероятностный методы синтеза решающих правил, предложенные автором в [6-8]. Совокупность этих правил ("генетических высказываний") записывается в терминах логических или многозначных предикатов, связанных с алфавитом оснований или аминокислот, и обладает необходимыми свойствами полноты и непротиворечивости при описании генетической БД.

7. Когнитивные модели генетического кода и генетические алгоритмы

Применение логико-вероятностного метода оптимального синтеза генетических БЗ к генетической БД, представляющую собой классическую табличную модель генетического кода [2], позволяет автоматически построить когнитивную модель генетического кода [8,11]. Эта модель в виде классифицирующего дерева аминокислот минимальной сложности представлена автором в [4]. Каждый путь на этом дереве с вероятностью 1 описывает соответствующую аминокислоту в виде логического "высказывания" определённого типа.

Примером могут служить следующие "генетические правила" классификации вида

- 1) ЕСЛИ 2 основание = А И 1 основание = С И 3 основание = А
ИЛИ 2 основание = А И 1 основание = С И 3 основание = G
ТО аминокислота a_{12} [Gln]
- 2) ЕСЛИ 2 основание = U И 1 основание = А И 3 основание = G
ТО аминокислота a_4 [Met].

Методы математического моделирования и вычислительного эксперимента играют важную роль в генетических исследованиях. Они позволяют формализовать генетические механизмы в виде математических и информационных моделей, генетических БД и БЗ и т.п. Учет биологических принципов обработки информации позволяет создавать генетические алгоритмы и развивать теорию клеточных автоматов, нейронных сетей и т.п.

В последние годы сформировались новые разделы генетики - математическая генетика и генетическое программирование [6,7]. В их основе лежит оригинальный математический аппарат и программное обеспечение. Этот новый инструмент ориентирован не только на собственно генетические исследования, но и на решение широкого класса задач дискретной оптимизации, эволюционного моделирования и т.п.

Сегодня генетические алгоритмы успешно используются для оптимизации расписаний, планирования поведения, оптимальной трассировки компьютерных плат, автоматического управления нелинейными процессами и т.п. [6,7]. Они особенно эффективны в многоэкстремальных задачах, связанных с поиском глобального экстремума. Весьма перспективно использование таких алгоритмов для управления генетическими БД и БЗ и обучения нейронных сетей на основе принципов самоорганизации и естественного отбора наилучших архитектур [8-13].

Отличительными чертами генетических алгоритмов является их разветвленность и параллелизм, связанные с использованием "вычислительных популяций", целенаправленная "селекция" с

“наследованием” наиболее важных признаков или фрагментов промежуточных результатов, многовариантное сравнение, “естественный отбор” наилучших решений и т.п. В этих алгоритмах используются принципиально новые вычислительные операторы. Примерами таких операторов, не имеющих аналогов в традиционных вычислительных моделях, являются нелинейные преобразования типа “мутации”, “инверсии” и “кроссингвера” [6,7].

8. Модели гетерогенных генно-нейронных сетей

Нейронные сети (НС) и нейросетевые технологии являются одним из наиболее эффективных средств массового распараллеливания и ускорения процессов обработки знаний и передачи потоков данных в задачах распознавания образов, классификации данных и диагностики состояний. Естественным прототипом искусственных НС является биологический мозг и центральная нервная система человека и животных как сложная гетерогенная нейронная сеть, обеспечивающая за счёт естественных био- и нанотехнологий высокую степень параллелизма, адаптации, самоорганизации и робастности при решении различных интеллектуальных задач (представление знаний, распознавание образов, классификация данных, поиск закономерностей, анализ изображений, диагностика состояний, прогнозирование явлений и т.п.). Возможности искусственных и биологических НС могут значительно расшириться при коллективном (мульти-агентном) решении сложных интеллектуальных задач.

Высокая сложность и размерность многих задач распознавания образов, классификации данных, анализа изображений и диагностики состояний, а также часто возникающая необходимость их решения в реальном времени требуют массового параллелизма и самоорганизации распределённых вычислений на базе НС. С этой точки зрения особый интерес и дополнительные возможности представляют гетерогенные полиномиальные нейронные сети (ПНС) с самоорганизующейся архитектурой и генно-нейронные сети (ГНС) [8-13].

Основные идеи, математические модели, методы оптимизации, алгоритмы обучения и принципы самоорганизации ПНС и ГНС были предложены автором в работах [8-13]. Они заключаются в следующем:

- архитектура НС гетерогенна и многослойна;
- наличие слоя полиномиальных нейронных элементов (П-нейронов);
- возможность обучения и адаптации НС к обучающим базам данных (БД);
- целесообразность самоорганизации и минимизации сложности архитектуры НС различных типов в процессе обучения;
- детерминированные, логические и вероятностные методы обучения и самоорганизации гетерогенных НС с самоорганизующейся архитектурой;
- принцип высокой экстраполяции (экстраполирующей силы) гетерогенных НС ;
- алгебраическое требование диофантовости (целочисленности синаптических весов) гетерогенных НС.

