

БАЙЕСОВСКАЯ ОПТИМИЗАЦИЯ ДЛЯ РЕКОНСТРУКЦИИ ДЕМОГРАФИЧЕСКОЙ ИСТОРИИ ПОПУЛЯЦИЙ

Носкова Е.Э. (Университет ИТМО)

Научный руководитель – к.т.н., доцент ФИТиП Ульяновцев В.И.
(Университет ИТМО)

Байесовская оптимизация — одна из наиболее используемых техник машинного обучения для оптимизации сложно-вычислимых функций. Мы разработали и применили алгоритм, основанный на байесовской оптимизации, для задачи поиска параметров демографической истории популяций. Предложенный алгоритм продемонстрировал свою эффективность для данных четырех и пяти популяций.

Введение.

Демографическая история популяций - это история развития популяций, которая включает в себя такие параметры как численность популяций, время разделения, темпы миграции и отбора. Развитие методов биоинформатики сделало возможным реконструкцию демографических историй популяций по генетическим данным.

Существующие методы для вывода демографической истории сводят задачу к оптимизационной: требуется найти параметры демографической истории, которые максимизируют значение правдоподобия. Наиболее эффективный алгоритм оптимизации был предложен в программном обеспечении GADMA и основан на генетическом алгоритме. Однако вычислительная сложность целевой функции правдоподобия является экспоненциальной от числа популяций. Вывод параметров демографических историй четырех и пяти популяций остается вычислительно-трудной задачей.

Основная часть.

Байесовская оптимизация является одним из наиболее популярных алгоритмов оптимизации в условиях ограниченного бюджета времени. На каждой итерации она использует суррогатную модель для аппроксимации целевой функции, а также функцию выбора для принятия решения о следующей точке вычисления.

Мы разработали и применили алгоритм, основанный на байесовской оптимизации для задачи вывода параметров демографической истории. Была проанализирована эффективность алгоритма на различных данных, включающие в себя демографические истории четырех и пяти популяций. Было проведено сравнение сходимости байесовской оптимизации и генетического алгоритма по итерациям и по времени.

Выводы.

Предложенный метод, основанный на байесовской оптимизации, для вывода параметров демографической истории по генетическим данным показал себя эффективным и быстрым алгоритмом для четырех и пяти популяций. На всех рассмотренных данных алгоритм показал более быструю сходимость на первых итерациях, чем генетический алгоритм.

Носкова Е.Э. (автор)

Ульянцев В.И. (научный руководитель)