10.37882/2223–2966.2021.04–2.11

КЛАСТЕРНЫЙ АНАЛИЗ ДАННЫХ МЕДИЦИНСКИХ ИССЛЕДОВАНИЙ С ИСПОЛЬЗОВАНИЕМ ЯЗЫКА R

Касюк Сергей Тимурович
К.т.н., доцент, ФГБОУ ВО «Южно-Уральский государственный медицинский университет» Министерства здравоохранения Российской Федерации (г. Челябинск)
sergey.kasyk@gmail.com
Диденко Галина Александровна
К.п.н., доцент, ФГБОУ ВО «Южно-Уральский государственный медицинский университет» Министерства здравоохранения Российской Федерации (г. Челябинск)
pga80@mail.ru
Степанова Оксана Александровна
К.п.н., доцент, ФГБОУ ВО «Южно-Уральский государственный медицинский университет» Министерства здравоохранения Российской Федерации (г. Челябинск)
okalst@mail.ru

CLUSTER ANALYSIS OF MEDICAL RESEARCH DATA WITH R

S. Kasyuk
G. Didenko
O. Stepanova

Summary. The article considers modern technologies of cluster analysis in medical research. Algorithms of fuzzy c-means clustering, hierarchical clustering, Kohonen neural network, and PAM are described. Formulas and appropriate functions of R language are considered. Examples of cluster analysis for breast cancer data sets are given.

Keywords: medical research, cluster analysis, fuzzy c-means clustering algorithm, hierarchical clustering algorithm, Kohonen neural network, PAM algorithm, R language.

Аннотация. В статье рассматриваются современные технологии кластерного анализа данных в медицинских исследованиях. Описываются алгоритмы нечетких с-средних, иерархической кластеризации, сетей Кохонена и PAM. Даются расчётные формулы и соответствующие функции языка R. Приводятся примеры кластерного анализа данных пациенток с раком молочной железы, решенные средствами языка R.

Ключевые слова: медицинские исследования, кластерный анализ, алгоритм нечетких с-средних, алгоритм иерархической кластеризации, сети Кохонена, алгоритм PAM, язык R.

Данная статья продолжает публикации [1, 2], посвященные современным технологиям статистического анализа данных в медицинских исследованиях. Цель статьи — дать обзор актуальных методов кластерного анализа с использованием статистического языка программирования R.

Кластерный анализ является начальным этапом статистического анализа, решающий задачу разбиения данных на группы «похожих» между собой объектов. В n-мерном метрическом пространстве признаков мерой «сходства» двух объектов считается расстояние между ними.

В статье приводятся различные примеры кластерного анализа данных пациенток с раком молочной железы, размещенные в онлайн репозитории машинного обучения The UCI Machine Learning Repository[1]. Однако авторы статьи не решали задачу выявления различий между полученными кластерами и выяснения причин

[1] https://archive.ics.uci.edu/ml/index.php

Таблица 1. Координаты центров кластеров

№	Age	BMI	Glucose	Insulin	HOMA	Leptin	Adipo-nectin	Resistin	MCP.1
1	–0,27	–0,64	–0,25	–0,36	–0,32	–0,51	–0,08	–0,15	–0,22
2	0,17	0,39	0,16	0,21	0,19	0,31	0,01	0,07	0,12
3	0,17	0,40	0,17	0,23	0,21	0,32	0,01	0,07	0,13

попадания пациенток в эти кластеры, поскольку данная задача лежит вне сферы их компетентности.

Для приведенных примеров *пропуски* в файлах данных были заменены на символы «NA». В дальнейшем, при обработке на языке R, данные *были очищены от пропусков* с помощью функции *na.omit* и *стандартизированы* с помощью функции *scale*.

Кластерный анализ с использованием алгоритма нечетких c-средних

Алгоритм кластеризации нечетких c-средних является нечеткой версией классического алгоритма k-средних и основан на минимизации целевой функции

$$\sum_i \sum_j w_i u_{ij}^m d_{ij}, \quad (1)$$

где w_i — вес наблюдения i; u_{ij} — членство наблюдения i в кластере j; d_{ij} — расстояние между наблюдением i и центром кластера j. Этот алгоритм использует такие метрики пространства, как евклидово расстояние и расстояние Манхэттена. Число кластеров C в алгоритме задается заранее. Параметр m определяет степень «нечеткости» [3, 4].

Шаги алгоритма нечетких *c*-средних следующие [4]

1. Инициализация матрицы **U**, определяющей принадлежность наблюдений к кластерам.
2. Вычисление координат центра для каждого кластера:

$$c_j = \sum_{i=1}^{N} u_{ij}^m x_i \Big/ \sum_{i=1}^{N} u_{ij}^m. \quad (2)$$

3. Корректировка матрицы **U** по следующей формуле:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}. \quad (3)$$

4. Повторение шагов 2 и 3 до схождения алгоритма. В итоге получается матрица **U** принадлежности i-го наблюдения j-му кластеру.

Пример кластеризации пациенток с раком молочной железы [5]

Университетский госпитальный центр г. Коимбры (Португалия) предоставил данные о 64 пациентках с раком молочной железы и 52 здоровых женщинах. В файле dataR2.csv[1]* содержатся следующие параметры:

1. *Age* — возраст;
2. *BMI* — индекс массы тела, кг/м^2;
3. *Glucose* — содержание сахара в крови, мг/дл;
4. *Insulin* — инсулин, мЕд/л;
5. *HOMA* — индекс HOMA;
6. *Leptin* — лептин, нг/мл;
7. *Adiponectin* — адипонектин, мг/мл;
8. *Resistin* — резистин, нг/мл;
9. *MCP*-1 — моноцитарный хемоаттрактантный белок 1, пг/дл;
10. *Classification* — классификационные метки (1 — здоровые; 2 — онкобольные).

