

3

Оценка скорости и эффективности эволюционных процессов*

В. Г. Редько

Насколько эффективны эволюционные процессы? Для того чтобы ответить на этот вопрос, целесообразно оценить скорость эволюции в четких опорных моделях. В настоящей работе излагаются результаты оценок скорости и эффективности эволюции для такой опорной модели, как модель квазивидов, и для нескольких близких к ней моделей эволюции. Модель квазивидов предложил Манфред Эйген в 1970-х гг. В этой модели предполагается существование оптимальной особи, а эволюция популяции особей в ней приводит к отбору квазивида – такого распределения особей, в которое входит как оптимальная особь, так и близкие к ней, чьи геномы отличаются от генома оптимальной особи мутационными заменами. Полученные аналитические оценки показывают, что при эволюционном поиске оптимальный геном длины N может быть найден в течение N поколений, при этом в эволюционном процессе участвуют порядка N^2 особей. Близкие оценки справедливы и для других моделей, сходных с моделью квазивидов. Аналитические оценки проверены путем компьютерного моделирования.

***Ключевые слова:** эволюция, геном, модель квазивидов, рулеточный отбор, нейтральный отбор, мутации, эволюционная модель.*

1. Введение

Оценки скорости сходимости и эффективности эволюционных процессов целесообразно провести для четких опорных примеров. В данной работе мы отталкиваемся от модели квазивидов, введенной лауреатом Нобелевской премии М. Эйгеном при изучении процессов предбиологической эволюции (Эйген 1973). Данная статья подводит итог цикла работ (Редько 1986; 1990; 2005; Редько, Цой 2005; Red'ko, Tsoy 2006), посвященных исследованию скорости сходимости простых моделей эволюции.

Оценки сделаны для метода рулеточного отбора, что позволяет использовать их при интерпретации биологических эволюционных процес-

* Работа выполнена при финансовой поддержке РФФИ (проект № 13-01-00399).

сов. Рулеточный отбор вероятностный, при таком отборе важно учитывать роль нейтрального отбора, то есть возможность случайной фиксации особей, не зависящей от их приспособленности. Учет нейтрального отбора подчеркивает важность рассмотрения численности популяции: при малой численности вероятность случайной фиксации особей возрастает. Для того чтобы нейтральный отбор не сказывался, необходимо, чтобы численность популяции была достаточно большой, и это явно учитывается в сделанных оценках.

В основной рассматриваемой модели эволюционного процесса (то есть для простого варианта модели квазивидов) предполагается, что имеется единственный максимум приспособленности. Также анализируются другие модели, близкие к основной: модель узкого канала, мажорирующая модель, для которых скорость эволюции постоянна, и модель эволюционного поиска минимумов энергии спинового стекла, для которой число локальных максимумов приспособленности экспоненциально растет с ростом длины генома.

Статья организована следующим образом. В разделе 2 изложена модель квазивидов. Раздел 3 характеризует качественную картину эволюции в этой модели. Роль нейтрального отбора в популяциях конечной численности представлена в разделе 4. Аналитические оценки скорости и эффективности эволюции в модели квазивидов сделаны в разделе 5, здесь же резюмируются основные принципы, которые заложены в метод получения оценок. Результаты компьютерных расчетов в рамках модели квазивидов приведены в разделе 6, компьютерные расчеты подтверждают полученные аналитические оценки. Модель узкого канала и близкая к ней мажорирующая модель излагаются в разделе 7. Модель эволюционного поиска минимумов энергии спинового стекла характеризуется в разделе 8. Раздел 9 заключает статью.

2. Модель квазивидов

Модель квазивидов была введена М. Эйгеном при изучении процессов предбиологической эволюции (Эйген 1973; Эйген, Шустер 1982). Содержательно эта модель описывает возможный процесс того, как могли бы эволюционно возникнуть простейшие макромолекулы, кодирующие наследственную информацию. В ней рассматривается эволюция популяции цепочек РНК. При наличии притока энергии в виде богатых энергией молекул типа АТФ цепочки РНК могут реплицироваться, то есть размножаться путем саморепликации или репликации с помощью простейших ферментных систем. При размножении происходит копирование наследуемой информации. Ошибки в процессе копирования приводят к мутациям РНК. Эволюция популяции приводит к отбору квазивида – такого распределения цепочек РНК, в которое входит как «наилучшая РНК» (размножающаяся с максимальной скоростью), так и близкие к ней цепочки, отличающиеся от этой наилучшей мутационными заменами.

