

## Глава 2

### Лекция 4. Эволюционные алгоритмы

# Оглавление

<b>2</b>	<b>Лекция 4. Эволюционные алгоритмы</b>	<b>1</b>
4.1	Теорема об отсутствии бесплатных обедов . . . . .	2
4.2	Теория эволюционных алгоритмов (кратко) . . . . .	3
4.2.1	Теорема шаблонов . . . . .	3
4.2.2	Модели Липенса-Воса и Никса-Воса . . . . .	4
4.2.3	Модель Пругеля-Беннета . . . . .	5
4.2.4	Подход с использованием стат. механики . . . . .	6
4.2.5	Время сходимости операторов . . . . .	6
4.2.6	Оценка времени сходимости ЭА . . . . .	7
4.3	Алгоритмы оценки распределений . . . . .	7
4.3.1	PBIL + общая идея . . . . .	7
4.3.2	АОР для вещественного кодирования . . . . .	7
4.3.3	СМА-ES . . . . .	7
4.4	Системы обучающихся классификаторов . . . . .	7
4.4.1	Питтсбургский подход . . . . .	7
4.4.2	Мичиганский подход . . . . .	7

## 4.1 Теорема об отсутствии бесплатных обедов

Теорема об отсутствии бесплатных обедов (No Free Lunch, NFL) является одной из фундаментальных при анализе методов оптимизации и доказана Д. Волпертом и М. МакРеди в 1995 г. [11].

В вольном пересказе суть теоремы в следующем: *На полном множестве задач оптимизации ни один алгоритм оптимизации не является лучшим.* Другими словами, в среднем все оптимизационные алгоритмы равны между собой.

Теорема получила широкую известность и, на мой взгляд, отлично дополняет т.н. теорему о гадком утенке [10], которая утверждает, что среди алгоритмов классификации нет наилучшего, если опять-таки рассматривать полное множество задач. Критика NFL-теоремы заключается в том, что полное множество задач оптимизации – это практически бесполезное множество, т.к. в нем больше бессмысленных задач, которые никогда не встретятся в реальности, чем задач осмысленных и имеющих практическую пользу. А раз так, то обычно неважно, как ведет себя тот или иной оптимизационный метод на бесполезной задаче. С другой стороны, здравый смысл подсказывает, что чудес не бывает и универсальных методов, одинаково хорошо справляющихся с любыми задачами нет. У каждого метода есть свой (под)класс задач, которые у него получается хорошо решать, и есть задачи, которые для этого метода являются сложными, тогда как для другого метода они могут не представлять особенных трудностей.

Известна несколько иная формулировка данной теоремы, в которой множество задач оптимизации уже не является полным, а представляет собой замыкание относительно операции перестановки. По-простому это значит, что если у нас есть целевая функция  $f(\mathbf{x})$ , то если мы зададим новую функцию  $g(\mathbf{x}) = f(\mathbf{x})$  для всех  $\mathbf{x}$ , кроме двух точек:  $g(\mathbf{x}_1) = f(\mathbf{x}_2)$  и  $g(\mathbf{x}_2) = f(\mathbf{x}_1)$ , то функция  $g$  будет получена перестановкой относительно функции  $f$ . Все возможные перестановки и образуют наше новое множество, для которого было доказано, что на нем NFL-теорема также выполняется [7].

## 4.2 Теория эволюционных алгоритмов (кратко)

### 4.2.1 Теорема шаблонов

Один из первых теоретических результатов в области эволюционных алгоритмов, полученный лично Д. Холландом [2]. Выводится очень просто, достаточно знать основы комбинаторики и теории вероятностей, но несмотря на это, к сожалению, для многих это единственная известная теорема из области эволюционных вычислений, доказанная за почти 40-летнюю историю. Сама теорема имеет отношение только к ГА с пропорциональным отбором, одноточечным кроссинговером и битовой мутацией, без каких-либо «излишеств нехороших» вроде элитаризма, ламарковской модели эволюции и пр.