В процессе дальнейшего развития теории гетерогенных ПНС и ГНС были предложены модели многозначных нейронных элементов (М-нейронов) и связанных с ними конъюнктивных, полиномиальных, дизъюнктивных и суммирующих нейронных элементов (МК-, МП-, МД- и МΣ-нейронов), а также новые разновидности гетерогенных ПНС (генно-нейронные сети, квантовые нейронные сети, мульти-агентные ПНС и т.п.).

Предложенные гетерогенные модели и быстрые алгоритмы обучения ПНС и ГНС разных типов обеспечивают высокий параллелизм и самоорганизацию нейровычислений в процессе решения многих интеллектуальных задач. Они успешно применялись для решения ряда прикладных задач распознавания образов (распознавание кораблей по отраженным радиолокационным сигналам, распознавание команд и дикторов по видеogramмам речи, распознавание и адресация деталей на конвейере, классификация дорожных ситуаций и т.д.), медицинской диагностики (диагностика и оценка эффективности лечения

артритов, векторная диагностика и расшифровка гастритов и т.д.), прогнозирования явлений (прогнозирование градоопасности облаков и исхода черепно-мозговых травм и т.д.) и нейросетевого представления генетического кода [4,8–13].

Заключение

Бурное развитие генетики и теории биологической эволюции привело к созданию новых научных направлений, связанных с разработкой “генетических алгоритмов”, “генетического программирования”, “эволюционного моделирования” и “генной инженерии”. Генетические принципы и механизмы породили новые подходы в теории кодирования и передачи информации, теории алгоритмов и теории автоматов. Они оказали глубокое влияние на компьютерную информатику и программирование.

Значительный интерес представляет использование принципов генетики и нейрофизиологии в теории нейронных сетей и нейрокомпьютеров, а также моделирование генно-нейронных сетей и их реализация на базе нанотехнологий.

Благодарности

Работа выполнена при частичной поддержке **грантов РФФИ № 08–08–12183-офи и № 09–08–00767-а** и **Программы № 1** Президиума РАН.

Литература

- [1]. Айала Ф., Кайгер Дж. Современная генетика (М.: Мир, 1968).
- [2]. Уотсон Дж. Молекулярная биология (М.: Мир 1979).
- [3]. Инге-Вечтомов С.Г. Генетика с основами селекции (М.: Высшая школа, 1988)
- [4]. Тимофеев А.В. Генетическая информация и национальный генотип. – В книге: В поисках парадигмы нации (нацио-логические очерки). Очерк 7. Москва - Нальчик, Изд.-во АМАН, 1997, с. 188-223.
- [5]. Колмогоров А.Н. Теория информации и теория алгоритмов (М.: Наука, 1987)
- [6]. Goldberg D.E. Genetic Algorithms in Search, Optimization and Machine Learning (Addison - Wesley, 1989).
- [7]. Koza J.R. Genetic Programming (Bradford/MIT Press, 1992).
- [8]. Тимофеев А.В. Адаптивные робототехнические комплексы (Л.: Машиностроение, 1988).
- [9]. Каляев А.В., Тимофеев А.В. Методы обучения и минимизации сложности когнитивных нейромодулей супермакро-нейрокомпьютера с программируемой архитектурой. - Доклады АН, 1994, т.337, №2, с.180-183.
- [10]. Тимофеев А.В. Методы синтеза диофантовых нейронных сетей минимальной сложности. - Доклады АН, 1995, т.337, № 1, с.32-35.
- [11]. Timofeev A.V. Intelligent Control Applied to Non-Linear Systems and Neural Networks with Adaptive Architecture. - Journal of Intelligent Control, Neurocomputing and Fuzzy Logic, 1996, v.1, № 1, pp.1-18.
- [12]. Тимофеев А.В. Оптимальный синтез и минимизация сложности генно-нейронных сетей по генетическим базам данных — Нейрокомпьютеры: разработка и применение, 2002, № 5-6, с. 34-39.
- [13]. Timofeev A. V. Parallel Structures and Self-Organization of Heterogeneous Polynomial Neural Networks for Pattern Recognition and Diagnostics of States. – Pattern Recognition and Image Analysis, 2007, Vol. 17, No. 1, pp. 163–169.

Информация об авторе

Тимофеев Адиль Васильевич – заведующий лабораторией информационных технологий в управлении и робототехнике Санкт-Петербургского института информатики и автоматизации Российской академии наук, доктор технических наук, профессор, Заслуженный деятель науки РФ, 199178, Россия, Санкт-Петербург, 14-я линия, д. 39, СПИИРАН, tav@ias.spb.su