Предварительно выполним отбор пациенток с раком молочной железы. Произведем разбиение данных на 3 кластера с помощью функции *cmeans* из пакета *e*1071. При этом максимальное количество итераций примем, равным 100, а степень «нечеткости» $m = 2$. Визуализируем полученные кластеры с помощью функции *clustplot* из пакета *cluster*.

Решение задачи на языке R:

```
> BCancer <- read.cv2("C:/Data/dataR2.csv")
> newBCancer <- subset(BCancer, Classification == 2)
> newBCancer <- newBCancer[,—10]
> result <- cmeans(scale(newBCancer), 3, 100, m = 2,
method = "cmeans")
> result
```

[1] http://archive.ics.uci.edu/ml/machine-learning-databases/00451/dataR2.csv

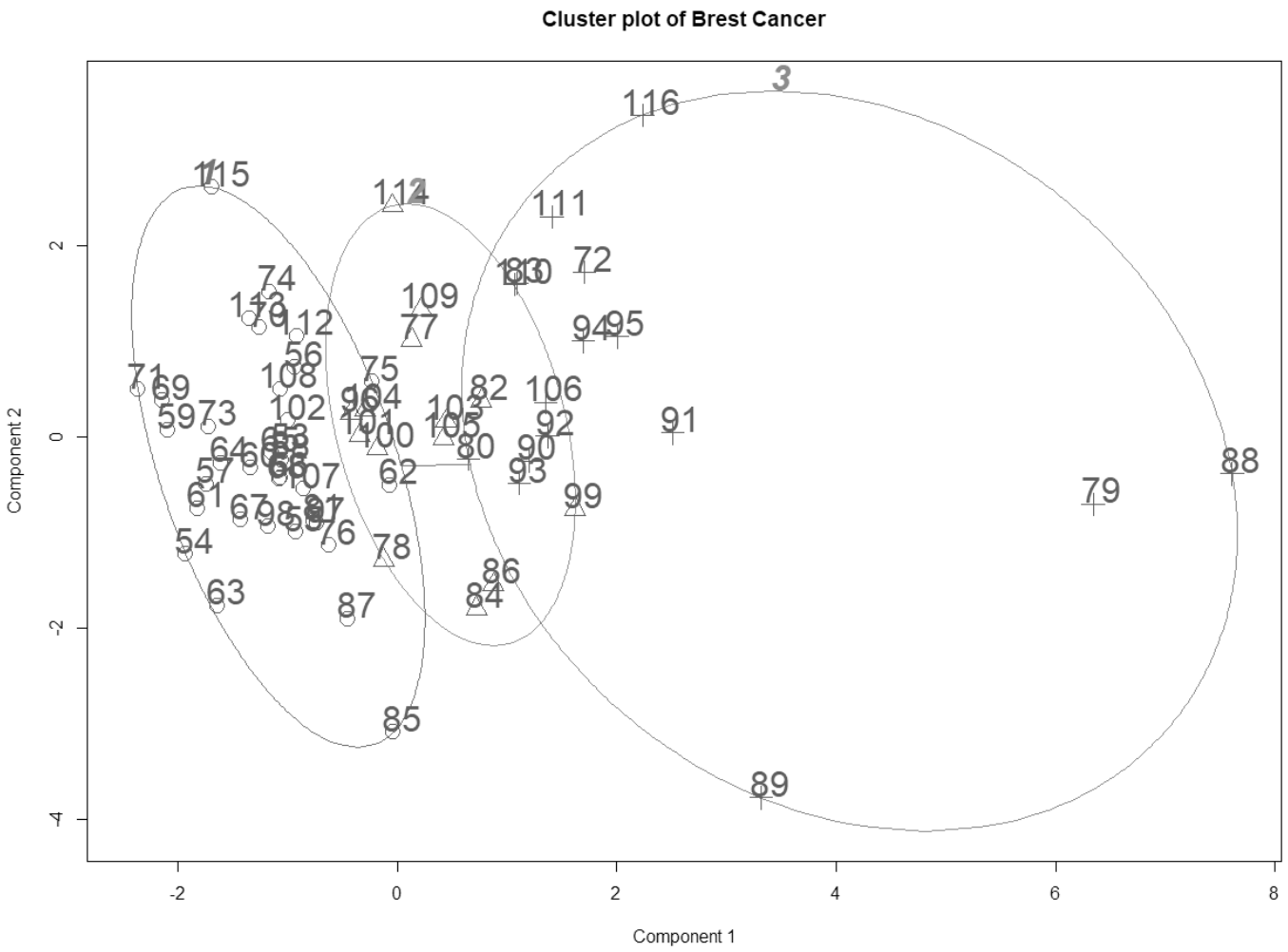

Рис. 1. Диаграмма разбиения данных пациенток на три кластера

> *clusplot*(*scale*(*newBCancer*), *result*$*cluster*,
main = "Cluster plot of Brest Cancer",
color = *TRUE*, *labels* = 2, *lines* = 2, *cex* = 1)

Визуализация результатов разбиения 64 пациенток с онкологией на 3 кластера приведена на рис. 1. В 1-й кластер попали 34 наблюдения, во 2-й — 14 наблюдений, в третий — 16 наблюдений. Полученные координаты центров кластеров представлены в табл. 1. Центры 2-го и 3-го кластеров практически совпали, и большинство наблюдений из этих кластеров объединяются в случае разбиения данных на 2 кластера.

