

УДК 303.722.4

ГРУППИРОВКА ОБЪЕКТОВ СО СХОЖИМИ ХАРАКТЕРИСТИКАМИ И ЛЮБЫМ ЧИСЛОМ НЕРАВНОЗНАЧНЫХ ПРИЗНАКОВ

П. А. МИХАЛЕНКО

Научный руководитель И. Г. ПЛИСКО

Белорусско-Российский университет

Могилев, Беларусь

Часто на практике встает задача группировки объектов со схожими признаками. Если число признаков мало и их количество не изменяется, и они разнозначные, то можно разбить объекты по группам, например, графически. Но чаще возникает проблема сгруппировать объекты с любым количеством неравнозначных характеристик.

Поэтому началась разработка приложения, позволяющего решать подобную задачу: разбить данные с любым числом признаков (характеристик), при этом можно не учитывать приоритет характеристик (вес характеристик может быть равным) либо учитывать приоритет характеристик (т. е. вес каждой характеристики можно задать).

Данное приложение написано на языке C#. Пользовательский интерфейс реализован с помощью технологии платформы .NET Windows Forms, а алгоритм метода кластеризации реализован в библиотеке классов .NET Framework.

Для разделения объектов со схожими признаками на группы (кластеры) использовался метод кластерного анализа, в частности – метод максимина.

Кластерный анализ – многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в сравнительно однородные группы.

Метод максимина предназначен для разделения объектов на кластеры, причем количество кластеров заранее неизвестно, оно определяется автоматически в процессе разбиения объектов.

Принцип работы метода следующий. Выбирается один из объектов (любой); он становится прототипом первого кластера. Находится объект, наиболее удаленный от выбранного; он становится прототипом второго кластера. Все объекты распределяются по двум кластерам; каждый объект относится к кластеру, представленному ближайшим прототипом. Затем в каждом из кластеров находится объект, наиболее удаленный от своего прототипа. Если расстояние между этим объектом и прототипом кластера оказывается значительным (превышающим некоторую предельную величину), то объект становится новым прототипом, т. е. образуется новый кластер. После этого распределение объектов по кластерам выполняется заново. Процесс продолжается, пока не будет получено такое разбиение на кластеры, при котором расстояние от каждого объекта до прототипа кластера не будет превышать предельную величину.

Разработанное приложение можно применить, например, для мониторинга успеваемости студентов кафедры.