Суть сводится к следующему: *Доля особей, соответствующих некоторому шаблону изменяется с течением времени пропорционально относительной приспособленности этого шаблона.*

Что такое шаблон? *Шаблон* (англ. *schema*) – это строка, в которой могут быть нули, единицы и символы маски. Например, «100\*1» – шаблон, задающий строки

$$100001, 100011, 100101, 100111$$

У шаблона есть *определяющая длина*  $\delta(H)$  – расстояние между крайним левым и правым значащими разрядами, и *порядок*  $o(H)$  – количество значащих разрядов.

Обозначим  $P_c$  – вероятность скрещивания,  $f(H, t)$  – приспособленность шаблона в поколении  $t$  (среднюю приспособленность особей, подходящих под шаблон  $H$ ),  $\bar{f}(t)$  – среднюю приспособленность в популяции,  $m(H, t)$  – доля хромосом (бинарных строк) в популяции, содержащих шаблон  $H$ . Тогда имеет место следующая формула:

$$\langle m(H, t+1) \rangle \geq m(H, T) \frac{f(H, t)}{\bar{f}(t)} \left[ 1 - p_c \frac{\delta(H)}{l-1} \right]. \quad (4.1)$$

т.е. по сути эта формула позволяет оценить снизу долю особей, содержащих шаблон  $H$ .

Одним из недостатков формулы (4.1) является то, что в ней отсутствует влияние мутации на создание и разрушение шаблонов. Если считать, что вероятность мутации одного разряда равна  $P_m$ , а порядок шаблона равен  $o(H)$ , то вероятность того, что мутация не разрушит шаблон равна  $(1 - P_m)^{o(H)}$ . Т.е. если мутирующий разряд не попадает ни на одну фиксированную позицию внутри шаблона, то она не изменяется. С учетом этого исправленная теорема шаблонов выглядит следующим образом [3]:

$$\langle m(H, t+1) \rangle \geq m(H, T) \frac{f(H, t)}{\bar{f}(t)} \left[ 1 - p_c \frac{\delta(H)}{l-1} \right] (1 - P_m)^{o(H)} \quad (4.2)$$

Теперь немного критики. Как уже было отмечено, классическая теорема шаблонов неприменима к ГА, модели, которых отличаются от канонического ГА. Кроме этого в теореме шаблонов практически не учитывается то обстоятельство, что кроссовер и мутация могут не только разрушать шаблон, но создавать её из других шаблонов. Поэтому в теореме шаблонов присутствует знак неравенства. Ещё одним недостатком теоремы шаблонов является то, что она позволяет рассчитать долю шаблонов в популяции только для следующего поколения. Т.е. при попытке подсчитать число строк, соответствующих данному шаблону через несколько поколений с использованием теоремы шаблонов к успеху не приведет. Так получается, в частности, из-за пропорциональной стратегии отбора.

#### 4.2.2 Модели Липенса-Воса и Никса-Воса

Модель представлена в 1991 году [9] и использует аппарат марковских цепей для описания динамики ГА.

В модели рассматривается популяция бинарных строк фиксированной длины  $L$ . Для селекции применяется пропорциональный отбор, в результате скрещивания получается один потомок<sup>1</sup>, используется битовая мутация.

Одно из допущений, упрощающих задачу: рассмотрение популяции бесконечного размера, что позволяет игнорировать случаи потери значащих разрядов.

