

2.2 АНАЛИЗ ДАННЫХ И ПРОГНОЗИРОВАНИЕ

Главным препятствием к научного прогресса во многих областях является неспособность обработать огромные выборки данных, которые были собраны экспериментальным путем или компьютерным моделированием. В сферах статистики и машинного обучения главные усилия прилагались для разработки автоматических методов для нахождения существенных и интересных образцов в сложных данных, и для прогнозирования будущего из этих данных; в целом, однако, успех таких попыток был ограничен, и автоматический анализ сложных данных остается открытой проблемой. Анализ данных и предсказание могут часто формулироваться как проблемы поиска, к примеру, поиска модели, представляющей данные, поиска прогностических правил, или поиска специфической структуры или сценария, хорошо предсказывающего по данным. В этом разделе я описываю два проекта, в которых генетический алгоритм используется для решения таких проблем поиска - один из прогнозирования динамических систем, и другой из предсказания структуры белков.

Прогнозирование динамических систем

Norman Packard (1990) разработал форму ГА, направленную на решение этой проблемы и применил свой метод к нескольким задачам анализа данных и прогнозирования. Общая проблема может быть заявлена следующим образом: ряд наблюдений некоторого процесса (например, физической системы или формальной динамической системы) принимают форму набора пар,

$$\{(x^{-1}, y^1), \dots, (x^{-N}, y^N)\},$$

где $x^{-i} = (x_1^i, \dots, x_N^i)$ является независимой переменной и y^i является зависимой переменной. Например, в задачах прогнозирования погоды, независимой переменной может быть некоторый набор параметров сегодняшней погоды (например, средняя влажность, среднее барометрическое давление, минимальная и максимальная температуры, есть ли дождь или нет), и зависимой переменной может быть характеристика завтрашней погоды (например, дождь). В задачах прогнозирования на фондовой бирже независимыми переменными могут быть $x = (x(t_1), x(t_2), \dots, x(t_n))$, представляющие значения стоимости отдельных акций ("переменная состояния") в последующие шаги времени, и зависимая переменная может быть $y = x(t_n + k)$, представляющей стоимость акции через некоторое время в будущем. (В этих примерах для каждого вектора независимых переменных x имеется только одна зависимая переменная, более общий случай задачи предполагает вектор зависимых переменных для каждого вектора независимых переменных.)

Packard использовал ГА для поиска в пространстве наборов условий независимых переменных для таких наборов условий, которые хорошо прогнозируют зависимую переменную. К примеру, в задаче прогнозирования на фондовой бирже, особь в популяции ГА может быть таким набором условий как:

$$\begin{aligned} C = & \{(\$20 \leq \text{Price of Xerox stock on day 1}) \\ & \wedge (\$25 \leq \text{Price of Xerox stock on day 2} \leq \$27) \\ & \wedge (\$22 \leq \text{Price of Xerox stock on day 2} \leq \$25)\} \end{aligned}$$

где " \wedge " – логический оператор "И". Эта особь представляет все наборы трех дней, в которых заданные условия встречаются (возможен пустой набор, если условия никогда не пересекутся). Набор условий S , таким образом, определяет специфическое подмножество точек данных (здесь, набор всех 3-дневных периодов). Цель Packard состояла в том, чтобы использовать ГА для поиска наборов условий, которые хорошо предсказывают что-то, другими словами, искать наборы условий, которые определяют подмножества точек данных, чьи значения зависимых переменных близки к тому, чтобы быть равномерными. В примере для фондовой биржи, если ГА нашел набор условий,

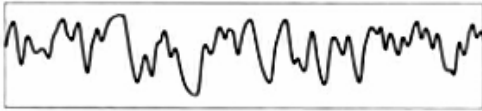


Рис. 2.11: Временной ряд для уравнения Mackey-Glass при $\ddot{A} = 150$. Время откладывается на горизонтальной оси; $x(t)$ откладывается по вертикальной оси. (Перепечатано из Martin Casdagli and Stephen Eubank, eds., *Nonlinear Modeling and Forecasting*; © 1992 Addison-Wesley Publishing Company, Inc. Печатается с разрешения издателя.)

