

ПРОГРАММНЫЙ КОМПЛЕКС ДЛЯ МНОГОМЕРНОГО АНАЛИЗА МИКРОЧИПОВЫХ ДАННЫХ: АЛГОРИТМЫ МНОГОМЕРНОГО АНАЛИЗА

Д. А. Полунин

Институт цитологии и генетики СО РАН
Новосибирский государственный университет

Существует большое количество специализированных программных средств, предназначенных для анализа микрочиповых данных, в том числе и многомерного. Однако, они не слишком удобны в использовании. Кроме того, некоторые из них являются коммерческими. Основная проблема таких программных средств – закрытая реализация, не позволяющая в полной мере реализовывать именно ту обработку, которая нужна пользователю. Кроме того, многие пакеты имеют собственные форматы данных, поэтому использование нескольких пакетов может быть крайне трудозатратным. Существуют и более универсальные средства, серьезным недостатком которых является тяжеловесный интерфейс. Поэтому разработка простого и удобного программного комплекса для анализа микрочиповых данных, доступного для пользователя-биолога без математического или технического образования, представляется весьма актуальной.

Исполняемые модули вызываются программой-диспетчером[1] во время исполнения пользовательского скрипта. Расширяемость комплекса обеспечивается благодаря тому, что каждому алгоритму соответствует исполняемый модуль, кроме того, предусмотрена совместимость с MS Excel, которая достигается за счет использования формата csv для хранения входных и выходных данных.

В процессе разработки были реализованы базовые алгоритмы многомерного анализа: метод главных компонент, метод главных координат, в основе которых лежит алгоритм сингулярного разложения матрицы, реализованный в отдельном модуле, дискриминантный анализ, неметрическое шкалирование, множественная регрессия.

1. И. А. Штайгер. Программный комплекс для многомерного анализа микрочиповых данных: интерфейс, входной язык, распараллеливание. МНСК-2012.

Научный руководитель – д-р биол. наук, доцент. В. М. Ефимов