Решая неравенство $F(n) \le Y_0(n)$, получим, что $L_E(\varepsilon, n) = 1$ для $0 < n \le 10$.

Можно использовать этот способ нахождения критического значения ледж-коэффициента и далее, но получение формул для количества бинарных цепочек с r ошибками $Y_r(k,m)$ при $r \ge 2$, оказалось достаточно трудоемким, поэтому от дальнейшего исследования было решено отказаться.

На основе изложенного алгоритма была разработана компьютерная программа на языке Java, вычисляющая критические значения ледж-коэффициента $L_E(\varepsilon, n)$. Она, в том числе, подтвердила полученный выше теоретический результат для $0 < n \le 10$. С помощью этой программы создана таблица.

Таблица – Критические значения ледж-коэффициента для малых n

| $n$ | $L_E(\varepsilon, n)$ | $n$ | $L_E(\varepsilon, n)$ | $n$ | $L_E(\varepsilon, n)$ | $n$ | $L_E(\varepsilon, n)$ | $n$ | $L_E(\varepsilon, n)$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.0 | 6 | 1.0 | 11 | 0.86 | 16 | 0.73 | 21 | 0.67 |
| 2 | 1.0 | 7 | 1.0 | 12 | 0.83 | 17 | 0.73 | 22 | 0.67 |
| 3 | 1.0 | 8 | 1.0 | 13 | 0.8 | 18 | 0.7 | 23 | 0.67 |
| 4 | 1.0 | 9 | 1.0 | 14 | 0.78 | 19 | 0.69 | 24 | 0.65 |
| 5 | 1.0 | 10 | 1.0 | 15 | 0.75 | 20 | 0.67 | 25 | 0.65 |

**Библиографический список**

1. Дронов С.В., Петухова Р.В. Один вид связи между номинальной и бинарной переменными // Известия АлтГУ. – 2010.– Вып. 1/2 (65).– С. 34–36.

3. Мирмоминов Р.М. Исследование крайних случаев для оценки степени связи типа «ступенька» // Ломоносовские чтения на Алтае: фундаментальные проблемы науки и образования : сб. научных статей международной конференции, Барнаул, 11–14 ноября 2014 г. – Барнаул: Изд-во Алт. ун-та, 2014. – С. 171–173.

3. Дронов С.В. , Бойко И. Ю. Метод оценки степени связи бинарного и номинального показателей // ПДМ. – 2015. – № 4(30). –С. 109–119.

**УДК 519.25**

## Post-hoc оценка силы кластерной связи

***Е. А. Евдокимов***
*АлтГУ, г. Барнаул*

Интересной проблемой для специалиста любой области науки, где проводится кластерный анализ или классификация, может оказаться выявление неявных связей между показателями, задействованными в исследовании. Обнаружение таких связей может способствовать получению неожиданных результатов в своей области исследования, а также привести к понижению размерности задачи, то есть уменьшению количества параметров, которые стоит подробно изучать: если, например, удалось выявить сильную связь между какими-то двумя или несколькими показателями, то можно заменить их одним, универсальным. Такой подход позволит упростить сбор данных в дальнейшем, ускорить работу алгоритма кластеризации и сделать итог его работы в той или иной мере нагляднее. При этом кластерная структура множества изучаемых объектов после сокращения размерности не должна существенно измениться.

Понятие различия кластерных разбиений и коэффициент их сходства

Вслед за [1] определим расстояние между двумя кластерными разбиениями $G_A$ , $G_B$ одного и того же множества объектов $X$ формулой:

$$d(G_A, G_B) = \sum_{x \in X} |A_x \Delta B_x|, \quad (1)$$

где $|A_x \Delta B_x|$ – число элементов симметрической разности тех кластеров, в которые отнесен элемент x в этих двух разбиениях.

В качестве меры сходства $K_{G_A,G_B}$ кластерных разбиений $G_B$ и $G_A$ для множества из n элементов примем коэффициент, определенный как

$$K_{G_A,G_B} = 1 - \frac{d(G_A, G_B)}{n(n-1)}.$$

Этот коэффициент тем больше, чем более похожи разбиения, и лежит в диапазоне {0:1}.

Допустим, что имеющиеся у нас объекты характеризуются набором P своих показателей. Построенное с помощью набора P показателей кластерное разбиение обозначим

При решении задачи сокращения размерности мы будем менять набор тех показателей, по которым строится кластерное разбиение нашего множества. Вследствие этого, по новому набору показателей может получаться разбиение, отличающееся от Пусть по сокращенному набору показателей R, будет построено кластерное разбиение

Качеством кластерного разбиения мы будем называть коэффициент сходства между разбиением и исходным кластерным разбиением :

Кластерную силу i-го показателя определим как коэффициент качества кластерного разбиения G({Xi}), построенного с использованием только этого показателя, и обозначим эту силу :

Введенные коэффициенты отражают способность совокупности или отдельного показателя заменить собой весь их исходный набор при построении разбиения X на кластеры. Подобным образом мы можем оценивать способность одного из показателей замещать другой при построении кластерных разбиений; назовем эту способность кластерной связью между этими показателями.