Кластерный анализ с использованием алгоритма иерархической кластеризации

Алгоритм иерархической кластеризации объединяет объекты в достаточно большие кластеры с использованием некоторой меры сходства или расстояния между объектами. Результатом такой кластеризации является дендрограмма, построенная в координатах шагов либо расстояний [6].

Алгоритмы иерархической кластеризации делятся на агломерационные и разделительные в зависимости от способа выбора начального кластера или кластеров.

Шаги агломерационного алгоритма иерархической кластеризации следующие [6]:

1. Вначале все объекты считают отдельными кластерами.
2. Затем выбирается пара ближайших кластеров, которые объединяют в один кластер.
3. В дальнейшем связываются вместе все большее число объектов и агрегируются большее количество кластеров.
4. Шаг 3 повторяется до тех пор, пока все объекты не объединятся в один кластер.

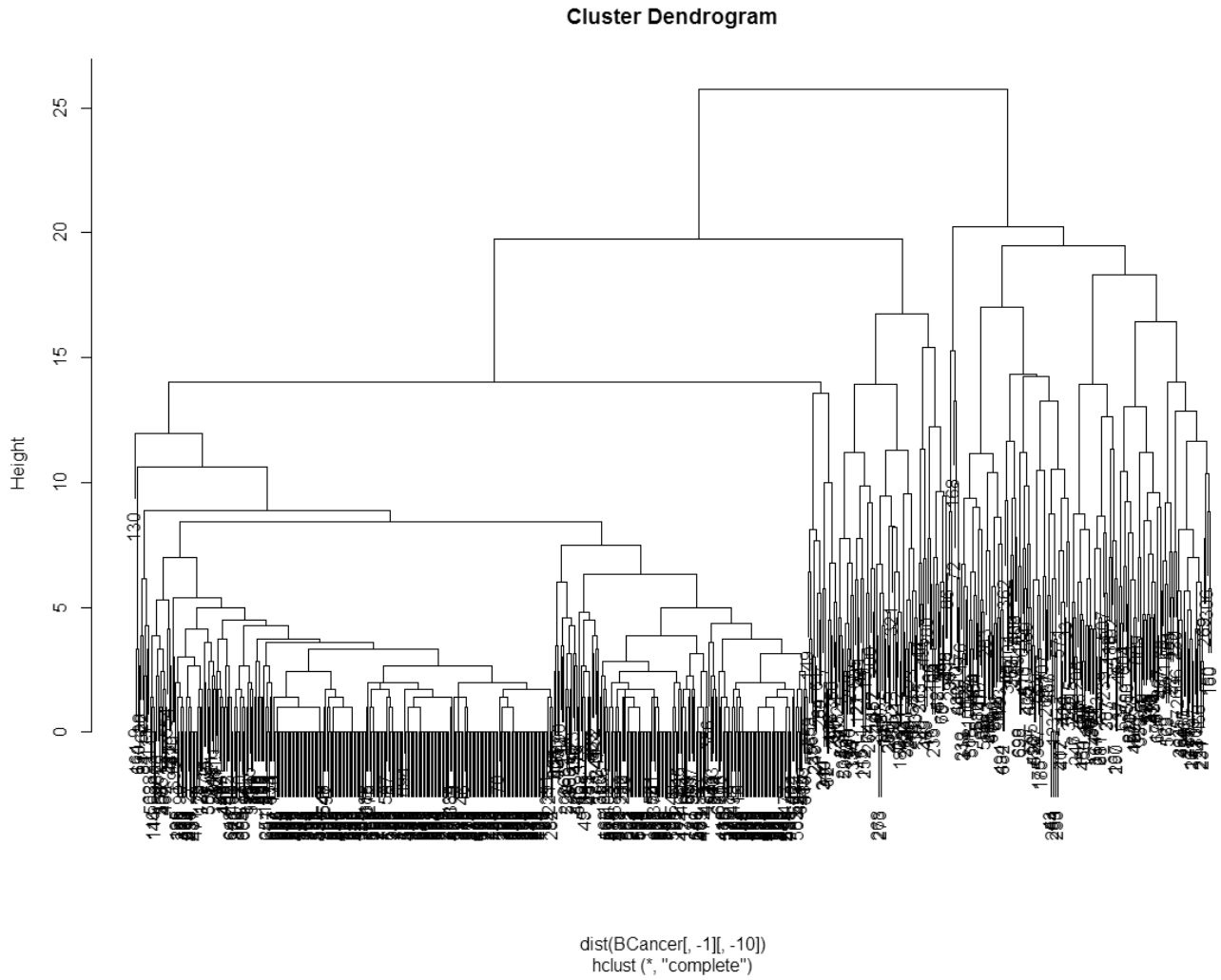

Рис. 2. Дендрограмма иерархической кластеризации

Пример кластеризации пациенток с раком молочной железы [7]

Клинический научный центр университета Висконсина (США) предоставил данные пациенток с раком молочной железы. В файле breast-cancer-wisconsin.data[1]* содержатся результаты биопсии с 9 ранговыми характеристиками новообразований для 1251 пациентки:

1. *ID number* — идентификационный номер;
2. *Clump Thickness* — размер образований (1–10);
3. *Uniformity of Cell Size* — однородность размера клетки (1–10);
4. *Uniformity of Cell Shape* — однородность формы клетки (1–10);
5. *Marginal Adhesion* — межклеточная мембранная адгезия (1–10);
6. *Single Epithelial Cell Size* — размер эпителиальной клетки (1–10);
7. *Bare Nuclei* — ядро клетки (1–10);
8. *Bland Chromatin* — деконденсированный хроматин (1–10);
9. *Normal Nucleoli* — нормальные ядра (1–10);
10. *Mitoses* — динамика митоза (1–10);
11. *Class* — диагноз (2 — доброкачественная опухоль; 4 — злокачественная опухоль).