Опишем формально модель квазивидов. Рассматриваем дарвиновскую эволюцию популяции модельных особей. Геномы особей представляют собой информационные последовательности S_k , для основного варианта модели будем считать, что символы последовательностей S_{ki} принимают два значения: $S_{ki} = +1$ либо $S_{ki} = -1$; k – номер особи; i – номер символа; $i = 1, 2, \dots, N$; $k = 1, 2, \dots, n$; N – длина последовательностей; n – численность популяции. Эволюция включает отбор особей S в соответствии с их приспособленностями $f(S)$ и мутации – равновероятные замены символов S_{ki} . Считаем, что N и n в процессе отдельного эволюционного процесса не меняются, они велики: $N, n \gg 1$. Для основного варианта модели считаем, что функция приспособленности $f(S)$ простая: имеется оптимальная особь S_m с максимальной приспособленностью, а приспособленность произвольной особи S экспоненциально уменьшается с ростом расстояния по Хеммингу (числа несовпадающих символов в соответствующих позициях последовательностей) $\rho(S, S_m)$ между S и S_m :

$$f(S) = \exp[-\beta\rho(S, S_m)], \quad (1)$$

где β – параметр интенсивности отбора. Это означает, что чем ближе геном рассматриваемой особи S к геному оптимальной особи S_m , тем больше приспособленность особи S . Понятие расстояния по Хеммингу иллюстрируется приведенными ниже примерами:

$$S_m = (-1, 1, 1, -1, 1), \quad S_1 = (-1, -1, 1, 1, -1), \quad \rho(S_1, S_m) = 3.$$

$$S_2 = (1, 1, 1, 1, 1), \quad \rho(S_2, S_m) = 2.$$

Эволюция приводит к формированию квазивида – популяции, в которую входит как оптимальная последовательность, так и отличающиеся от нее мутационными заменами.

Формальная схема эволюции в модели квазивидов представлена в Таблице 1.

Таблица 1. Схема эволюции в модели квазивидов

| |
|--|
| Шаг 0. <i>Формирование начальной популяции</i> $\{S_k(0)\}$. Для всех $k = 1, 2, \dots, n$, $i = 1, 2, \dots, N$ выбираем символ S_{ki} случайно, полагая $S_{ki} = +1$ либо $S_{ki} = -1$. |
| Шаг 1. <i>Отбор</i> |
| Подшаг 1.1. <i>Расчет приспособленностей особей</i> . Для каждого $k = 1, 2, \dots, n$ вычисляем величину $f(S_k)$ согласно выражению (1). |
| Подшаг 1.2. <i>Формирование новой популяции</i> $\{S_k(t+1)\}$. Отбираем n особей в новую популяцию $\{S_k(t+1)\}$ с вероятностями, пропорциональными $f(S_k)$. |
| Шаг 2. <i>Мутации</i> . Для $k = 1, 2, \dots, n$; $i = 1, 2, \dots, N$ меняем знак $S_{ki}(t+1)$ на противоположный с вероятностью P ; P – параметр интенсивности мутаций. |
| <i>Организация последовательности поколений</i> . Повторяем шаги 1, 2 для $t = 1, 2, \dots$; t – номер поколения. |

В этой схеме начальное распределение особей случайное, а затем в каждом поколении происходит отбор и мутации особей.

Подшаг 1.2 требует пояснения. На этом подшаге происходит вероятностный отбор особей в следующее поколение рулеточным методом. Формирование новой популяции происходит следующим образом. Представим, что имеется рулетка. Для каждого поколения размечаем рулетку на n секторов, долю k -го сектора (отнесенную ко всей площади круга) полагаем равной $q_k = f_k [\sum_i f_i]^{-1}$ (рис. 1); $f_k = f(S_k)$. Далее n раз крутится рулетка, каждый раз определяется номер сектора, на котором останавливается стрелка рулетки, и соответствующая этому номеру особь выбирается в популяцию следующего поколения. Таким образом, в следующее поколение будут отобраны ровно n особей. При этом для каждого вращения рулетки вероятность k -й особи попасть в следующее поколение равна q_k , то есть пропорциональна ее приспособленности f_k . Некоторые особи могут быть отобраны в новое поколение несколько раз, это означает, что в новой популяции будет несколько потомков данной особи.

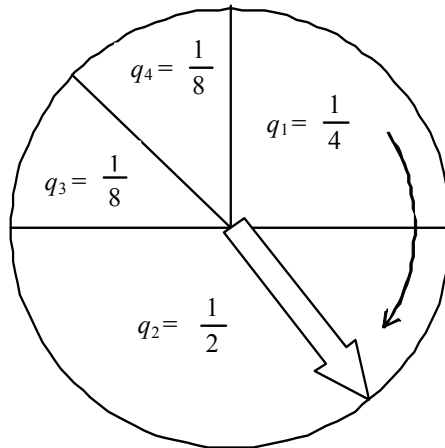


Рис. 1. Схема рулеточного отбора, при котором особи выбираются в популяцию нового поколения с вероятностями q_k , пропорциональными их приспособленностям f_k . Показан пример, для которого $n = 4$, $f_1 = 2$, $f_2 = 4$, $f_3 = 1$, $f_4 = 1$

3. Качественная картина эволюции

В работах (Редько 1986; Редько, Цой 2005) был приведен цикл компьютерных расчетов в рамках рассматриваемой модели квазивидов с хемминговой мерой близости между последовательностями. Результаты расчетов демонстрируют, что при достаточно естественно выбранных параметрах ($N, n \gg 1$, $\beta \geq PN$, $PN \leq 1$, то есть когда интенсивность отбора достаточно

велика, а интенсивность мутаций ограничена) эволюцию можно охарактеризовать следующим образом (пример расчета представлен на Рис. 2):

