Библиографический список

1. Белякин Н.В. Усиление одной теоремы Мостовского // Вестник Сибирского независимого института. – Новосибирск, 2010. –№1. – С. 51–74.

2. Роджерс Х. Теория рекурсивных функций и эффективная вычислимость. – М., 1972.

3. Ганов В.А., Карымов В.Р. Ограниченные вычисления с оракулами // Известия АлтГУ. – Барнаул, 2009. – №1(61).

4. Ганов В.А., Белякин Н.В. Общая теория вычислений с оракулом. – Новосибирск: ИМ СОАН СССР, 1989.

5. Карымов В.Р. Вычисления с оракулом и фиксированным ограничением на число вопросов // Известия АлтГУ. – Барнаул, 2010. – №1/2(65).

Нечеткие задачи кластеризации и представление их результатов

А.С. Герасимова

АлтГУ, г. Барнаул

Кластеризация – это задача разбиения заданного набора объектов на подмножества, причем требуется чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно различались.

Пусть при изучении n объектов $X_1,...,X_n$ у каждого из них измеряется p показателей. Если все показатели числовые, то, не теряя общности, можно отождествить объекты с точками в p-мерном евклидовом пространстве. В зависимости от метода кластеризации объектов, могут возникать разные результаты. В основе понятия нечетких множеств лежит представление о том, что составляющие данное множество элементы, обладающие общим свойством, могут обладать этим свойством в различной степени и, следовательно, принадлежать к данному множеству с различной степенью.

Допустим, что известно количество k кластеров, на которые требуется разбить заданные объекты. Каждому из объектов X_j ставится в соответствие номер кластера $f(j)$, к которому он относится. Если j-й объект отнесен к i-му кластеру, то это можно закодировать цепочкой символов 0 или 1 длины k, где единственная единица стоит на i-м месте. Переходя на язык нечетких множеств, такие цепочки можно интерпретировать как задание k просто устроенных функций принадлежности – вероятность того, что объект относится к i-му кластеру, есть 1, к любому другому – 0.

Итак, для каждого объекта X_j требуется указать k чисел $\mu_{j1},\ldots,\mu_{jk}$, которые являются значениями k функций принадлежности, соответствующим каждому из строящихся кластеров. Число μ_{ji} будет интерпретироваться нами как вероятность попадания j-го объекта в i-й кластер.

На данный момент существует много алгоритмов нечеткой кластеризации, но на взгляд автора, ни один из них никак не учитывает различие результатов кластеризации при подходе к этой задаче с различных точек зрения. Предлагается новый алгоритм, основанный на выделении «основных объектов» кластеров.

Пусть, как и раньше, имеется n объектов $X_1,\ldots,X_n$, каждый из которых задан p признаками. Рассмотрим различные «правдоподобные» разбиения наших объектов на k кластеров. Если при применении N разных способов j-й объект оказался в i-м кластере mij раз, то положим

$$\mu_{ij} = \frac{m_{ij}}{N}. \qquad (1)$$

Это число является естественной оценкой соответствующей вероятности попадания j-го объекта в i-й кластер.

Одной из проблем, с которой здесь может столкнуться исследователь, является то, что номера кластеров задаются достаточно произвольно. Поскольку мы собираемся использовать разные способы разбиения и хотим их сравнивать, возникает проблема «универсальной нумерации кластеров». Рассмотрим один вариант такой нумерации более подробно.

Переберем все «стартовые» разбиения заданных объектов на k кластеров. Располагая априорной информацией о данных либо из наглядного восприятия картины множества объектов, выделим k основных объектов, заведомо относящихся к разным кластерам, и обозначим их $\hat{X}_1,\ldots,\hat{X}_k$. Кластеру, к которому принадлежит объект $\hat{X}_j$ в l-том разбиении, присвоим номер j, $j=1,\ldots,k$, $l=1,\ldots,N$. Далее находим числа mij как количество попаданий j-го объекта в i-й кластер, после чего строим функцию принадлежности по формуле (1).

Пусть мы успешно построили соответствующие каждому объекту и каждому кластеру функции принадлежности. Нарисуем двумерную картинку объектов согласно одному из современных методов визуализации и будем красить точки-объекты в разные цвета. По мере увеличения значения μ_{ji} от 0 до 1 цвет точек будем менять по шкале радуги от фиолетового до красного.

При этом возможно построить, по крайней мере, k+1 раскрашивание. Из них k раскрашиваний получаются путем использования для раскраски функции принадлежности каждого из кластеров. Смысл таких раскрашиваний – определение степени четкости того из кластеров, который для нас служит основным. Для построения (k+1)-го раскрашивания в качестве функции принадлежности будем использовать наибольшее значение из k функций принадлежности каждого объекта. При этом если объект с большой вероятностью постоянно относится к одному и тому же кластеру, то ему будет соответствовать точка цвета, близкого к красному. По результатам итогового раскрашивания можно судить об устойчивости выбранного кластерного алгоритма по отношению к изучаемому набору данных.

Имитационное моделирование оценок полезностей инвестиционных проектов при проведении экспертизы

Е.В. Данько
АлтГУ, г. Барнаул

Важным элементом при определении информационной полезности экспертизы инвестиционных проектов, является моделирование оценок рассматриваемого проекта до непосредственного проведения экспертизы.

Пусть у инвестора имеется инвестиционный проект, границы его чистого приведенного дохода определяются отрезком $[NPV_1; NPV_2]$, для упрощения считаем, что распределение дохода внутри отрезка – равномерное и $NPV_1 < 0$, и $NPV_2 > 0$. Имеется возможность провести экспертизу, уменьшающую длину этого отрезка неопределенности. Инвестору известна стоимость проведения экспертизы E и погрешность Δ, равная половине длины отрезка $[NPV_1^E; NPV_2^E]$.

Для оценки инвестиционного проекта используется функция:

$$U = (1+\gamma)\frac{NPV_2^E}{2}\frac{NPV_2^E}{NPV_2^E - NPV_1^E} + (1+2\beta)\frac{NPV_1^E}{2}\frac{-NPV_1^E}{NPV_2^E - NPV_1^E}, \quad (1)$$

где γ – коэффициент упущенной выгоды $0 \le \gamma \le 1$, β – коэффициент, учитывающий дополнительный страх риска $\beta \ge 0$.