

УДК 004

МЕТОДЫ КЛАСТЕРНОГО АНАЛИЗА ДЛЯ ОБРАБОТКИ МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ

М. В. АЛЕКСЕЙКОВ

Научный руководитель А. И. ЯКИМОВ, д-р техн. наук, доц.

Белорусско-Российский университет

Могилев, Беларусь

Для формирования групп пациентов со сходными признаками заболевания или их отсутствием можно применить методы кластерного анализа многомерных данных. Большинство алгоритмов, при всём их разнообразии, основаны на гипотезе компактности, состоящей в том, что объекты одного класса по результатам измерений близко расположены по отношению друг к другу, а объекты разных классов заметно отличаются между собой.

Методы разбиения множеств на кластеры можно разделить на иерархические (подразделяют на агломеративные: Tree Clustering, CURE, CACTUS и дивизимные: Смита – Макнаотона и Кауфмана – Роузева) и неиерархические (подразделяют по статистикам: EM (Expectation Maximization), Scalable EM и по расстоянию, которые в свою очередь разделяют на сравнение по расстоянию до центра кластера: семейство K-Means и на сравнение по расстоянию между образцами: Fuzzy Relation Clustering). Семейство K-Means в свою очередь подразделяют на числовые: K-Means (Hard C-Means), Scalable K-Means, Fuzzy C-Means, Гюстафсона – Кесселя; категорийные: K-Modes; смешанные: K-prototypes. Для исследования отобраны методы кластеризации Fuzzy Relation Clustering, Tree Clustering, Fuzzy C-Means, EM (Expectation Maximization), Смита – Макнаотона, которые являются представителями основных видов методов кластеризации для разделения множества исследуемых данных.

Метод нечеткого кластерного анализа Fuzzy Relation Clustering (FRC) имеет следующие особенности: не требует охвата каждого кластера отдельным выпуклым множеством, т. е. не накладывает ограничения на геометрическую форму кластеров; характеризуется трудоемкостью $O(m^4)$ от числа элементов m .

Метод нечеткого кластерного анализа Fuzzy C-Means является представителем семейства неиерархических методов кластерного анализа группы K-Means и, следовательно, накладывает ограничения на геометрию кластеров, требуя охвата каждого кластера отдельным выпуклым множеством. Трудоемкость метода меньше трудоемкости метода нечеткого кластерного анализа Fuzzy Relation Clustering (FRC) и, следовательно, скорость работы Fuzzy C-Means превысит скорость работы FRC на больших объемах данных.