- начальное случайное распределение по ρ в популяции близко к нормальному распределению со средним $\langle \rho \rangle = N/2$ и дисперсией $N/4$ ($\langle \rho \rangle$ – среднее по популяции расстояние по Хеммингу до оптимальной последовательности S_m);

- процесс эволюции можно характеризовать двумя стадиями: первой – быстрой и второй – медленной;

- на первой стадии происходит отбор особей, расположенных «на левом крыле» исходного распределения, и распределение сжимается (это происходит быстро за счет достаточно большой интенсивности отбора);

- на второй стадии появление новых особей в популяции ограничено мутациями, из-за этого ограничения скорость уменьшения $\langle \rho \rangle$ значительно ниже, чем на первой стадии;

- окончательное распределение характеризует квазивид – распределение в окрестности оптимальной последовательности S_m .

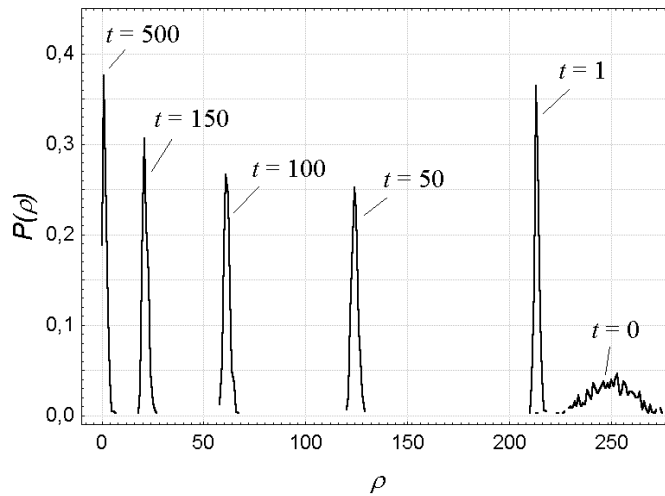


Рис. 2. Эволюция распределения особей, $N = 500$, $n = N$, $\beta = 1$, $P = 0,002$. По оси ординат отложена доля особей $P(\rho)$, имеющих рассматриваемое значение ρ , t – номер поколения

Отметим, что для представленного на Рис. 2 расчета первая стадия за счет достаточно интенсивного отбора ($\beta = 1$) действительно проходит быстро, она длится всего лишь одно поколение эволюции. Основное время эволюции занимает вторая стадия, на которой скорость эволюции существенно ограничена появлением новых особей в результате мутаций. Уменьшение величины $\langle \rho \rangle$ со временем происходит неравномерно: в начале второй

стадии скорость уменьшения $\langle \rho \rangle$ выше, чем в дальнейшем, так как число благоприятных мутаций, приводящих к понижению ρ , постепенно уменьшается. Окончательное распределение (близкое к показанному при $t = 500$) формируется в окрестности оптимальной последовательности S_m , для которой $\rho = 0$.

4. Стохастический характер эволюционного процесса. Роль нейтрального отбора

Далее предполагается, что длина N последовательностей S (модельных ДНК) достаточно велика, так что $2^N \gg n$. Это неравенство означает, что общее число возможных видов особей S , равное 2^N , значительно больше численности популяции. При этом число отдельных видов в популяции будет невелико, а многие виды в ней вообще будут отсутствовать (см., например, процесс, представленный на Рис. 2, для которого $2^N \sim 10^{150}$, $n = 500$). В силу ограниченности численности популяции существенны флуктуации числа видов, и рассматриваемые эволюционные процессы имеют стохастический характер. В частности, необходимо учитывать нейтральный отбор, то есть фиксацию особей, не зависящую от их приспособленностей (Кимура 1985; Редько 2005).

Для того чтобы продемонстрировать особенности нейтрального отбора явно, рассмотрим «чисто нейтральную» эволюционную игру, которую определим следующим образом.

Имеется популяция черных и белых шаров, общее количество шаров в популяции равно n . Эволюция состоит из последовательности поколений. Каждое поколение включает в себя два шага. На первом шаге дублируются все шары с сохранением их цвета: черный шар имеет два черных потомка, белый шар имеет два белых потомка. На втором шаге из популяции случайным образом удаляется ровно половина шаров, независимо от их цвета.

Подчеркнем, что отбор шаров в следующее поколение происходит совершенно нейтрально, независимо от цвета шара.