такой что все дни удовлетворяют этому набору, следуют за днями, в которые цена акций Ксерокса росла приблизительно до 30\$, тогда мы можем с уверенностью предсказать, что если эти условия удовлетворяют сегодняшнему дню, цена акций Хегох повысится.

Фитнесс каждой особи S вычисляется пропуском всех данных (\bar{x}_y) в обучающем наборе через S и для каждого \bar{x} , который удовлетворяет S , запоминается соответствующий y . После того как это сделано, измерение состоит из равномерных результирующих значений y . Если все значения y близки к соответствующему значению \dot{A} , тогда S является кандидатом, хорошо прогнозирующим y – то есть, можно надеяться, что новый \bar{x} , удовлетворяющий S , также будет соответствовать значению y , близкому к \dot{A} . С другой стороны, если значения y очень отличаются одно от другого, тогда x , удовлетворяющий S , не представляется прогнозирующим что-то соответствующее значению y .

В качестве иллюстрации этого подхода, я опишу работу, сделанную Thomas Meyer и Norman Packard (1992) в обнаружении "области предсказуемости" во временном ряду, сгенерированном уравнением Mackey-Glass, хаотической динамической системы, созданной как модель для кровотока (Mackey и Glass 1977):

$$\frac{dx}{dt} = \frac{ax(1-\tau)}{1+[x(t-\tau)]^c} - bx(t)$$

Здесь $x(t)$, это переменная состояния, t – время в секундах и a, b, c и \ddot{A} константы. Временной ряд из этой системы (с \ddot{A} принятым 150) изображен на рис. 2.11.

Для создания выборки данных Meyer и Packard сделали следующее: для каждой i -той точки данных, независимая переменная \bar{x}^i является 50-ю последовательными значениями $x(t)$ (одно в секунду):

$$\bar{x}^i = (x_1^i, x_2^i, \dots, x_{50}^i)$$

Зависимой переменной y^i для каждой i -той точки данных является переменная состояния t через $y^i = x_{50+t}^i$, временных интервалов в будущем. Каждая точка данных (\bar{x}^i, y^i) была сформирована вычислением уравнения Mackey-Glass с разными начальными условиями, состоящими из значений для $\{x_1, \dots, x_0\}$.

Meуer и Packard использовали следующую фитнесс-функцию:

$$f(C) = -\log_2 \left(\frac{\sigma}{\sigma_0} \right) - \frac{\alpha}{N_C}$$

где σ – стандартное отклонение набора y для точек данных, удовлетворяющих C , σ_0 – стандартное отклонение распределения y по всему набору данных, N_C – число точек данных, удовлетворяющих условиям C , и \pm – константа. Первое составляющее фитнесс-функции измеряет количество информации в распределении y для точек, удовлетворяющих C , и второе составляющее – это штраф за плохую статистику – если число точек, удовлетворяющих C , является маленьким, то первое составляющее менее достоверно, таким образом, C будет иметь более низкий фитнесс. Константа \pm может быть приспособлена для каждого отдельного применения.

Meуer и Packard использовали следующий вариант ГА:

1. Инициализировать популяцию случайным набором C .
2. Вычислить фитнесс каждого из C .
3. Оценить популяцию по фитнессу.
4. Отбросить некоторых малопригодные особи и заменить их новыми C , полученными применением кроссовера и мутации к оставшимся C .
5. Перейти к шагу 2.