Для оценки силы кластерной связи нужно провести разбиение по каждому показателю по отдельности и вычислить коэффициент сходства между ними. Коэффициент такой связи между показателями и будем вычислять:

По смыслу он похож на коэффициент парной корреляции, но описывает именно кластерную связь: между показателями может отсутствовать функциональная (и даже достаточно сильно выраженная корреляционная) связь, но, тем не менее, по отношению к кластерному разбиению они могут быть фактически идентичны. На рисунке 1 представлена простейшая ситуация, когда коэффициент связи равен единице: при проецировании объектов на оси показателей получаются одинаковые разбиения.

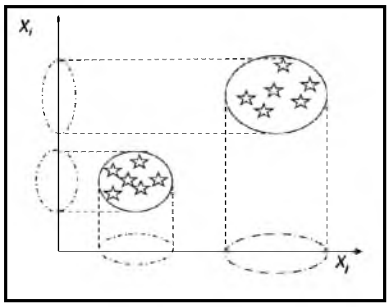


Рисунок 1 – $K_{\{i\},\{j\}} = 1$

Данный подход к понятию связи по сравнению с обычным обладает следующими преимуществами:

1. Оценка кластерной связи проводится post-hoc, то есть после выполнения конкретного алгоритма кластеризации, — это даёт нашему подходу большее прикладное значение.

2. С введением расстояния между разбиения по формуле (1) становится безразлична нумерация кластеров: мы получаем объективную оценку связи между показателями и их совокупностями в задачах, где кластеры не имеют порядка и заранее определенного смысла.

**Простой алгоритм снижения размерности**

С помощью определения кластерной силы и связи показателей мы можем выстроить простой алгоритм снижения размерности, настроенный на минимальное искажение исходной структуры кластеров.

Алгоритм:

1. Для каждой пары показателей вычисляем коэффициент кластерной связи.
2. Выбираем пару показателей с максимальным значением коэффициента связи.
3. Исключаем тот показатель, без которого качество разбиения страдает меньше.
4. Если качество разбиения получается ниже требуемого, то отменяем последнюю операцию и выходим из алгоритма.
5. Повторяем с пункта 2 для оставшейся совокупности показателей.

Приоритетной перспективой работы считаем нахождение формулы, приблизительно вычисляющей качество разбиения совокупности показателей по их коэффициентам силы и связи. Это по-

зволит находить нужную совокупность показателей, не используя для промежуточных вычислений алгоритм кластеризации.

**Библиографический список**

1. Дронов С.В., Дементьева Е.А. On a Coefficient of Cluster Differences and its Usage for Post-hoc Analysis in Clusterization. – Biomedical Soft Computing and Human Sciences. – 2012. – Vol. 18, №1. – С. 27–31.

**УДК 514.177.2**

## Замкнутая кривая данной длины, выпуклая оболочка которой имеет наибольший объем

***К.О. Кизбикенов***
*АлтГПА, г. Барнаул*

Широко известна задача о незамкнутой кривой данной длины, выпуклая оболочка которой имеет наибольший объем. Эта задача решена и ответом является виток винтовой линии. Аналогичная задача для замкнутых кривых, по-моему, до сих пор не решена.

С помощью программы Mathematica, было проведено численное моделирование. Пусть n это количество звеньев замкнутой ломаной. Искомую кривую будем искать как предел вписанных замкнутых ломаных, когда длины звеньев стремятся к нулю. Для $n = 5,6, \ldots 20$ были исследованы замкнутые ломаные с одинаковыми по длине звеньями, координаты вершин, которых считались неизвестными, кроме трех базисных. Был вычислен объем ее выпуклой оболочки, как функция координат вершин. При этом, длина каждого звена имела длину $\frac{1}{n}$. Задача свелась к поиску условного экстремума объема выпуклой оболочки при ограничении на длину ломаной. Эта задача была решена численными методами в программе Mathematica. При этом оказалось, что при всех значениях $n = 5,6, \ldots,20$ все концы ломаных лежали на некоторой винтовой линии.

Рисунок 1 – Выпуклая оболочка ломаной

Поэтому есть основания предполагать, что искомая кривая есть один виток винтовой линии вместе с отрезком, соединяющим концы этой линии. Осталось найти параметры этой кривой. Векторное уравнение винтовой линии имеет вид $r = \{a\cos t, a\sin t, b\, t\}$, где $t = 0 \ldots 2\pi$. Выпуклая оболочка одного витка винтовой линии имеет вид (рисунок 2).

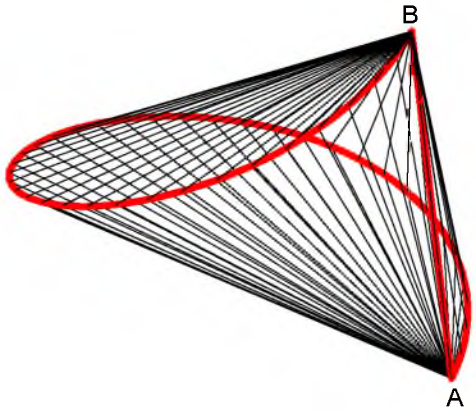


Рисунок 2 – Выпуклая оболочка кривой

По условию длина всей кривой равна 1. Поэтому

$$l = 2\pi\sqrt{a^2 + b\hat{}2} + 2\, b\, \pi = 1. \qquad (1)$$

Чтобы вычислить объем этой фигуры, заметим, что эта фигура описывается треугольником с основанием AB (рисунок 2) и вершиной, которая движется по винтовой линии. Площадь этого треугольника