Считаем, что популяция находится в l -состоянии, если число черных и белых шаров для рассматриваемого поколения равны l и $n - l$ соответственно. Будем характеризовать эволюцию вероятностями P_{lm} переходов из l -состояния в m -состояние в течение одного поколения. Используя простой комбинаторный расчет, можно определить значения P_{lm} :

$$P_{lm} = \begin{cases} \binom{2l}{m} \times \binom{2n-2l}{n-m} / \binom{2n}{n}, & \text{если } 2l-n \leq m \leq 2l \\ 0, & \text{если } m > 2l \text{ или } m < 2l-n, \end{cases} \quad (2)$$

$$\text{где } \binom{a}{b} = \frac{a!}{(a-b)!b!}.$$

Матрица P_{lm} задает случайный марковский процесс, который может рассматриваться как пример простого стохастического эволюционного процесса (Карлин 1975). Используя общие методы анализа таких процессов, можно показать, что:

1) рассматриваемый процесс всегда сходится к одному из поглощающих состояний, а именно – к 0-состоянию (все шары белые) либо к n -состоянию (все шары черные);

2) при больших n характерное число поколений T_n , требуемое для сходимости к какому-либо из поглощающих состояний, равно $2n$:

$$T_n = 2n. \quad (3)$$

Таким образом, хотя данный эволюционный процесс чисто нейтральный (черные и белые шары имеют равные шансы «выжить»), однако в результате эволюции отбирается только один вид шаров. Величина T_n представляет собой оценку характерного времени (числа поколений) нейтрального отбора.

Итак, характерное число поколений, за которое происходит нейтральный отбор, порядка численности популяции n .

5. Оценка скорости и эффективности эволюционного процесса

Теперь, учитывая качественный характер эволюции и роль нейтрального отбора, оценим скорость и эффективность рассматриваемого эволюционного процесса в модели квазивидов с хемминговой мерой близости.

Предполагаем, что численность популяции n достаточно велика:

$$T_n \geq T, \quad [1 - (1 - P)^N]^n \ll 1, \quad (4)$$

где T_n – характерное время (число поколений) нейтрального отбора; T – характерное время сходимости всего эволюционного процесса. Первое неравенство в формуле (4) означает, что мы считаем не слишком сильным влияние нейтрального отбора, так как характерное время всей эволюции меньше или порядка T_n . Отбор в соответствии с приспособленностями особей происходит быстрее (точнее, не медленнее), чем нейтральный отбор. Второе неравенство соответствует пренебрежению мутационными потерями уже найденных удачных особей в популяции.

При больших N величина T определяется второй (медленной) стадией эволюции, на которой эволюционный поиск происходит следующим образом. Для того чтобы появились новые особи с меньшими значениями ρ , должно произойти достаточное количество мутаций, а затем фиксация этих новых особей в результате отбора. Оценим характерное время t_{-1} , за которое $\langle \rho \rangle$ уменьшается на 1. Оно составляет: $t_{-1} \sim t_m + t_{от}$, где $t_m \sim (NP)^{-1}$ – характерное время, за которое особи популяции промутируют; $t_{от} \sim \beta^{-1}$ – характерное время, за которое особи, для которых $\rho = \langle \rho \rangle - 1$, в результате отбора вытеснят из популяции особи, для которых $\rho = \langle \rho \rangle$. Оценим

общее число поколений всего эволюционного процесса. Учитывая, что за время всего эволюционного процесса $\langle \rho \rangle$ должно уменьшиться на величину, приближенно равную $N/2$, полагаем $T \sim t_{-1} N$ и имеем:

$$T \sim P^{-1} + N\beta^{-1}. \quad (5)$$

Общее число особей, участвующих в эволюции, составляет $n_{\text{общ.}} = nT$. Оценим величину $n_{\text{общ.}}$ для заданного N при некоторых достаточно разумно выбранных остальных параметрах β, P, n . Эти параметры выбираем таким образом, чтобы по возможности уменьшить величины T и $n_{\text{общ.}}$. Интенсивность отбора считаем достаточно большой: $\beta \geq PN$, тогда можно пренебречь вторым слагаемым в формуле (5). Полагаем, что интенсивность мутаций P соответствует примерно одной мутации в каждой последовательности в каждом поколении, $P \sim N^{-1}$; при такой частоте мутаций, с одной стороны, появление новых особей в популяции в результате мутаций происходит достаточно быстро, с другой – можно пренебречь мутационными потерями (выполняется второе неравенство в формуле (4)). С учетом сказанного имеем:

$$T \sim N. \quad (5a)$$

Также полагаем, что первое из неравенств (4) выполняется «на пределе» ($T_n \sim T$), то есть для численности популяции имеем оценку:

$$n \sim T_n \sim T \sim N. \quad (6)$$

Оценки (6) предполагают минимальную допустимую численность популяции, при которой еще не очень существенны потери особей в результате нейтрального отбора. С учетом сделанных предположений получаем оценки характерного времени эволюции T и общего числа особей, участвующих в эволюции $n_{\text{общ.}}$ ($n_{\text{общ.}} = nT \sim NT \sim N^2$):

$$T \sim N, n_{\text{общ.}} \sim N^2. \quad (7)$$

Выражение (7) определяет оценки скорости сходимости всего эволюционного процесса (значение T) и эффективности эволюционного поиска оптимальной особи S_m (значение $n_{\text{общ.}}$).

Резюмируем наиболее существенные сделанные предположения и полученные оценки. Оценки сделаны в следующих предположениях:

- 1) $\beta \geq 1$ – интенсивность отбора достаточно велика;
- 2) $P \sim N^{-1}$ – мутации «оптимальны» (порядка одной мутации на геном);
- 3) $n \sim N$ – условие пренебрежения нейтральным отбором выполняется «на пределе».