(Их метод селекции был, как используемый в проекте клеточных автоматов, описанном выше, подобен методу эволюционной стратегии " $\frac{1}{4} + >>$ ") Meуer и Packard использовали форму кроссовера, известного в литературе по ГА как "однородный кроссовер" (Syswerda 1989). Этот оператор берет два C и меняет приблизительно половину "генов" (условий). Таким образом, в каждой позиции гена в родителе A и родителе B , принимается случайное решение, должен ли этот ген войти в потомок A или потомок B . Пример приведен ниже:

Родитель A: $\{(3.2 \leq x_6 \leq 5.5) \wedge (0.2 \leq x_8 \leq 4.8) \wedge (3.4 \leq x_9 \leq 9.9)\}$

Родитель B: $\{(6.5 \leq x_2 \leq 6.8) \wedge (1.4 \leq x_4 \leq 4.8) \wedge (1.2 \leq x_9 \leq 1.7) \wedge (4.8 \leq x_{16} \leq 5.1)\}$

Потомок A: $\{(3.2 \leq x_6 \leq 5.5) \wedge (1.4 \leq x_4 \leq 4.8) \wedge (3.4 \leq x_9 \leq 9.9)\}$

Потомок B: $\{(6.5 \leq x_2 \leq 6.8) \wedge (0.2 \leq x_8 \leq 4.8) \wedge (1.2 \leq x_9 \leq 1.7) \wedge (4.8 \leq x_{16} \leq 5.1)\}$

Здесь потомок A имеет два гена из родителя A и один ген из родителя B . Потомок B имеет один ген из родителя A и три гена из родителя B .

В дополнение к кроссоверу, было использовано четыре различных оператора мутации:

Добавление условия:

$$\{(3.2 \leq x_6 \leq 5.5) \wedge (0.2 \leq x_8 \leq 4.8)\}$$

$$\rightarrow \{(3.2 \leq x_6 \leq 5.5) \wedge (0.2 \leq x_8 \leq 4.8) \wedge (3.4 \leq x_9 \leq 9.9)\}$$

Удаление условия:

$$\{(3.2 \leq x_6 \leq 5.5) \wedge (0.2 \leq x_8 \leq 4.8) \wedge (3.4 \leq x_9 \leq 9.9)\}$$

$$\rightarrow \{(3.2 \leq x_6 \leq 5.5) \wedge (0.2 \leq x_8 \leq 4.8)\}$$

Расширение или уменьшение диапазона:

$$\{(3.2 \leq x_6 \leq 5.5) \wedge (0.2 \leq x_8 \leq 4.8)\}$$

$$\rightarrow \{(3.9 \leq x_6 \leq 4.8) \wedge (0.2 \leq x_8 \leq 4.8)\}$$

Смещение диапазона вверх или вниз:

$$\{(3.2 \leq x_6 \leq 5.5) \wedge (0.2 \leq x_8 \leq 4.8)\}$$

$$\rightarrow \{(3.2 \leq x_6 \leq 5.5) \wedge (1.2 \leq x_8 \leq 5.8)\}$$

Результаты запуска ГА, использующим эти данные для $\ddot{A} = 150$ из временной серии при $t = 150$ проиллюстрированы на [рис. 2.12](#) и [рис. 2.13](#). [Рис 2.12](#) приводит четыре набора условий с самым высоким фитнесом найденные ГА, и [рис 2.13](#) показывает четыре результата этих наборов условий. Каждый из четырех графиков на [рис 2.13](#) показывает траектории, соответствующие точкам данных (x^i, y^i) , которые удовлетворяют набору условий. Крайняя левая белая область - начальные 50 временных интервалов, в течение которых были взяты данные. Вертикальные линии в той области представляют различные условия для x .

\bar{x} представлен набором условий. Например, на графике *a* крайняя левая вертикальная линия представляет условие для x_{20} (этот набор траекторий отображен, начиная с 20-ого временного интервала), и самая правая вертикальная линия в этой области представляет условие для x_{49} . Заштрихованная область представляет промежуток времени между шагами 50 и 200, и самая правая вертикальная линия отмечает шаг времени 200 (точка, в которой был сделано наблюдение y^i). Заметьте, что в каждом из этих графиков значение y^i попадает в очень узкий диапазон, что означает успешность применения ГА в обнаружении подмножеств данных, для которых возможен очень точный прогноз. (О других результатах в том же направлении сообщено Meyer в 1992.)

