2. Если $I_{x_i} = I_{x_{i-1}}$, то $i = i + 1$,и переходим к шагу 1, иначе:
3. Заменяем все значения в I_x левее I_{x_i} на I_{x_0}, справа – на $1 - I_{x_0}$.
4. В категорию X с номером I_{x_0} относим объекты с начала шкалы до последнего измененного значения. Найдем S_{00}. Заменяем все элементы правее $I_{x_{i-1}}$ на элементы из I_x' правее $I_{x_{i-1}}'$.

5. Увеличилось ли $|S_{00}|$? Если да, то запоминаем его, $i = i + 1$. К шагу 1. Иначе конец алгоритма.

В докладе рассмотрены соответствующие примеры и представлена программа для ЭВМ, реализующая описанный алгоритм.

Библиографический список

1. Ростовцев П.С. Черно-белый анализ связи переменных // Социология: 4М (методология, методы, математические модели). – 1998. – №10.

УДК 519.237

Оценивание силы пост-кластерной связи между формирующими показателями

С.В. Дронов, К.А. Леонгардт
АлтГУ, г. Барнаул

Предположим, что в результате работы некоторого кластерного алгоритма или путем экспертных оценок множество из n объектов $A_i, i = 1,...,n,$ каждый из которых задан совокупностью своих формирующих показателей $x_i, i = 1,...,k,$ разбито на p кластеров. Полученное кластерное разбиение будем далее называть основным и считать, что оно является объективно правильным в рамках решаемого круга задач. Другими словами, мы полностью доверяем тому алгоритму или той экспертной группе, в результате работы которых было построено основное разбиение.

Каждому кластеру поставлено в соответствие некоторое число которое условимся называть его меткой. Обозначим этот (внешний для решаемой задачи) набор меток **A**. Будем считать, что набор меток **A** значимо связан с основным кластерным разбиением и назовем этот набор основным. Предположение о связи основного кластерного разбиения с основным набором меток всегда справедливо, если и разбие-

ние, и набор меток возникают в результате работы одного и того же алгоритма, и крайне правдоподобно, если присвоение меток производится добросовестным и квалифицированным экспертом уже по готовому разбиению.

После построения системы меток мы можем ввести кластерную переменную *f*, которая каждому из объектов ставит в соответствие метку того кластера, к которому этот объект принадлежит. Понятие кластерной переменной, по-видимому, впервые было введено в работе [1].

В [2] предложен метод построения, другой системы **B** меток кластеров $\overline{f}_i, i=1,...,p$. Если, как и выше, по набору **B** определить кластерную переменную, то **B** является решением задачи на максимум для суммы квадратов коэффициентов корреляции:

$$\sum_{i=1}^{k} \rho^2(x_i, \overline{f}) \to \max. \qquad (1)$$

Сам способ построения системы меток **B** организован, следовательно, так, что она наилучшим образом согласована с совокупностью формирующих показателей. Ясно, что она не обязана совпадать с системой меток **A**, и даже может оказаться совсем не похожей на нее. Этот факт можно интерпретировать как несогласованность набора формирующих показателей напрямую со структурой основного кластерного разбиения. Основной задачей работы является попытка устранения этого несогласования.

Для этого будем искать некоторое преобразование *g* (или систему преобразований – свое преобразование для каждого показателя) такое, что если метки некоторого нового набора **C** будут формироваться с помощью решения задачи (1) после замены x_i на $g_i(x_i), i=1,...,k$, то наборы меток **C** и **A** будут, если не совпадать, то быть близкими.

Тогда без ограничения общности можно считать, что преобразованные показатели уже хорошо согласуются с основным набором кластерных меток **A**, а, следовательно, их можно будет считать достаточно хорошо согласованными с основной кластерной структурой. Более того, и корреляционные связи между преобразованными показателями в определенном смысле можно рассматривать как новый вид связи, порождаемой этой структурой.

Тогда $\rho^2(g_i(x_i), g_j(x_j))$ – мера силы (степени) этой новой связи между и . Фактически, рассматривая такую меру, мы математически строго вводим новый вид связи между показателями. Это – корреляционная связь их специально подобранных преобразований. Назовем такой тип связи пост-кластерной связью.