Результат оценок:

- 1) число поколений эволюции порядка длины генома: $T \sim N$;
- 2) численность популяции также этого порядка: $n \sim N$;
- 3) общее число особей, участвующих в эволюционном поиске, порядка квадрата длины генома: $n_{\text{общ.}} = nT \sim N^2$.

6. Результаты численного моделирования

Приведенные оценки были получены при использовании ряда допущений. Поэтому эти оценки были проверены путем компьютерного расчета (Редько, Цой 2005). Расчет проводился для соотношений между параметрами, которые соответствуют условиям получения оценок (7): $n = N$, $P = N^{-1}$, $\beta = 1$.

Расчет выполнялся следующим образом. Были получены зависимости среднего по популяции расстояния до оптимума от времени $\langle \rho \rangle(t)$ для разных значений N (Рис. 3), и по этим зависимостям оценивалось характерное время сходимости эволюции T двумя способами: 1) рассчитывалось характерное время релаксации T_R в зависимостях $\langle \rho \rangle(t)$ по начальному наклону этих кривых; 2) определялось время выхода T_S на стационарное значение $\langle \rho \rangle$, которое получается при больших t (см. Рис. 3).

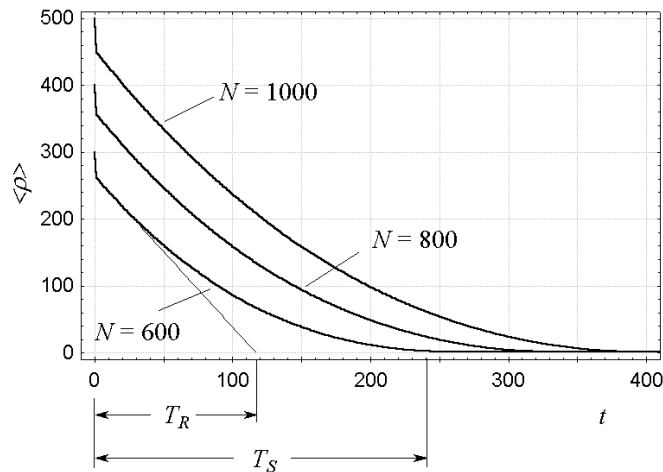


Рис. 3. Зависимости $\langle \rho \rangle(t)$ для разных значений N ; t – номер поколения. Показана схема оценки значений T_R и T_S для случая $N = 600$. Зависимости усреднены по 50 расчетам

Полученные в результате зависимости $T_R(N)$ и $T_S(N)$ представлены на Рис. 4. Кроме того, определялись значения времени T_O первого появления оптимальной особи S_m в популяции. Соответствующая зависимость $T_O(N)$ также представлена на Рис. 4.

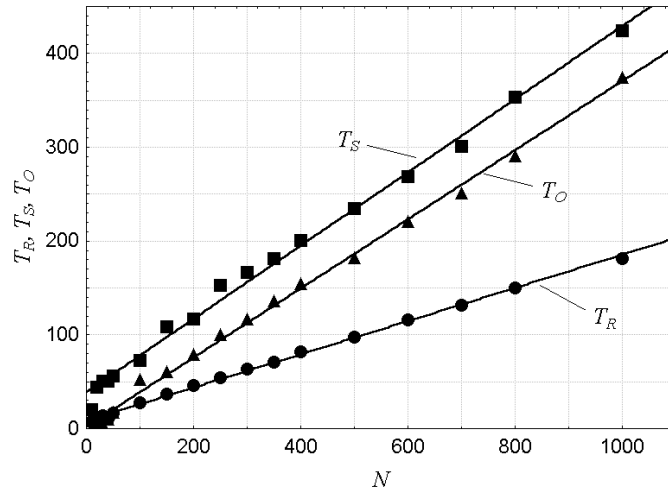


Рис. 4. Зависимость времени релаксации T_R , времени выхода на стационар T_S и времени нахождения оптимальной особи T_O от длины генома N . Приведенные результаты усреднены по 50 расчетам

Видно, что при достаточно больших N все три зависимости линейны:

$$T_R(N) = k_R N + T_{R0}, \quad T_S(N) = k_S N + T_{S0}, \quad T_O(N) = k_O N + T_{O0},$$

где $k_R = 0,1772$; $k_S = 0,3903$; $k_O = 0,3685$; $T_{R0} = 8,2709$; $T_{S0} = 38,7356$; $T_{O0} = 2,1288$, что вполне согласуется с оценками (7). Значения k_R , k_S , k_O показывают, что оценки скорости сходимости эволюционных процессов были получены с некоторым «запасом», что неудивительно, так как в рассматриваемой модели перемещение к оптимальной последовательности S_m возможно разными способами за счет большого количества благоприятных мутаций, приводящих к уменьшению ρ ; эта возможность приводит к увеличению скорости уменьшения $\langle \rho \rangle$ по сравнению со сделанными оценками. Отметим, что продемонстрированная компьютерными расчетами линейность зависимостей $T_R(N)$, $T_S(N)$, $T_O(N)$ неочевидна и примечательна.

Итак, получены и проверены путем численного моделирования аналитические оценки скорости сходимости и эффективности эволюционных процессов для модели квазивидов с хемминговой мерой близости.

