

# ОБ ОДНОМ СПОСОБЕ НАХОЖДЕНИЯ ТОЧЕК МНОГОКРАТНОЙ РАЗЛАДКИ В БИОЛОГИЧЕСКИХ ПОСЛЕДОВАТЕЛЬНОСТЯХ

Полушина Т.В.

Марийский Государственный Университет, Россия, Марий Эл, 424001, г. Йошкар-Ола,  
пл. Ленина, 1, tvpolushina@inbox.ru

Геномы сложных организмов содержат информацию, передающуюся из поколения поколение и обеспечивающую реализацию генетической программы развития организма. Основной ролью ДНК в клетках является хранение информации о структуре РНК и белков. ДНК представляет собой длинную полимерную молекулу, построенную из повторяющихся блоков, нуклеотидов. Каждый нуклеотид состоит из азотистого основания, сахара и фосфатной группы. В ДНК встречается четыре вида азотистых оснований: аденин, гуанин, тимин, цитозин. Азотистые основания одной из цепей соединены с азотистыми основаниями другой цепи водородными связями согласно принципу комплементарности: аденин соединяется с тимином, гуанин — с цитозином. Вышеописанные свойства означают, что последовательность азотистых оснований, входящих в ДНК, может рассматриваться как бинарная. В позиции, на которых размещены гуанин или цитозин будем записывать 1, а в позиции, на которых размещены аденин или тимин, — 0. Получим бинарную последовательность, состоящую из 0 и 1. Считаем что, что последовательность состоит из нескольких сегментов с неизвестными границами, распределение на каждом из которых различно. С математической точки зрения эта задача может быть описана как задача обнаружения многократной разладки [2]. В данной работе рассматривается применение генетического алгоритма [1] для обнаружения точек разладки. Опишем более формально данную задачу. Пусть  $b_1, b_2, \dots, b_L$  - бинарная последовательность длины  $L$ . Последовательность делится на несколько сегментов, где границы сегментов  $c_1, c_2, \dots, c_N$ . Эту последовательность будем использовать для формирования популяции в генетическом алгоритме. Для оценки целевой функции рассматриваем оценку максимального правдоподобия на каждом сегменте. В качестве примера генерируем случайную последовательность с известным распределением на каждом сегменте.

## Литература

1. *Goldberg D.* Genetic algorithms in search, optimization, and machine learning. – Massachusetts, Addison-Wesley, 1989. 414 p.
2. *G.E.Evans, G.Yu. Gofronov, J.M. Keith, D.P. Kroese.* Estimating change-points in biological sequence via the cross-entropy method // Ann.Oper.Res. DOI 10.1007/s10479-010-0687-0.