Это очень выдающиеся результаты, но немедленно возникают некоторые вопросы. Первый и самый важный, правильно ли прогнозируют обнаруженные условия для точек данных, не входящих в обучающую выборку (то есть, набор данных использовался для вычисления фитнеса), или они просто описывают случайные статистические колебания в данных, которые были изучены ГА? Мейер и Паккард выполнили множество "типовых" испытаний с данными не из обучающей выборки, которые удовлетворяли полученным наборам условий и обнаружили, что результаты были адекватными - значения y' для этих данных также имели тенденцию находиться в узком диапазоне (Томас Мейер, персональное общение).

Как точно ГА решает проблему? Каковы схемы, которые обрабатываются? Какова роль кроссовера в обнаружении хорошего решения? Однородный кроссовер использовавшегося типа имеет совсем другие свойства, нежели одноточечный кроссовер, и его использование усложняет выявление рекомбинированных схем. Мейер (персональное общение) обнаружил, что отказ от кроссовера и использование исключительно четырех операторов мутации не приводило к большим различиям в работе ГА; как в случае генетического программирования, это поднимает вопрос того, является ли ГА лучшим методом для этой задачи. Интересным расширением этой работы могло бы стать проведение экспериментов, сравнивающих выполнение ГА с таким другим методом поиска как поиск экстремума.

$$C_1 = \left\{ \begin{array}{l} (x_{20} > 1.122) \wedge (x_{25} < 1.330) \wedge (x_{28} > 1.168) \wedge \\ (x_{39} < 1.342) \wedge (x_{41} > 1.304) \wedge (x_{49} > 1.262) \end{array} \right\} \rightarrow y = 0.18 \pm 0.014$$

$$C_2 = \left\{ \begin{array}{l} (x_{25} < 1.330) \wedge (x_{26} > 1.177) \wedge (x_{31} > 1.127) \wedge \\ (x_{38} < 1.156) \wedge (x_{40} < 1.256) \wedge (x_{46} > 1.194) \wedge \\ (x_{47} < 1.311) \wedge (x_{49} > 1.070) \end{array} \right\} \rightarrow y = 0.27 \pm 0.019$$

$$C_3 = \left\{ \begin{array}{l} (x_{24} > 0.992) \wedge (x_{29} < 1.190) \wedge (x_{30} > 1.020) \wedge \\ (x_{34} < 1.090) \wedge (x_{40} < 0.951) \wedge (x_{42} > 0.599) \wedge \\ (x_{43} > 0.591) \wedge (x_{49} < 0.763) \wedge (x_{50} > 0.576) \end{array} \right\} \rightarrow y = 1.22 \pm 0.024$$

$$C_4 = \left\{ \begin{array}{l} (x_{20} < 0.967) \wedge (x_{22} < 1.049) \wedge (x_{26} > 0.487) \wedge \\ (x_{29} < 1.066) \wedge (x_{31} > 0.416) \wedge (x_{34} < 1.008) \wedge \\ (x_{37} < 1.331) \wedge (x_{40} < 0.941) \wedge (x_{41} > 0.654) \wedge \\ (x_{42} > 0.262) \wedge (x_{48} > 0.639) \wedge (x_{49} < 0.814) \end{array} \right\} \rightarrow y = 1.34 \pm 0.034$$

Рис. 2.12: Четыре наиболее пригодных набора условий, найденных ГА для системы Mackey-Glass при $\dot{A} = 150$ (Приведено из Meyer и Packard 1992.)

До какой степени результаты ограничены фактом, что только определенные условия применимы (то есть, условия, которые являются конъюнкциями диапазонов для независимых переменных)? Packard (1990) предложил более общую форму для условий, которая также разрешает дизъюнкцию; примером может служить:

$$\{[(3.2 \leq x_6 \leq 5.5) \vee (1.1 \leq x_6 \leq 2.5)] \wedge [0.2 \leq x_8 \leq 4.8]\}$$

где представлено два непересекающихся выбора условий для x_6 . Дальнейшее обобщение, предложенное Packard, предполагало разрешение дизъюнкции между наборами условий.