В поколении  $t$  популяция описывается вещественным вектором  $\mathbf{p}(t)$  длины  $2^L$ ,  $i$ -й компоненте которого соответствует  $i$ -я возможная комбинация нулевых и единичных разрядов, а сама компонента равна доле соответствующих особей в текущей популяции. Зная, как распределены особи в популяции, можно определить вероятность выбора каждой особи в результате отбора. Эта вероятность записывается в вектор  $\mathbf{s}(t)$ . В модели Воса и Липинса выводится оператор  $\mathcal{G}$  перехода, такой что:

$$\mathcal{G}(\mathbf{s}(t)) = \mathbf{s}(t+1),$$

и производится анализ свойств полученной модели. В частности производится анализ точек равновесия:

$$\mathcal{G}(\mathbf{s}(t)) = \mathbf{s}(t),$$

т.е. таких точек, в которых эволюция прекращается. Примером могут являться вырожденные популяции. Вос и Липинс показали, что в случае вырождения популяции неподвижная точка будет соответствовать локальному оптимуму. Ещё одним результатом стало то, что если к популяции применять только операторы скрещивания и мутации, то со временем популяция достигнет состояния, в котором все возможные строки будут присутствовать в популяции с равными долями<sup>2</sup>.

Модель Липинса-Воса является точной, однако она влечет громоздкие вычисления и есть трудности применения этой модели для других алгоритмов.

Модель для ГА, учитывающая конечный размер популяции, была предложена Никсом и Восом [4]. Эта модель использует аппарат цепей Маркова. В ней производится моделирование вероятности перехода от популяции, описываемой вектором  $\phi_i(t)$ , к популяции, описываемой вектором  $\phi_j(t+1)$ , где  $i, j$  – индексы в пространстве

<sup>1</sup>На самом деле, их два, но выбирается случайным образом только один. Это один из практически не встречающихся на практике вариантов, который однако позволяет упростить анализ, т.к. в результате для популяции из  $n$  особей будет необходимо  $n$  операций скрещивания.

<sup>2</sup>Предположение о бесконечной популяции все еще в силе.

популяций (не обязательно различные). Построенная модель позволила получить ряд интересных свойств ГА. В частности было показано, что при увеличении размера популяции модель Никса и Воса сходится по вероятности к модели Липенса-Воса. Если в модели Липенса-Воса имеются устойчивые равновесные точки (алгоритм сходится), то и модель Никса и Воса имеет такие же устойчивые точки.

### 4.2.3 Модель Пругеля-Беннета

Источник: [6]

- Опетак
- Фиксированный размер популяции
- Бинарная турнирная селекция
- Сначала битовая мутация
- Потом скрещивание (!), однородный кроссинговер
- Моделируемые параметры: средняя приспособленность и дисперсия
- Выделяются три режима
  - Начальный (переходный)
  - Медленная эволюция
  - Стационар
- Для решения проблемы флуктуаций результатов при разных запусках одновременно моделируется множество (ансамбль) популяций и рассматриваются макроскопические параметры.
- Модель нулевого порядка
  - Популяция бесконечна
  - Приспособленность распределена по гауссу
  - Нет кроссинговера
  - Получены уравнения, описывающие изменение среднего и дисперсии
- Уточнение модели
  - Для анализа используются кумулянты (полуинварианты), 1 (ср), 2 (дисп) и 3 (скос) порядка
  - Получены уравнения, описывающие изменение среднего и дисперсии
- Учет конечности популяции
  - Влияние на селекцию, но не на мутацию
  - Sampling
  - Получены уравнения, описывающие изменение среднего и дисперсии

- Последствия конечности – уменьшение дисперсии приспособленности
  - \* Можно сравнить с несмещенной формулой для дисперсии
  - \* Коррекция дисперсии зависит от модели ГА и типа селекции
- Учет флуктуаций
  - Ансамбль
  - Влияние флуктуаций и необходимые коррекции для модели не очень велики
- Рекомбинация
  - Усложняет анализ
  - Существенно уменьшает величину кумулянтов высокого порядка
  - Особи с одинаковой приспособленностью могут быть устроены по-разному
  - Анализ корреляции между строками
- Общий результат
  - Хорошее согласие с экспериментом для средней (100) и большой (1000) популяции
  - Расхождения для малой популяции (10 особей) из-за роста кумулянтов

#### 4.2.4 Подход с использованием стат. механики

*Для обобщения*

Следят за макроскопическими параметрами  
 Распределения!  
 Функционалы энергии и энтропии  
 Анализ влияния операторов на макро-параметры  
 Построение синтетической модели