Произведем *иерархическую кластеризацию* данных с помощью функции *hclust*, используя *метод полной связи (complite linkage)*. Выделим 4 кластера пациенток, используя функцию *cutree*.

Решение задачи на языке R:

> *BCancer* <- *read.table*("C:/Data/breast-cancer-wisconsin.data", *header* = *FALSE*, *sep* = ",")

[1] https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/breast-cancer-wisconsin.data

```
> BCancer <- na.omit(BCancer)
> result <- hclust(dist(BCancer[,-1][,-10]),
method = "complete", members = NULL)
> plot(result)
> resultcut <- cutree(result, 4)
> table(resultcut, BCancer[,11])
resultcut 2 4
1 452 50
2 2 71
3 3 75
4 1 45
```

После удаления пропусков количество наблюдений уменьшилось до 683. Среди пациенток с *доброкачественной опухолью* в 1-й кластер попали 98,7%. Пациентки со *злокачественной опухолью* были разбиты на четыре кластера следующим образом: 1-й кластер — 20,7%; 2-й кластер — 29,5%; 3-й кластер — 31,1%; 4-й кластер — 18,7%.

На дендрограмме, представленной на рис. 2, отчетливо видны пациентки с доброкачественной опухолью из 1-го кластера.

Кластерный анализ с использованием сетей Кохонена

Нейронные сети Кохонена позволяют распознавать кластеры в данных, а также устанавливать близость этих кластеров. Эти сети имеют два слоя: входной слой, содержащий по одному нейрону для каждой входной переменной; выходной слой, нейроны которого упорядочены, как правило, в одномерную или двухмерную решетку прямоугольной формы.

Обучается сеть Кохонена методом последовательных приближений. Начиная со случайным образом выбранного исходного расположения центров, алгоритм обучения постепенно улучшает его так, чтобы уловить кластеризацию обучаемых данных. Алгоритм обучения является итерационным, при этом нейроны входного слоя не участвуют в процессе обучения [8].

В языке R существует множество пакетов с различными алгоритмами самоорганизующихся карт: *kohonen*, *Multi-SOM*, *SOMbrero*, *som* и др. Так, в пакете *kohonen* реализованы как стандартные самоорганизующиеся карты Кохонена, работающие с числовыми данными, а также и «суперорганизующиеся карты», построенные на множественных параллельных картах. Функция *som* из этого пакета позволяет строить сети Кохонена по имеющимся данным для заданной топологической карты, при этом функции *somgrid* задает размер и структуру этой топологической карты, а результаты кластеризации выводятся при помощи функции *map* [9, 10, 11].

Пример кластеризации пациенток с раком молочной железы [12]

Национальный институт биомедицинской инженерии в г. Порто (Португалия) предоставил данные пациенток с раком молочной железы. В файле BreastTissue.xls[1]* содержатся результаты обследования 106 пациенток, включающие 9 характеристик электрического импеданса образцов ткани молочной железы:

1. *Class* — классы (*car* — карцинома; *fad* — фиброаденома; *mas* —местопатия; *gla* — железестый; *con* — соединительный; *adi* — жировой);
2. *I*0 — инпеданс на нулевой частоте;
3. *PA*500 — фазовый угол на частоте 500 кГц;
4. *HFS* — высокочастотный наклон (крутизна) фазового угла;
5. *DA* — расстояние импеданса между спектральными концами;
6. *AREA* — площадь области под спектром;
7. *A/DA* — площадь, отнесенная к величине *DA*;
8. *MAX IP* — максимум спектра;
9. *DR* — расстояние между *I*0 и реальной частью точки максимальной частоты;
10. *P* — длина спектральной кривой.

Предварительно исходный файл BreastTissue.xls преобразуем в текстовый с расширением «txt». Затем с помощью функции *som* из пакета *kohonen* произведем обучение сети Кохонена с прямоугольной топологической картой размерностью 3×2. Визуализируем кластеры с помощью функции *clustplot* из пакета *cluster*.

Решение задачи на языке R:

```
> BTissue <- read.table("C:/Data/BreastTissue.txt",
header = TRUE, sep = "\t")
> Class <- BTissue[, 1]
> BTissue <- BTissue[,—1]
> set.seed(1000)
> som.BTissue = som(scale(BTissue),
grid = somgrid(3, 2, "rectangular"))
> som.BTissue
> som.cluster = map(som.BTissue)
> som.cluster
> table(Class, som.cluster$unit.classif)
> plot(som.BTissue, main = "Breast Tissue: SOM")
> clusplot(scale(BTissue), som.cluster$unit.classif,
main = "Cluster plot of Brest Tissue",
color = TRUE, labels = 2, lines = 2, cex = 1)
```

[1] http://archive.ics.uci.edu/ml/machine-learning-databases/00192/BreastTissue.xls