Сравним эволюционный метод оптимизации рассматриваемой функции приспособленности (1) с двумя простейшими методами: последовательный поиск и случайный перебор.

Последовательный поиск организуем следующим образом. Исходим из произвольной последовательности S , символы которой равны $S_i = 1$ либо -1 . Далее последовательно для каждого i ($i = 1, 2, \dots, N$) меняем знак символа ($S_i \rightarrow -S_i$) и при увеличении приспособленности $f(S)$ со сменой

знака принимаем новое значение символа, а при уменьшении $f(S)$ – возвращаемся к старому. В результате после N испытаний найдем оптимальную последовательность S_m . Итак, для последовательного поиска имеем: $n_{\text{общ.}} = N$.

При случайном переборе для нахождения оптимальной последовательности необходимо испытать порядка 2^N последовательностей:

$$n_{\text{общ.}} \sim 2^N.$$

Полученные оценки приведены в Таблице 2.

Таблица 2. Оценки эффективности методов поиска

| Метод поиска | $n_{\text{общ.}}$ | $n_{\text{общ.}}$ при $N = 1000$ |
|------------------|-------------------|----------------------------------|
| Последовательный | N | 1000 |
| Эволюционный | $\sim N^2$ | $\sim 10^6$ |
| Случайный | $\sim 2^N$ | $\sim 10^{300}$ |

Приведенные оценки показывают, что эволюционный процесс, как алгоритм поиска S_m , в рассмотренной модели в определенном смысле «суб-оптимален»: он не обеспечивает максимальную скорость поиска (например, для рассматриваемой задачи возможен более эффективный последовательный перебор), тем не менее он значительно эффективнее случайного поиска, при котором необходимо испытать порядка 2^N последовательностей: $n_{\text{общ.}} \sim 2^N$.

Хотя оценки (7) характерного времени эволюции и общего числа особей, участвующих в ней, получены для модели квазивидов, в которой символы геномов бинарны, эта же методика получения оценок скорости и эффективности эволюции применима к нескольким близким эволюционным моделям.

В частности, в работе (Red'ko, Tsoy 2006) аналогичная методика была применена к модели квазивидов, в которой символы геномов S принимали λ различных значений (λ – небольшое целое число, например $\lambda = 4$). Как и в изложенной выше модели, предполагается существование оптимальной особи S_m и то, что приспособленность произвольной особи S экспоненциально уменьшается с ростом расстояния по Хеммингу $\rho(S, S_m)$ между S и S_m (см. выражение (1)).

Поскольку для этого случая методика получения оценок и проверки их численным расчетом совершенно аналогична вышеизложенной, приведем соответствующий результат без вывода. При небольших целых λ вместо оценок (7) имеем: характерное время эволюции T и общее число особей, участвующих в эволюции $n_{\text{общ.}}$, определяются выражением:

$$T \sim (\lambda - 1)N, \quad n_{\text{общ.}} \sim [(\lambda - 1)N]^2. \quad (8)$$

Отметим, что формула (8) обобщает выражение (7): при $\lambda = 2$ формула (8) переходит в выражение (7).

Далее рассмотрим эволюционные модели, близкие к изложенной модели квазивидов с хемминговой мерой близости.

7. Модель узкого канала и мажорирующая модель

В настоящем разделе рассмотрим эволюционную модель узкого канала и близкую к ней мажорирующую модель.

В модели узкого канала предполагается эволюция популяции особей, геномы которых представляют собой последовательности S_k . Как и в модели с хемминговой мерой близости, символы последовательностей S_{ki} принимают два значения: $S_{ki} = +1$ либо -1 ; $i = 1, 2, \dots, N$; $k = 1, 2, \dots, n$; N – длина последовательностей; n – численность популяции. Также имеется оптимальная последовательность-особь S_m , приспособленность которой максимальна. Однако число последовательностей S_k , приспособленность которых отлична от нуля, ограничено. Имеется только $N + 1$, такая последовательность S_k , $k = 0, 1, 2, \dots, N$, причем $S_0 = S_m$, число несовпадающих символов в S_k и S_m равно k , а приспособленность последовательности S_k равна $f(S_k) = \exp(-\beta k)$. При этом для того, чтобы перейти от последовательности S_k к S_{k-1} , нужно изменить знак только одного определенного символа. Следовательно, чтобы перейти от последовательности S_N к S_0 , нужно поочередно изменить N символов в строго заданном порядке, то есть пройти по «узкому каналу». И в процессе эволюции мутации должны произойти именно в этом порядке. При «неправильных» мутациях будут появляться нежизнеспособные последовательности-особи с нулевой приспособленностью. Для наглядности приведем простой пример таких особей ($N = 4$): $S_4 = -1, -1, -1, -1$; $S_3 = -1, -1, -1, 1$; $S_2 = -1, -1, 1, 1$; $S_1 = -1, 1, 1, 1$; $S_0 = 1, 1, 1, 1$. Все остальные последовательности имеют нулевую приспособленность. Видно, что для того, чтобы пройти от S_4 к S_0 , нужно идти по «узкому каналу», поочередно меняя символы в нужном порядке.