Какова степень успешности этого метода в других типах задач прогнозирования? Packard (1990) предлагает применить этот метод к таким задачам, как прогнозирование погоды, прогнозирование финансовых рынков, распознавание речи и визуальное распознавание образов. Интересно, в 1991 Packard оставил Физический факультет в Университете Иллинойса, чтобы помочь создать компанию для прогнозирования финансовых рынков (Prediction Company, в Санта-Фе, Нью-Мексико). На момент когда я пишу это (середина 1995), компания еще не предала гласности их результаты, но продолжала разработку.

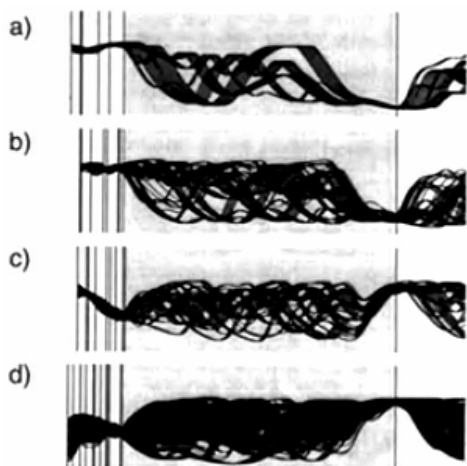


Рис. 2.13: Результаты четырех наиболее пригодных наборов условий, найденных ГА. (См. рис. 2.12.). Каждый график показывает траектории точек, удовлетворяющих набору условий. Самая левая белая область представляет начальные 50 шагов, на которых берутся данные. Вертикальные линии в этой области представляют различные условия \bar{x} для данного набора условий. Вертикальные линии в правой части представляют время для которого осуществляется прогноз. Заметьте, как сужаются траектории в этой области, означая что ГА нашел условия для хорошего прогноза. (Перепечатано из Martin Casdagli and Stephen Eubank, eds., *Nonlinear Modeling and Forecasting*; © 1992 Addison-Wesley Publishing Company, Inc. Печатается с разрешения издателя.)

Задача предсказания структуры белка

Одна из самых многообещающих и быстро растущих областей применения ГА – анализ данных и прогнозирование в молекулярной биологии. ГА использовался, между прочим, для интерпретации данных ядерно-магнитного резонанса, при определении структуры ДНК (Lucasius и Kateman в 1989), находя правильный порядок для неупорядоченной группы фрагментов ДНК (Parsons, Forrest, и Burks, в прессе), и предсказании структуры белка. Здесь я опишу один специфический проект, в котором ГА использовался, чтобы предсказать структуру белка.

Белки – фундаментальные функциональные структурные блоки всех биологических клеток. Главная цель ДНК в клетке состоит в том, чтобы закодировать инструкции для создания белков из аминокислот; белки в свою очередь выполняют большинство структурных и метаболических функций клетки. Белок составлен из последовательности аминокислот, связанных в соответствии с пептидными цепями. Длина последовательности изменяется от белка до белка, но обычно в пределах 100 аминокислот. Вследствие электростатических и других физических сил, последовательность "складывается" в специфическую трехмерную структуру. Именно эта трехмерная структура, прежде всего, определяет функцию белка. Трехмерная структура белка Крамбина (белок спермы, состоящий из 46 аминокислот) представлена на [рисунке 2.14](#). Трехмерная структура белка определена специфической последовательностью его аминокислот, но это в настоящее время не известно точно, как данная последовательность приводит к данной структуре. Фактически, способность предсказывать структуру белка из последовательности его аминокислот – одна из самых важных нерешенных проблем молекулярной биологии и биофизики. Мало того, что успешный алгоритм предсказания был бы огромным прогрессом в понимании биохимических механизмов белков, но такой алгоритм мог бы очевидно использоваться,

чтобы *проектировать* белки, выполняющие определенные функции, это будет иметь глубокие, далеко идущие эффекты на биотехнологию и лечение болезни.

