

ЗАДАЧА КЛАСТЕРИЗАЦИИ В УСЛОВИЯХ НЕЧЕТКИХ ИСХОДНЫХ ДАННЫХ

Раскин Л.Г., Карпенко В.В.

*Национальный технический университет
«Харьковский политехнический институт»,
г. Харьков*

Стандартная задача кластерного анализа состоит в получении разбиения множества объектов кластеризации на подмножества. При этом качество кластеризации тем лучше, чем дальше друг от друга находятся полученные подмножества и чем ближе друг к другу объекты внутри каждого из них. Различные методы кластеризации различаются способом расчета расстояний между объектами и технологией организации процедуры кластеризации. Эти методы реально усложняются в связи с тем, что положения объектов в пространстве контролируемых, измеряемых координат могут быть заданы нечетко, причем законы распределения ошибок наблюдения, как правило, не известны. Адекватная модель этой ситуации будет получена, если аналитическое описание неопределенных координат объектов выполнить в терминах нечеткой математики, задав нечеткие числа их функциями принадлежности. Введем необходимые формальные соотношения.

Пусть x_{ij} , $i = 1, 2, \dots, n$, $j = 1, 2, \dots, n$, - нечеткое значение j -й координаты i -го объекта и $\mu(x_{ij})$ - функция принадлежности x_{ij} . Простейшую процедуру кластеризации построим следующим образом. На первом этапе выбирается рациональное число кластеров и отыскивается их положение. На втором этапе осуществляется последовательное присоединение каждого из объектов к тому из кластеров, расстояние до центра которого – минимально. Единственный проблемный элемент этой процедуры – сравнение нечетких расстояний от очередного объекта до центров кластеров и выбор минимального из них. Известные технологии решения этой задачи [1,2] сложны и неудобны при сравнении расстояний для большого числа объектов. Альтернативный подход состоит в следующем. Пусть $\mu(R_{ok})$ - функция принадлежности нечеткого расстояния R_{ok} от объекта i_o до центра k -го кластера. Задача состоит в сравнении нечетких чисел R_{ok_1} и R_{ok_2} , отображающих расстояния от кластеризуемого объекта i_o до центров кластеров k_1 и k_2 . Функций принадлежности $\mu(R_{ok_1})$ и $\mu(R_{ok_2})$ нормируются. Полученные неотрицательные функции, удовлетворяющие условию нормировки, могут быть интерпретированы как плотности распределения некоторых случайных величин и могут быть использованы для расчета вероятности того, что одно из этих чисел больше другого.

Литература:

1. Костенко. Ю.Г., Раскин Л.Г. Прогнозирование технического состояния систем управления. Ч.: Основа, 1996. – 303с.
2. Раскин Л.Г., Кириченко И.О., Серая О.В. Прикладное континуальное линейное программирование. – Х.: 2014.-292с.