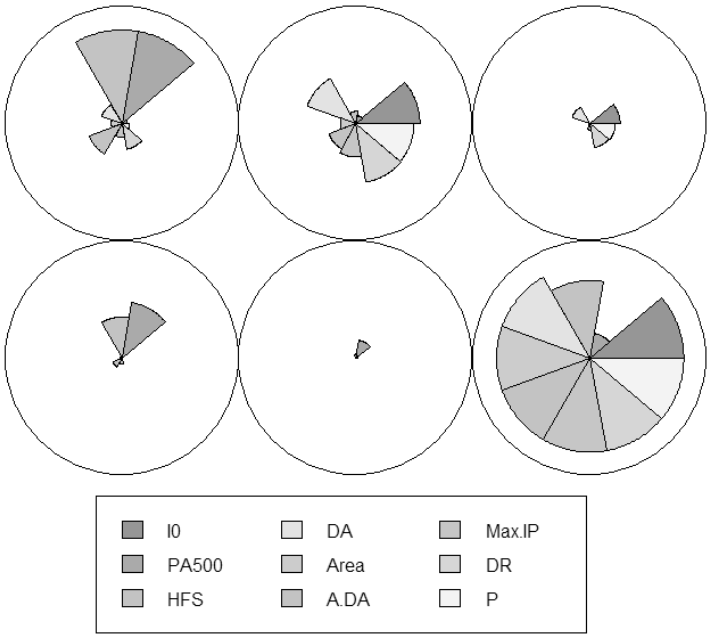

Рис. 3. Топологическая карта сети Кохонена

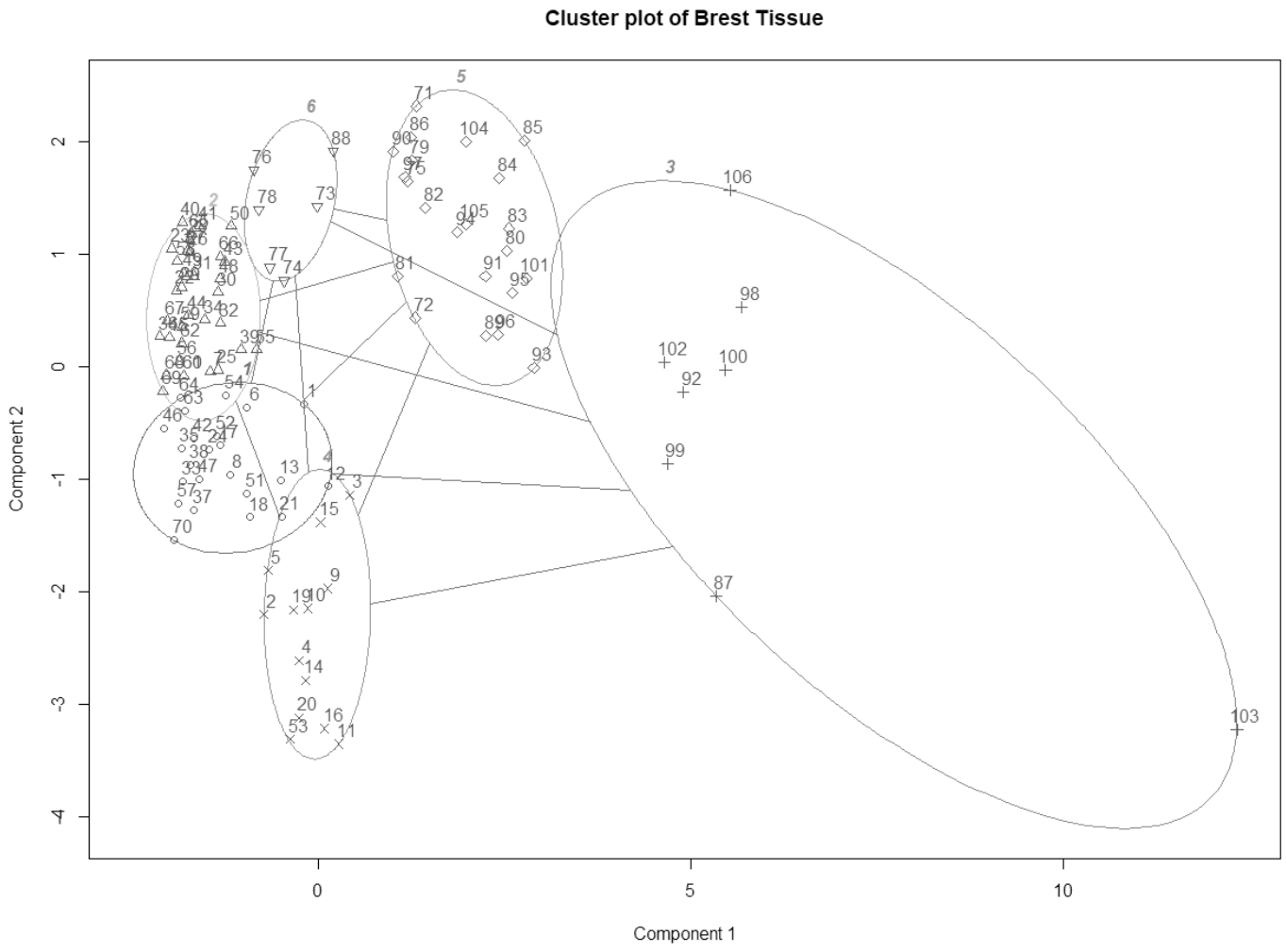

Рис. 4. Диаграмма разбиения данных пациенток на 6 кластеров

Class 1 2 3 4 5 6
adi 0 0 8 0 13 1
car 8 1 0 12 0 0
con 0 0 0 0 9 5
fad 3 12 0 0 0 0
gla 4 12 0 0 0 0
mas 8 9 0 1 0 0

В результате данные пациенток были разбиты на 6 кластеров и получена таблица соответствия классов образцов ткани и найденных кластеров. Для этого примера сеть Кохонена не дала такого разбиения объектов на кластеры, которое бы соответствовало классам образцов тканей молочной железы. Например, класс *gla* представлен во 1-м и 2-м кластере, класс *car* — в 1-м и 4-м кластере.