Недавно было приложено значительное усилие к разработке методов, типа ГАмов и нейронных сетей для того, чтобы автоматически предсказывать структуры белка (см., к примеру, Hunter, Searls, и Shavlik 1993). Относительно простой ГА проект предсказания Steffen Schulze-Kremer (1992) иллюстрирует один путь, которым ГА может быть применен в этой задаче; это также иллюстрирует некоторые потенциальные ловушки.

Schulze-Kremer взял последовательность аминокислоты белка Крамбина и использовал ГА, для поиска в пространстве возможных структур ту, которая будет хорошо соответствовать последовательности аминокислоты Крамбина. Самый прямой способ описать структуру белка состоит в том, чтобы перечислить трехмерные координаты каждой аминокислоты, или даже каждого атома. В принципе, ГА может использовать такое представление, оперируя векторами координат, для нахождения одного, результирующего вероятную структуру. Но, из-за множества трудностей с таким представлением (например, обычный кроссовер и операторы мутации, слишком вероятно, создали бы физически невозможные структуры), Schulze-Kremer вместо этого описал структуры белка, используя "углы скручивания" – грубо говоря, углы, образованные пептидными цепями, соединяющими аминокислоты и углы, образованные цепями в аминокислотах ответвлений. (см. Dickerson и Geis (1969) для краткого обзора того, как трехмерная структура белка измеряется.) Schulze-Kremer использовал 10 углов скручивания, чтобы описать каждую из N (46 в случае Крамбина) аминокислот в последовательности для данного белка. Эта коллекция из N наборов 10 углов скручивания полностью определяет трехмерную структуру белка. Хромосома, представляя структуру кандидата с N аминокислотами, таким образом, содержит N наборов десяти вещественных чисел. Это представление проиллюстрировано на рисунке 2.15.

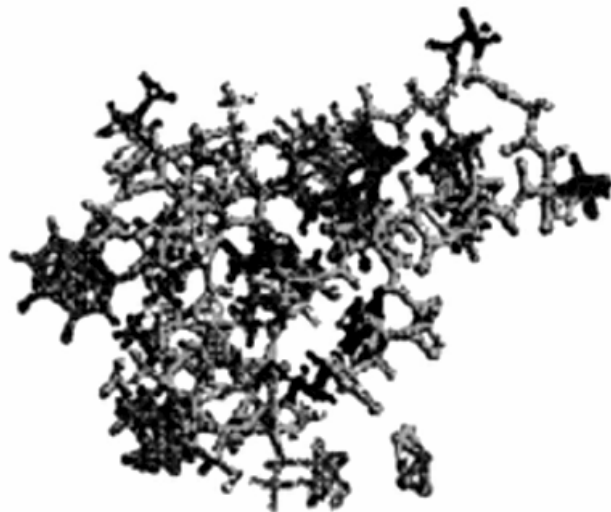


Рис. 2.14: Представление трехмерной структуры белка Крамбина. (Из "PDB at a Glance" URL страницы в Интернете http://www.nih.gov/molecular_modeling/pdb_at_a_glance.)

Следующим шагом является определение фитнес-функции в пространстве хромосом. Цель состоит в том, чтобы найти структуру, которая имеет низкую потенциальную энергию для данной последовательности аминокислот. Эта цель базируется при условии, что последовательность аминокислот свернется к состоянию минимальной энергии, где энергия – функция физических и химических свойств индивидуальных аминокислот и их пространственных взаимодействий (например, электростатические парные взаимодействия между атомами в двух пространственно смежных аминокислотах). Если бы полное описание существенных сил было известно и разрешимо, то в принципе

структура с минимальной энергией могла бы быть вычислена. Однако, практически эта проблема тяжела, и биологи вместо этого развивают приблизительные модели, чтобы описать потенциальную энергию структуры. Эти модели – чрезвычайно разумные предположения относительно того, какие силы являются наиболее значимыми. Начальные эксперименты Schulze-Kremer использовал чрезвычайно упрощенную модель, в которой потенциальная энергия структуры предполагалась как функция только углов скручивания, электростатических парных взаимодействий между атомами, и парными взаимодействиями van der Waals`а между атомами (Schulze-Kremer 1992). Цель ГА состояла в нахождении структуры (определенной в условиях углов скручивания), которая минимизировала функцию потенциальной энергии для последовательности аминокислот Крамбина.