#### 4.2.5 Время сходимости операторов

**Селекция** Источник [8]

- Пропорциональная (рулеточная):  $O(L \ln L)$
- Усечением:  $O(k\sqrt{L})$
- Турнирная:  $O(\sqrt{L})$

**Скращивание** Источник [5, 1]

- 1-точечный:  $O(L \ln L)$
- 2-точечный:  $O(L \ln L)$
- Однородный:  $O(\ln L)$

## Мутация

- Битовая мутация:  $O(\ln L)$

### 4.2.6 Оценка времени сходимости ЭА

## 4.3 Алгоритмы оценки распределений

### 4.3.1 RBIL + общая идея

### 4.3.2 АОР для вещественного кодирования

### 4.3.3 CMA-ES

## 4.4 Системы обучающихся классификаторов

### 4.4.1 Питтсбургский подход

Поиск набора правил поведения в явном виде

#### Samuel

1. Тестирование особи  $n$  раз и обновление приспособленности отдельных правил по результатам
2. Улучшение правил (мутация)
  - Удаление
  - Уточнение
  - Обобщение
  - Покрытие
  - Слияние
3. Тестирование особи  $m$  раз
4. Обновление приспособленности особи в целом

Информация о правиле

- Полезность
- Дисперсия полезности («стабильность»)
- Активность (как давно использовалось в последний раз)

### 4.4.2 Мичиганский подход

Популяция эволюционирующих правил

**ZCS**

- Устойчивая схема алгоритма
- Каждая особь – правило
- Для определения каждого действия агента формируется набор правил и набор действий
- Результативность правила влияет на приспособленность остальных правил. Вознаграждаются правила из набора с таким же действием
- Результат – вся популяция



# Литература

- [1] Цой Ю.Р. Один способ вычисления времени смешивания для генетических операторов скрещивания. In *Десятая национальная конференция по искусственному интеллекту с международным участием КИИ-2006 (25-28 сентября 2006 г., Обнинск): Труды конференции. В 3-х т.*, volume 3, pages 1047–1054, Москва, 2006. Физматлит.
- [2] J.H. Holland. *Adaptation in Natural and Artificial Systems*. The University of Michigan Press, 1975.
- [3] J.H. Holland. *Adaptation in Natural and Artificial Systems*. MIT Press, Cambridge, MA, second edition, 1992.
- [4] A.E. Nix and M.D. Vose. Modelling genetic algorithms with markov chains. *Annals of Mathematics and Artificial Intelligence*, 5:79–88, 1992.
- [5] A. Prugel-Bennett. The mixing rate of different crossover operators. In *Foundations of Genetic Algorithms 6*, page 261–274, 2001.
- [6] Adam Prugel-Bennett and Alex Rogers. Modelling ga dynamics. In *Theoretical Aspects of Evolutionary Computing*, pages 59–86, 1999.
- [7] C. Schumacher, M. Vose, and D. Whitley. The no free lunch and problem description length. In *Proceedings of Genetic and Evolutionary Computation Conference (GECCO-2001)*, pages 565–570. Morgan Kaufmann, 2001.
- [8] Dirk Thierens and David E. Goldberg. Convergence models of genetic algorithm selection schemes. In *Proceedings of the International Conference on Evolutionary Computation. The Third Conference on Parallel Problem Solving from Nature: Parallel Problem Solving from Nature*, PPSN III, pages 119–129, 1994.
- [9] M. D. Vose and G. E. Liepins. Punctuated equilibria in genetic search. *Complex Systems*, (5):31–44, 1991.
- [10] Satoshi Watanabe. *Knowing and Guessing: A Quantitative Study of Inference and Information*. New York: Wiley, 1969.
- [11] D. Wolpert and M. McReady. No free lunch theorems for search. Technical report SFI-TR-95-02-010, Santa Fe Institute, July 1995.