Результаты обучения нейронной сети в виде топологической карты размерностью 3×2 и визуализация полученных кластеров представлены на рис. 3 и 4.

Кластерный анализ категориальных данных с использованием алгоритма PAM

Кластерный анализ категориальных данных осуществляется с использованием метрики пространства с расстоянием Гауэра, для которой сходство между объектами i и j оценивается как среднее значение по всем возможным сравнениям:

$$S_{ij} = \sum_{k=1}^{v} S_{ijk} \Big/ \sum_{k=1}^{v} \ddot{a}_{ijk}, \quad (4)$$

где S_{ijk} — значение сходства между объектами i и j по параметру k, лежащее в диапазоне от 0 до 1; $\ddot{a}_{ijk}$ — весовой коэффициент, равный 1, если объекты i и j сравнимы по параметру k, и равный 0 в противном случае; v — количество параметров у объектов [13].

Алгоритм кластеризации PAM (partitioning around medoids) является менее чувствительной к выбросам модификацией алгоритма k-средних. Этот алгоритм работает модоидами и основан на минимизации целевой функции:

$$\sum_i \sum_j d(i, j) z_{ij}, \quad (5)$$

где $d(i, j)$ — мера расстояния между объектами j и i; z_{ij} — дихотомическая переменная, равная 1, если объект j назначен в кластер, к которому принадлежит объект i, и равная 0 в противном случае [14].

Шаги алгоритма PAM следующие [15]:

1. Выбор случайным образом k объектов в качестве медоидов.
2. Назначение для каждого объекта кластера, представленного ближайшим к этому объекту медоидом.
3. Нахождение для каждого кластера наблюдения, которое минимизирует среднее расстояние в случае, если бы его назначили медоидом, и последующая замена медоида в кластере на искомое наблюдение.
4. Возвращение к шагу 2, если хотя бы один медоид изменился, или завершение алгоритма в противном случае.

Пример кластеризации пациенток с раком молочной железы [16]. Институт онкологии г. Любляна (Югославия) предоставил данные о 286 пациенток с раком молочной железы. В файле breast-cancer.data[1]* содержатся следующие категориальные данные:

1. *Class* — классы (*no-recurrence-events* — без повторения; *recurrence-events* — повторяющееся событие);
2. *age* — возрастные группы (10–19, 20–29, 30–39, 40–49, 50–59, 60–69, 70–79, 80–89, 90–99);
3. *menopause* — предклимактерический или климактерический период (lt40, ge40, premeno);
4. *tumor-size* — размер новообразования (0–4, 5–9, 10–14, 15–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59);
5. *inv-nodes* — количество подмышечных лимфатических узлов, содержащих метастатический рак молочной железы, видимых при гистологическом исследовании (0–2, 3–5, 6–8, 9–11, 12–14, 15–17, 18–20, 21–23, 24–26, 27–29, 30–32, 33–35, 36–39).
6. *node-caps* — метастазы рака в лимфатические узлы (yes — да; no — нет);
7. *deg-malig* — степень злокачественности (1, 2, 3);
8. *breast* — грудь (*left* — левая; *right* — правая);
9. *breast-quad* — зоны груди (*left-up* — слева вверху; *left-low* — слева внизу; *right-up* — справа вверху; *right-low* — справа внизу; *central* — по центру);
10. *irradiat* — проведение лучевой терапии (*yes* — да; *no* — нет).

Предварительно выполним преобразование всех переменных к категориальному типу с помощью функции *as.factor*. Измерим расстояния Гауэра между объектами, используя функцию *daisy* из пакета *cluster*; построим матрицу отличий между объектами, используя функцию *as.matrix*. Разобьём данные на 5 кластеров с помощью функции *pam* из пакета *cluster* и визуализируем результат кластеризации с помощью функции *ggplot* из пакета *ggplot*, как это показано в работе [15].