В ГА Schulze-Kremer, кроссовер был произвольно двухточечный (то есть, выполненный в двух точках по хромосоме, вероятней чем в одной точке) или равномерный (то есть, вместо того, чтобы брать смежные гены от каждого родителя, для формирования потомка, каждый "ген" выбирался из одного или другого родителя, с 50%-ой вероятностью для каждого родителя). Здесь "ген" состоял из группы 10 углов скручивания; точки кроссовера выбирались лишь в границах аминокислоты. Два оператора мутации разрабатывались для работы с вещественными числами, а не с битами: первый заменял произвольно выбранный угол скручивания новым значением, произвольно выбранным из 10 наиболее часто встречающихся угловых значений для этой цепи, и второй увеличивал произвольно выбранный угол скручивания на небольшое значение.

Углы скручивания

аминокислота 1	аминокислота 2	...	аминокислота 46
φ : 66.3°	φ : -27.2°		
ψ : 45.2°	ψ : 23.1°		•
ω : 180.0°	ω : 180.0°		•
χ^1 : -22.7°	χ^1 : 111.4°		•
χ^2 : 127.1°	χ^2 : 120.2°		
χ^3 : -100.0°	χ^3 : -22.1°		
χ^4 : 32.2°	χ^4 : 32.2°		
χ^5 : -125.9°	χ^5 : -87.3°		
χ^6 : 55.4°	χ^6 : -95.2°		
χ^7 : 76.6°	χ^7 : -54.1°		

хромосома:

[66.3 45.2 180.0 -22.7 127.1 -100.0 32.2 -125.9 55.4 76.6] ...

Рис. 2.15: Иллюстрация представления для структуры белка, используемой в экспериментах Schulze-Kremer. Каждая из аминокислот N в последовательности представлена 10 углами скручивания (см. Schulze-Kremer 1992 для подробностей того, что представляют эти углы.) Хромосома - список этих N наборов из 10 углов. Точки кроссовера выбраны только в границах аминокислоты.

ГА, начатый на произвольно сгенерированной начальной популяции многих структур и запускался для 1000 поколений. В каждом поколении вычислялся фитнес (здесь, высокий фитнес означает низкую потенциальную энергию), популяция сортировалась по фитнесу, некоторое количество особей с самым высоким фитнесом отбиралось, чтобы стать родителями для следующего поколения (это, снова, форма отбора по ранжиру). Потомство создавалось кроссовером и мутацией. Использовалась схема, в которой вероятности различных мутаций и операторов кроссовера увеличивались или уменьшались в течение запуска. При разработке этой схемы, Schulze-Kremer полагался на свою интуицию, что операторы, вероятно, будут наиболее полезными на каждом из этапов запуска.

Поиск ГА произвел множество структур с весьма низкой потенциальной энергией фактически, намного ниже у фактической структуры Крамбина! К сожалению, однако, ни одна из произведенных особей не была структурно подобна Крамбину. Препятствием было то, что это было слишком легко для ГА, найти структуры низкой энергии для функции упрощенной потенциальной энергии; то есть, фитнес-функция не была в достаточной мере эффективной для того, чтобы вынудить ГА находить реальную целевую структуру. Факт, что начальные эксперименты Schulze-Kremer не были очень успешны, демонстрирует, как важно получить фитнес-функцию именно тут, понимая модель потенциальной энергии правильно (сложная биофизическая проблема), или по крайней мере получить достаточно хорошее приближение для ведения ГА в правильном направлении.

Эксперименты Schulze-Kremer – первый шаг в процессе "правильного понимания". Я предсказываю, что довольно скоро ГА и другие методы машинного обучения помогут биологам сделать по-настоящему крупные достижения в свертывании белка и в других областях молекулярной биологии. Я даже рискну предположить, что этот тип применения будет намного более выгодным (и с научной точки зрения и материально) чем использование ГА, для предсказания финансовых рынков.