Было проведено компьютерное моделирование (Редько 1999) в рамках модели узкого канала. При компьютерных расчетах скорость сходимости эволюционного процесса была меньше, чем для модели с хемминговой мерой близости, а величина скорости уменьшения среднего по популяции значения k была практически постоянна в течение всего эволюционного процесса. Скорость сходимости эволюции в модели узкого канала мажорирует скорости эволюции квазивидов (см. Рис. 2 лекции 4 в [Редько 1999]). Этот факт привел к построению более простой мажорирующей модели, скорость эволюции в которой мажорирует таковую в модели квазивидов, в ней можно не учитывать информационную структуру последовательностей-особей.

В мажорирующей модели предполагаем, что имеется всего $N + 1$ типов особей: S_k , $k = 0, 1, \dots, N$. Особь k -го типа может в результате мутации

перейти в особь $k - 1$ -го типа либо в особь $k + 1$ -го типа. Вероятность любой из этих мутаций равна P_1 . Другие мутации отсутствуют. Приспособленность особей k -го типа равна $f_k = \exp(-\beta k)$. Численность популяции равна n . Эволюционный процесс организуется так же, как и в модели квазивидов с хемминговой мерой близости. Были проведены компьютерные расчеты по данной модели. На Рис. 5 приведены зависимости средних по популяции значений от времени $\langle \rho \rangle (t)$ (для модели квазивидов с хемминговой мерой близости) и $\langle k \rangle (t)$ (для мажорирующей модели) при $n = N = 1000$, $P = P_1 = N^{-1}$, $\beta = 1$.

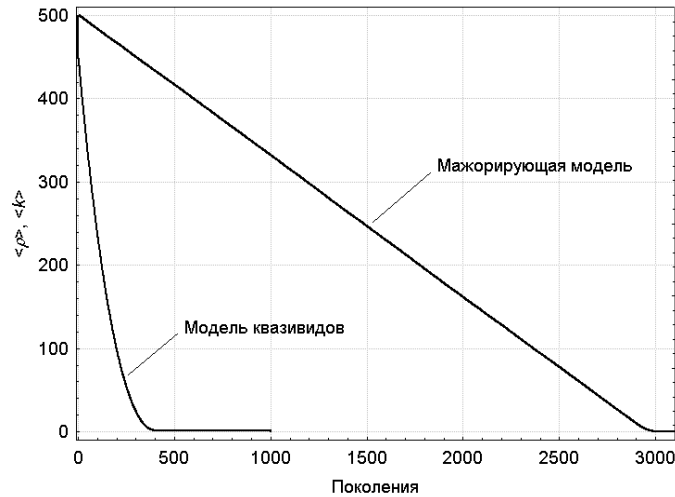


Рис. 5. Зависимости $\langle \rho \rangle (t)$ и $\langle k \rangle (t)$ при $N = n = 1000$, $P = P_1 = 0,001$, $\beta = 1$; t – номер поколения. Результаты усреднены по 50 расчетам

Видно, что в мажорирующей модели скорость уменьшения $\langle k \rangle$ практически постоянна и число поколений по порядку величины близко к N , хотя и превышает это значение. Так как эволюционный процесс в данной модели, как и для основной модели, определяется мутациями и отбором, то для мажорирующей модели приближенно справедливы оценки скорости и эффективности эволюционного процесса (7).

Рассмотренные эволюционные модели соответствуют случаю единственного максимума приспособленности, что является их ограничением. В следующем разделе излагается «спин-стекольная» модель эволюции (Редько 1990; 2005), в которой число максимумов приспособленности экспоненциально растет с ростом длины генома особи.

8. Модель эволюционного поиска минимумов энергии спинового стекла

Построим модель для многоэкстремальной функции приспособленности $f(S)$, используя известную модель спиновых стекол Шеррингтона – Киркпатрика. Эта модель (Sherrington, Kirkpatrick 1975; Kirkpatrick, Sherrington 1978) описывает систему попарно взаимодействующих спинов. Взаимодействия между спинами предполагаются случайными. Формально модель сводится к следующему.

1. Имеется система S , состоящая из N спинов: $S = (S_1, S_2, \dots, S_N)$; число спинов предполагается большим, $N \gg 1$. Спины принимают значения 1 либо -1 : $S_i = 1, -1$.

2. Взаимодействия между спинами случайны. Энергия спиновой системы есть:

$$E(S) = -\sum_{i < j} J_{ij} S_i S_j, \quad i, j = 1, \dots, N, \quad (9)$$

где J_{ij} – элементы матрицы случайных взаимодействий между спинами. Величины J_{ij} нормально распределены, плотность распределения $P(J_{ij})$ определяется выражением:

$$P(J_{ij}) = (2\pi)^{-1/2} (N-1)^{1/2} \exp[-J_{ij}^2 (N-1)2^{-1}]. \quad (10)$$

Для модели (9), (10) известно, что число локальных минимумов энергии M экспоненциально растет с увеличением N (Tanaka, Edwards 1980):

$$M \sim \exp(\alpha N), \quad \alpha \approx 0,2. \quad (11)$$

Локальный минимум есть такое состояние системы спинов S_L , для которого переворот любого одного спина ($S_i \rightarrow -S_i$) приводит к повышению энергии.