[1] https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer/

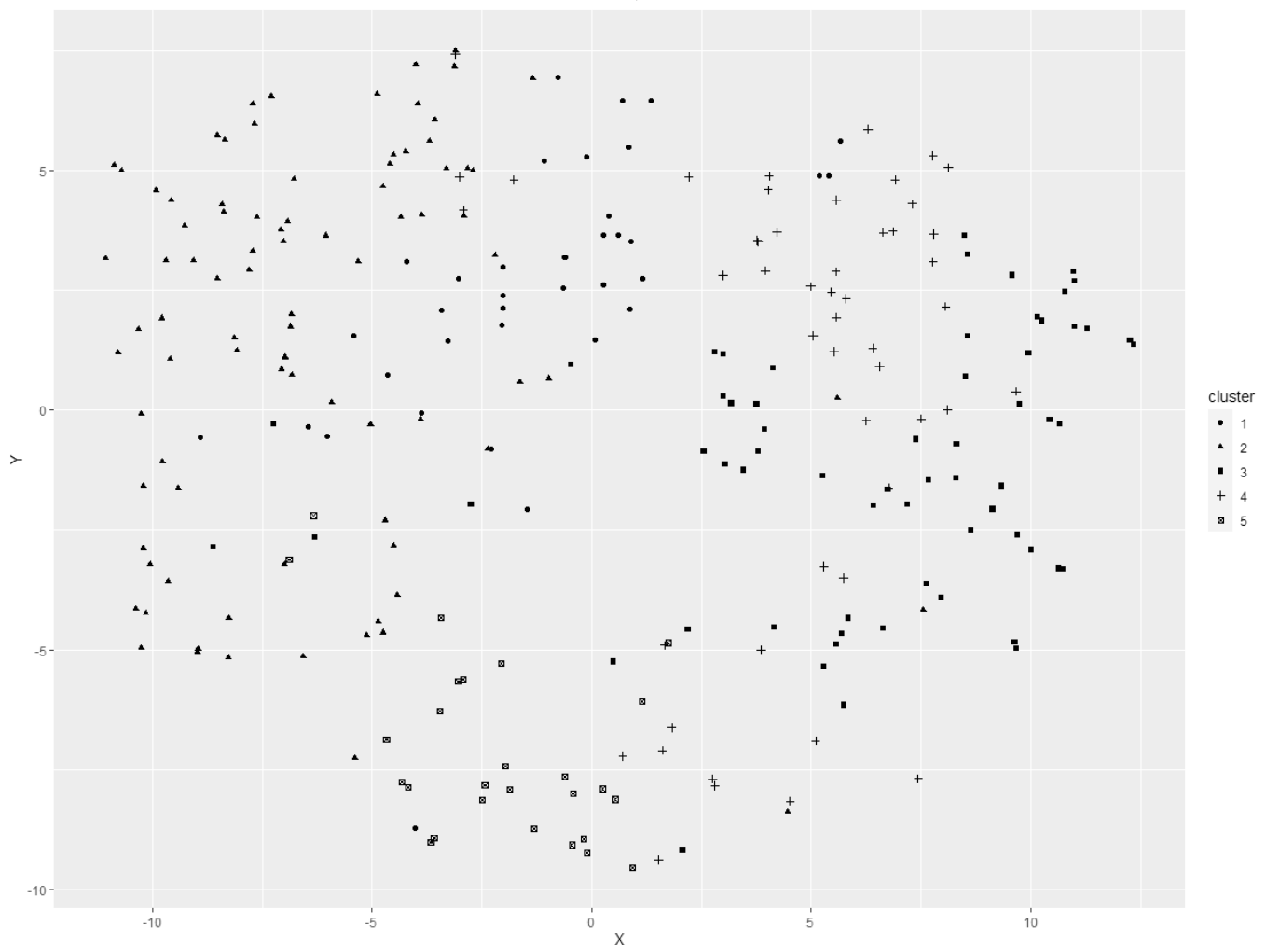

Рис. 5. Диаграмма распределения пациенток по пяти кластерам

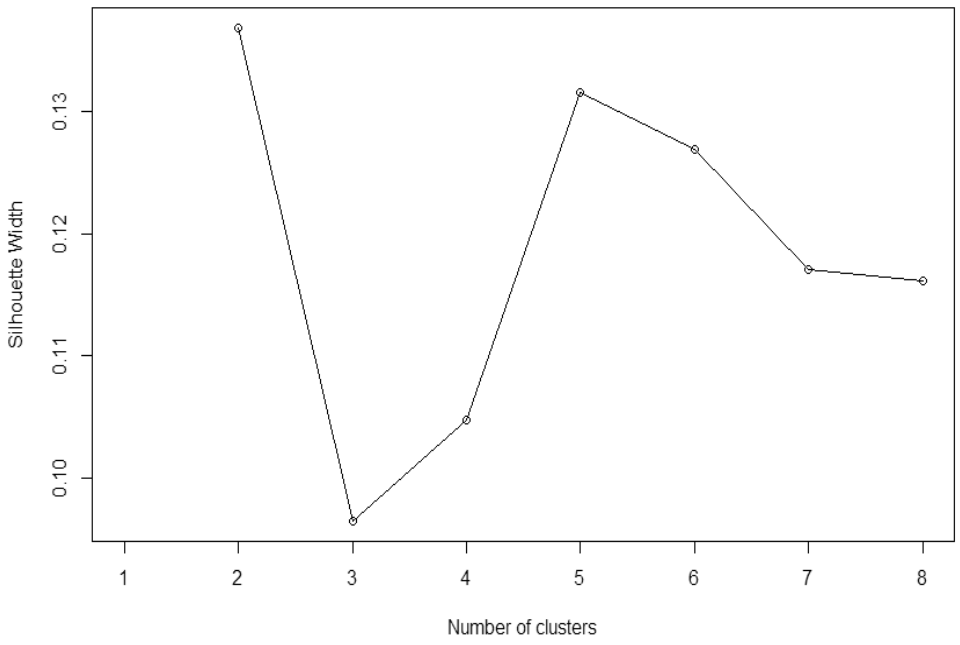

Рис. 6. Диаграмма зависимости ширины силуэта от количества кластеров

Решение задачи на языке R:

```
> BCancer <- read.table("C:/Data/breast-cancer.data",
header = FALSE, sep = ",")
> BCancer <- na.omit(BCancer)
> BCancer$V1 <- as.factor(BCancer$V1)
…
> BCancer$V10 <- as.factor(BCancer$V10)
> gower_dist <- daisy(BCancer, metric = "gower")
> gower_mat <- as.matrix(gower_dist)
> pam_fit <- pam(gower_dist, diss = TRUE, 5)
> pam_fit
> tsne_obj <- Rtsne(gower_dist, is_distance = TRUE)
> tsne_data <- tsne_obj$Y%>% data.frame()%>%
setNames(c("X", "Y"))%>%
mutate(cluster = factor(pam_fit$clustering))
> ggplot(aes(x = X, y = Y), data = tsne_data) +
geom_point(aes(shape = cluster))
```

После очистки данных количество наблюдений уменьшилось до 277. В первый кластер попало 40 наблюдений (14%), во 2-й — 90 наблюдений (32%), в 3-й — 66 наблюдений (24%), в 4-й — 53 наблюдения (19%), в 5-й — 28 наблюдений (10%).

