Секция 2. ГЕОМЕТРИЯ И АНАЛИЗ

О римановых многообразиях с тривиальной целой частью в разложении тензора кривизны

О.П. Гладунова, Е.Д. Родионов, В.В.Славский
АлтГУ, АлтГПА, г. Барнаул

Данная работа продолжает исследования, начатые в [1, 2]. В ней изучаются римановы многообразия с теми или иными ограничениями на целую часть разложения тензора кривизны в прямую сумму произведения Кулкарни-Номидзу тензора одномерной кривизны с метрическим тензором и тензора Вейля. Строятся примеры римановых метрик отрицательной кривизны Риччи и осциллирующей одномерной кривизны.

Настоящие исследования подержаны РФФИ (гранты № 08-01-98001, №10-01-90000-Бел_а), Советом по грантам Президента РФ для поддержки молодых ученых и ведущих научных школ РФ (грант № НШ-5682.2008.1), а также ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 гг. (гос. контракт № 02.740.11.0457).

Библиографический список

1. Rodionov E.D., Slavsky V.V. Conformal deformation of the Riemannian metrics and homogeneous Riemannian spaces // Comm. Math. Univ. Carolinae. – 2002. – V. 43, – №2. – P. 271–282.

2. Гладунова О.П., Родионов Е.Д., Славский В.В. Об операторе кривизны на четырехмерных группах Ли с левоинвариантной римановой метрикой // Известия АлтГУ: математика и механика – Барнаул, 2010. – №1/2. – С. 29–33.

Об одном подходе к post-hoc проблеме в кластерном анализе

С.В. Дронов
АлтГУ, г. Барнаул

Алгоритмы кластерного анализа сегодня широко используются в качестве инструмента для обработки так называемых «сырых» данных. Хорошо обоснованное, значимое с точки зрения практика разбиение конечного множества объектов X на кластеры, как правило, использует достаточно большое число p характеристик изучаемых объектов. После построения соответствующего разбиения возникает вопрос – все ли использованные для построения кластеров характеристики одинаково важны для формирования полученной кластерной структуры? Ответ на этот вопрос обычно отрицательный, и возможность определить и отбросить малозначимые характеристики позволяет решить две проблемы:

1) оценить потери информации при практическом отнесении того или иного объекта к какому-то из кластеров в случае, если некоторые из характеристик оказываются недоступными для измерения;

2) выбрав две или три наиболее информативные характеристики из использовавшихся, и рассматривая их как координаты объектов множества X, получить визуализацию этих объектов с наименьшими искажениями имеющейся кластерной структуры.

Задачу определения информационной значимости каждой из характеристик, степени ее влияния на кластерную структуру множества X, будем называть post-hoc проблемой в кластерном анализе.

К этой проблеме ранее были предложены разные подходы (см. [1–3]). Наиболее достойный, на взгляд автора, алгоритм предполагает рассмотрение номера (или нечисловой метки) кластера как уровня значения некоторого нечислового фактора. После этого степень влияния такого фактора на каждую из рассматриваемых характеристик изучается с помощью аппарата однофакторного дисперсионного анализа (ANOVA). Этот подход, в частности, реализован в статистическом пакете PASW (SPSS в ранних версиях). Можно, тем не менее, указать несколько очевидных недостатков этого подхода, таких, как необходимая нормальность распределений характеристик, однородность их дисперсий, а также тот факт, что изучается зависимость среднего значения характеристики от метки кластера, а не наоборот. При работе с медицинскими данными предположения ANOVA часто нарушаются, а незначительные объемы этих данных не позволяют подобные нарушения игнорировать.

В докладе предложен более прямой и естественный подход к проблеме, не требующий проверки никаких априорных предположений и нечувствительный к незначительности объемов экспериментальных данных.

Предлагаемый подход использует числовой коэффициент, оценивающий степень различия разбиений множества X на кластеры. Для его определения рассмотрим два кластерных разбиения этого множества – A и B. Положим

$$d(\mathrm{A},\mathrm{B}) = \sum_{x\in X} \rho(A_x, B_x)\,,$$

где A_x, B_x – множества из А и В соответственно, содержащие элемент x, а через $\rho(A,B)$ обозначено количество элементов, входящих в одно из множеств А или В, но не в другое (число элементов их симметрической разности).

