

КЛАСТЕРИЗАЦИЯ ПРИЗНАКОВ В ЗАДАЧЕ ДИАГНОСТИКИ МИТОХОНДРИАЛЬНЫХ ЗАБОЛЕВАНИЙ

Борисов¹ А.П., Поворознюк¹ А.И., Филатова¹ А.Е., Гречанина² Ю.Б.

¹*Национальный технический университет «ХПИ», г. Харьков*

²*Харьковский специализированный медико-генетический центр,
г. Харьков*

Митохондриальные заболевания – одни из многих генетических заболеваний, исследованием которых занимается современная медицина. Диагностика этих заболеваний включает этап построения диагностической модели на основе анализа медицинских показателей.

Рассмотрена задача кластеризации множества медицинских показателей. Исходные экспериментальные данные для рассматриваемой задачи предоставлены Харьковским специализированным медико-генетическим центром. Особенность этих данных в том, что они узкоспециализированные и могут быть неполными. Для решения поставленной задачи обосновано применение алгоритма «Clore», характерными особенностями которого являются: работа с объектами, описанными номинальными признаками; возможность работы с объектами, описанными неполным набором признаков; зависимость трудоемкости реализации алгоритма только от мерности пространства признаков (трудоемкость не зависит от объема экспериментальной выборки).

Исходные данные представлены разного рода шкалами: числовыми, ординальными и номинальными. Используемый алгоритм кластеризации требует, чтобы все данные были представлены в номинальной шкале, к которой необходимо привести все анализируемые числовые данные. Рассмотрены и проанализированы следующие методы приведения числовой шкалы к номинальной:

- простое деление на равные интервалы на основе минимального и максимального значений;
- деление на интервалы на основе базовых статистических характеристик – математического ожидания m и среднеквадратического отклонения σ ;
- деление на интервалы на основе базовых статистических характеристик, но с предварительным выделением в отдельный интервал значений x_i , удовлетворяющих условию:

$$x_i > m + 3\sigma \text{ или } x_i < m - 3\sigma.$$

В результате реализации алгоритма «Clore» даны рекомендации по использованию методов приведения числовой шкалы к номинальной.