Глобальный минимум энергии E_0 приближенно составляет величину $-0,8N$ (Young, Kirkpatrick 1982):

$$E_0 \approx -0,8N. \quad (12)$$

Из формул (9), (10) следует, что среднее вариации энергии ΔE при перевороте одного спина ($S_i \rightarrow -S_i$) равно 2:

$$\langle \Delta E \rangle = 2. \quad (13)$$

Среднее значение энергии при случайной ориентации спинов согласно (9), (10) равно 0:

$$\langle E \rangle = 0. \quad (14)$$

Рассмотрим эволюцию популяции, в которой геномы модельных особей S представляют собой информационные последовательности N символов $S = (S_1, S_2, \dots, S_N)$. Приспособленность особи S_k определяем как:

$$f(S_k) = \exp[-\beta E(S_k)], \quad (15)$$

где $E(S_k)$ – энергия спинового стекла. Локальные минимумы энергии $E(S_k)$ соответствуют локальным максимумам приспособленности $f(S_k)$, число этих максимумов экспоненциально растет с ростом длины генома N .

Эволюционный процесс строим так же, как и в модели квазивидов с хемминговой мерой близости: исходная популяция состоит из случайных особей, отбор и мутации производятся в соответствии с Таблицей 1.

Подчеркнем, что удаление средней энергии от глобального минимума в начальном поколении согласно (12), (14) порядка N . Вариация энергии при мутации $S_i \rightarrow -S_i$ согласно (13) порядка 1. Таким образом, эти величины, определяющие скорость эволюции, близки к таковым для модели квазивидов с хемминговой мерой близости. Поэтому и для спинстекольной модели эволюции справедливы оценки скорости сходимости и эффективности эволюционного процесса, определяемые выражением (7).

Эти оценки согласуются с результатами компьютерных расчетов (Редько 1990; 1999; 2005). Эти расчеты также показали, что в спинстекольной модели эволюции находится не единственный, а один из локальных (близкий по величине E к глобальному) минимумов энергии спинового стекла. То есть находится не единственный максимум приспособленности, а один из возможных локальных максимумов, достаточно близкий к глобальному максимуму.

9. Заключение

Итак, получены количественные оценки скорости сходимости и эффективности в нескольких сходных эволюционных моделях. Эти оценки таковы.

Общее число поколений эволюционного процесса, требуемого для нахождения оптимального генома, составляет величину порядка длины генома: $T \sim N$; численность популяции имеет тот же порядок: $n \sim N$; общее число особей, участвующих в эволюционном поиске, составляет величину порядка квадрата длины генома: $n_{\text{общ.}} = nT \sim N^2$.

Библиография

- Карлин С. 1975. *Основы теории случайных процессов*. М.: Мир.
- Кимура М. 1985. *Молекулярная эволюция: теория нейтральности*. М.: Мир.
- Редько В. Г. 1986. Оценка скорости эволюции в моделях Эйгена и Куна. *Биофизика* 31(3): 511–516.
- Редько В. Г. 1990. Спиновые стекла и эволюция. *Биофизика* 35(5): 831–834.
- Редько В. Г. 1999. Эволюционная кибернетика. Интернет-лекции. URL: <http://www.keldysh.ru/pages/BioCyber/Lectures.html>; <http://www.keldysh.ru/pages/BioCyber/Lecture4.html>.
- Редько В. Г. 2005. *Эволюция, нейронные сети, интеллект: Модели и концепции эволюционной кибернетики*. М.: КомКнига.
- Редько В. Г., Цой Ю. Р. 2005. Оценка эффективности эволюционных алгоритмов. *Доклады АН* 404(3): 312–315.

- Эйген М. 1973. *Самоорганизация материи и эволюция биологических макромолекул*. М.: Мир.
- Эйген М., Шустер П. 1982. *Гиперцикл. Принципы самоорганизации макромолекул*. М.: Мир.
- Kirkpatrick S., Sherrington D. 1978. Infinite Range Model of Spin-glass. *Physical Review B* 17(11): 4384–4403.
- Red'ko V. G., Tsoy Yu. R. 2006. Estimation of the Evolution Speed for the Quasispecies Model: Arbitrary Alphabet Case. *Artificial Intelligence and Soft Computing – ICAISC 2006. 8th International Conference, Proceedings / Eds. L. Rutkowski et al. Zakopane, Poland: Springer*. Pp. 460–469.
- Sherrington D., Kirkpatrick S. 1975. Solvable Model of Spin-glass. *Physical Review Letters* 35(26): 1792–1796.
- Tanaka F., Edwards S. F. 1980. Analytic Theory of the Ground State of a Spin Glass: 1. Ising Spin Glass. *Journal of Physics F: Metal Physics* 10(12): 2769–2778.
- Young A. P., Kirkpatrick S. 1982. Low-temperature Behavior of the Infinite-range Ising Spin-glass: Exact Statistical Mechanics for Small Samples. *Physical Review B*. 25(1): 440–451.