Рассмотрим пример. Пусть основное кластерное разбиение строится с применением агломеративного кластерного алгоритма. При этом в качестве расстояния между кластерами Q_i, Q_j будем использовать

$$d(Q_i, Q_j) = \frac{n_i n_j}{n_i + n_j} \| Z_i - Z_j \|^2, \qquad (2)$$

где n_i, n_j количества объектов кластеров, Z_i, Z_j – их центры, а через $\|.\|$ обозначена евклидова норма в *k*-мерном пространстве признаков. Предположим, что на старте алгоритма, когда каждый кластер состоит из одного объекта, все они имеют метки 0. На каждом следующем шаге новый кластер образуется как объединение каких-то двух из имевшихся. Метку нового кластера тогда определим как сумму меток объединяющихся кластеров и расстояния *d* между ними.

Использование расстояния (2) вместо евклидова рекомендовано в [3] для того, чтобы визуализирующая работу алгоритма дендрограмма не имела самопересечений (в [3] они называются инверсиями). Для нашего примера *d* выбрано потому, что полученные описанным образом метки имеют прозрачный смысл.

Поясним это. Для кластера *Q* рассмотрим величину

$$S(Q) = \sum_{x \in Q} \| x - Z \|^2,$$

где *Z* – центр кластера. Будем называть ее изменчивостью набора формирующих показателей внутри кластера *Q*. Название можно объяснить, например, тем, что в случае, когда размерность *k*=1 и кластер *Q* состоит из *q* элементов, то $S(Q)$ совпадает с дисперсией единственного показателя с точностью до множителя 1/*q*.

Теорема 1. *При описанном построении системы меток метка каждого кластера будет равна изменчивости набора формирующих показателей внутри этого кластера,*

Имея в виду доказанную теорему, определим систему преобразований признаков соотношением

$$g_i(x_i) = (x_i - \bar{X}_i)^2, i = 1, ..., k, \qquad (3)$$

где $\bar{X}_i$ – среднее значение *i*-го показателя по всем изучаемым объектам. Следующее утверждение представляется очевидным.

Теорема 2. *Система меток, построенная путем максимизации суммы квадратов коэффициентов корреляции с преобразованными по формулам (3) показателями совпадает с системой меток иерархического алгоритма с точностью до масштабного множителя.*

В докладе приводятся наглядные примеры, иллюстрирующие

утверждение этой теоремы.

Как вытекает из теоремы 2, в случае, когда метки кластеров присваиваются в соответствии с выбранным вариантом иерархического алгоритма, пост-кластерная связь между показателями это корреляционная связь между их преобразованными вариантами (3).

Особо отметим, что фактически вид этой связи определяется не алгоритмом, а именно основной системой меток, а способ построения меток конкретизирован в примере только для того, чтобы обосновать выбор вида преобразований (3).

Библиографический список

1. Сазонова А.С, Дронов С.В. Обратная post-hoc задача кластерного анализа и ее применение к дискриминации данных // Вестник Тюменского государственного университета. – 2014. – № 7.

2. Dronov S.V., Sazonova A.S. Two approaches to cluster variable quantification. // Model Assisted Statistics and Applications. – 2015. – V. 10.

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

УДК 579.64

Псевдоримановы эйнштейново-подобные метрические группы Ли с метрикой алгебраического солитона Риччи

П.Н. Клепиков, Е.Д. Родионов
АлтГУ, г. Барнаул

Многообразия Эйнштейна являются важным классом (псевдо)римановых многообразий, которые широко используются в геометрии и физике. Известно, что каждое многообразие Эйнштейна имеет параллельный тензор Риччи (т.е. $\nabla_X r = 0$). В последнее время активно изучаются различные обобщения многообразий Эйнштейна, одними из которых являются эйнштейново-подобные (псевдо)римановы многообразия в смысле А. Грея [1].

Определение 1. (Псевдо)риманово многообразие имеет циклично параллельный тензор Риччи (принадлежит к классу A), если

$$(\nabla_X r)(Y,Z) + (\nabla_Y r)(Z,X) + (\nabla_Z r)(X,Y) = 0$$