На рис. 5 представлена визуализация кластерного анализа с помощью алгоритма *t-SNE (стохастического вложения соседей с t-распределением)*, позволяющего вкладывать многомерные данные в двухмерное пространство. На диаграмме отчетливо различимы 1-й и 2-й кластеры, 1-й и 3-й кластеры, 5-й и 4-й кластеры, а кластеры 4-й и 3-й практически сливаются.

Выбор пяти кластеров основывался на методе *ширины силуэта* (*silhouette width*), оценивающего качество кластеризации [17]:

$$s_i = \frac{b(i) - a(i)}{\max[b(i), a(i)]},$$

где $a(i)$ — среднее расстояние между объектами i-го кластера; $b(i)$ — среднее расстояние от объектов i-го кластера до самого близкого кластера.

На основе кода R, приведенного в работе [15], предварительно была построена диаграмма силуэтов (рис. 6). Разбиение на 2 и 5 кластеров дают в данном случае самые высокие значения ширины силуэта.

Заключение

Язык R является эффективным средством, в котором реализованы все актуальные методы кластерного анализа, использующиеся в медицинских исследованиях. Кластеризация данных в R сводится к использованию функций из различных пакетов. В языке R имеются средства визуализации состава кластеров и методы оценки качества кластеризации.

ЛИТЕРАТУРА

1. Касюк, С.Т. Современные информационные технологии в медицинских исследованиях: сравнение данных по качественному признаку с использованием языка R / С.Т. Касюк, Т.Н. Шамаева // Современная наука: актуальные проблемы теории и практики. Естественные и технические науки. — 2019. — № 5. — С. 60–66.
2. Касюк, С.Т. Современные информационные технологии в медицинских исследованиях: непараметрические методы сравнения данных по количественному признаку с использованием языка R / С.Т. Касюк, Т.Н. Шамаева // Cloud of Science. — 2020. — Т.7. — № 2. — С. 320–333.
3. Package «e1071», October 14, 2020, Version 1.7–4 [Электронный ресурс]. — Режим доступа: https://cran.r-project.org/web/packages/e1071/e1071.pdf (дата обращения: 09.03.2021).
4. Pamulaparty, L. Cluster analysis of medical research data using R / L. Pamulaparty, C.V Guru Rao, M. Sreenivasa Rao // Global journal of computer science and technology: C software & data engineering. — 2016. — V. 16. — № 1. — P. 17–22.
5. Breast Cancer Coimbra Data Set [Электронный ресурс]. — Режим доступа: http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra (дата обращения: 09.03.2021).
6. Трухачёва, Н.В. Математическая статистика в медико-биологических исследованиях с применением пакета Statistica / Н.В. Трухачёва. — М.: ГЭОТАР-Медиа, 2013. — 384 с.
7. Breast Cancer Wisconsin (Diagnostic) Data Set [Электронный ресурс]. — Режим доступа: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29 (дата обращения: 09.03.2021).
8. Касюк, С.Т. Анализ и прогнозирование спортивных данных в нейронных сетях: учеб.-метод. дособие / С.Т. Касюк,. — Челябинск: Уральская Академия, 2014. — 72 с.
9. Ciaburro, G. Neural networks with R / G. Ciaburro, B. Venkateswaran. — Birmingham: Packt Publishing, 2017. — 314 p.
10. Wehrens, R. Self- and super-organizing maps in R: The kohonen package / R. Wehrens, L. Buydens // Journal of Statistical Software. — October 2007. — V. 21. — № 5 [Электронный ресурс]. — Режим доступа: https://www.jstatsoft.org/issue/view/v021 (дата обращения: 09.03.2021).
11. Package «kohonen», December 26, 2019, Version 3.0.10 [Электронный ресурс]. — Режим доступа: https://cran.r-project.org/web/packages/kohonen/kohonen.pdf (дата обращения: 09.03.2021).

12. Breast Tissue Data Set [Электронный ресурс]. — Режим доступа: http://archive.ics.uci.edu/ml/datasets/Breast+Tissue (дата обращения: 09.03.2021).
13. Gower, J.C. A general coefficient of similarity and some of its properties / J.C. Gower // Biometrics. — 1971. — Dec. — V. 27. — № 4. — P. 857–871.
14. Kaufman, L. Clustering by means of medoids / L. Kaufman, P.J. Rousseeuw // Statistical data analysis based on the L1-norm and related methods. — Springer US. — 1987. — P. 405–416.
15. Clustering mixed data types in R [Электронный ресурс]. — Режим доступа: https://dpmartin42.github.io/posts/r/cluster-mixed-types (дата обращения: 09.03.2021).
16. Breast Cancer Data Set [Электронный ресурс]. — Режим доступа: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer (дата обращения: 09.03.2021).
17. Алгоритмы кластеризации, основанные на разделении [Электронный ресурс]. — Режим доступа: https://ranalytics.github.io/data-mining/101-Partitioning-Algos.html (дата обращения: 09.03.2021).

© Касюк Сергей Тимурович (sergey.kasyk@gmail.com),
Диденко Галина Александровна (pga80@mail.ru), Степанова Оксана Александровна (okalst@mail.ru).
Журнал «Современная наука: актуальные проблемы теории и практики»

«Южно-Уральский государственный медицинский университет» Министерства здравоохранения Российской Федерации