Лемма. *Введенная величина d(А,В) является метрикой на семействе всех кластерных разбиений n-элементного множества X. При этом максимальное ее значение равно n(n-1) и достигается тогда и только тогда, когда одно из этих разбиений состоит лишь из одноэлементных множеств, а другое имеет в своем составе только множество X целиком.*

Теперь определим числовой коэффициент, который условимся называть коэффициентом кластерных различий. Положим для этого

$$K(\mathrm{A,B}) = \frac{d(\mathrm{A,B})}{n(n-1)}.$$

Теорема. *Коэффициент кластерных различий принимает значения от 0 до 1. Он равен нулю лишь для совпадающих разбиений. Чем больше его величина, тем менее похожи кластерные разбиения.*

Теперь сформулируем алгоритм решения post-hoc проблемы для кластерных разбиений. Пусть при помощи набора p характеристик объектов множества X построено его разбиение А. Исключим из набора характеристик j-ю характеристику, и, используя тот же кластерный алгоритм, что был использован для построения А, по сокращенному набору построим новое разбиение A_j. Вычислим $K_j = K(\mathrm{A,A}_j)$ и повторим эту процедуру для j=1,…,p. Расположим вычисленные коэффициенты по убыванию и будем называть полученный ряд и тот ряд, что будет строиться из него в соответствии с излагаемым далее алгоритмом, рядом сравнений.

Кажется совершенно понятным, что чем раньше в ряду сравнений встречается коэффициент, соответствующий какой-либо характеристике, тем более важную роль в построении кластеров она играет. Но из этого правила есть одно исключение. Действительно, если несколько характеристик тесно связаны по отношению к влиянию на кластерную структуру, то при удалении одной из них остальные компенсируют ее влияние, а значит, коэффициент K_j, соответствующей этой характеристике, окажется близким к нулю (или точно равным ему).

Поэтому далее рассмотрим набор S всех тех характеристик, которые обладают малыми или нулевыми коэффициентами K_j. Будем последовательно исключать из набора исходных характеристик всевозможные подмножества характеристик $T \subset S$, состоящие, по крайней мере, из двух элементов, и строить по оставшимся новые кластерные разбиения. Коэффициенты кластерных различий между ними и исходным разбиением А будем обозначать K_T. Признаком того, что мы нашли набор связанных между собой и значимо влияющих на кластерную структуру характеристик будет то, что коэффициент K_T окажется в ряду сравнений между двумя «индивидуальными» коэффициентами: $K_i > K_T > K_j$, причем последние существенно отличаются от 0.

В этой ситуации следует считать все характеристики, составляющие T, равно значимыми и помещающимися по значимости между i-й и j-й характеристиками. Дополним ряд сравнений «коллективными» коэффициентами K_T. Первая из поставленных задач решена.

Если же решается задача сокращения размерности с точки зрения наименьшего искажения кластерной структуры, то сначала следует определиться с размерностью итоговой задачи q (обычно q=2, реже 3). Затем в качестве новых координат выбираются q характеристик, коэффициенты кластерных различий которых возглавляют ряд сравнений. Если в число выбранных попали «групповые» коэффициенты K_T, то из каждого такого множества T характеристик в список новых координат берется одна (любая).

Библиографический список

1. Burns R.P., Burns R. Business Research Methods and Statistics Using SPSS. SAGE Publications Ltd, 2008.

2. StatSoft Electronic Statistics Textbook http://www.statsoft.com/textbook/cluster-analysis/.

3. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: классификация и снижение размерности. М.: Финансы и статистика, 1989.

Формула взаимной зависимости сторон треугольника с вершинами в замечательных точках ортоцентрического симплекса

А.А. Дудкин

АлтГПА, г. Барнаул

Получено следующее утверждение для ортоцентрического симплекса произвольной размерности $n \geq 2$ в евклидовом пространстве.

Теорема. Пусть R и O – радиус и центр описанной сферы ортоцентрического симплекса размерности n в евклидовом пространстве, r и T