

МОРФОЛОГИЧЕСКИЙ ПОДХОД К СЕГМЕНТАЦИИ МЕДИЦИНСКИХ ДАННЫХ ПОСРЕДСТВОМ КЛАСТЕРИЗАЦИИ

A MORPHOLOGICAL APPROACH TO SEGMENTATION OF MEDICAL DATA THROUGH CLUSTERING

**K. Kraynov
E. Polyakov
E. Kharchenko**

Summary. Typical problems specific to the segmentation of tabular medical data are highlighted (for example, the complexity of interpreting the results due to the need to modify the data). It is established that classical cluster analysis methods are often used to solve the problem of dividing medical data into homogeneous groups. An alternative method for clustering multidimensional data, isolated from the morphological method of making management decisions, is proposed. A method of its modification for working with large amounts of data is proposed. To confirm the effectiveness and convenience of the method, an example of processing one open near-medical dataset is given.

Keywords: system analysis, non-tagged data, structured data, mixed type data, big data.

Крайнов Кирилл Алексеевич

Московский политехнический университет
krainov.2003@yandex.ru

Поляков Евгений Алексеевич

Московский политехнический университет
polevg2015@yandex.ru

Харченко Елена Алексеевна

старший преподаватель,
Московский политехнический университет
elenakhaa@yandex.ru

Аннотация. Выделены типичные проблемы, специфичные для сегментации табличных медицинских данных (например, сложность интерпретации результатов из-за необходимости видоизменения данных). Установлено, что для решения задачи разделения медицинских данных на однородные группы зачастую применяют классические методы кластерного анализа. Предложен альтернативный метод кластеризации многомерных данных, выделенный из морфологического метода принятия управленческих решений. Предложен способ его модификации для работы с большими массивами данных. Для подтверждения действенности и удобства метода приводится пример обработки одного открытого околomedical датасета.

Ключевые слова: системный анализ, неразмеченные данные, структурированные данные, данные смешанного типа, большие данные.

Введение

Под сегментацией понимают разделение исследуемых объектов на группы (сегменты) по критерию схожести по одному или нескольким признакам. Одна и та же совокупность объектов может подвергаться сегментации многократно — по разным группам признаков. В целом, сегментация позволяет организовать большой объем информации и выделить в нем внутреннюю структуру, представляемые профилями объектов. Обособление объектов со схожими характеристиками способствует более предметному их изучению и направленному на них воздействию. Как правило, с целью сегментации данных используют методы кластерного анализа (их относят к т.н. разведочным методам машинного обучения) [1]. Сегментация по всем признакам формально является кластеризацией.

В процессе кластеризации структурированные данные разбиваются на категориально однородные группы (кластеры) по фиксированному набору признаков, каждый из которых характеризует конкретное свойство объекта. По существу, кластеризация состоит в поиске «естественной» топологической группировки объектов. Уточнение числа кластеров и их границ в пространстве

признаков осуществляется исходя из «близости», «похожести» или «различия» унифицированных описаний объектов [2]. Отметим, что совокупность изучаемых объектов должна отвечать статистическому определению выборки и алгебраическому определению отношения.

Как базовый механизм машинного обучения кластеризация находит свое прикладное применение во многих предметных областях. Контекстом данной работы являются табличные медицинские данные (текстовые и числовые данные из анамнезов, анкет и результатов анализов пациентов) [3] и потребность в их разделении на однородные группы [4–10]. Анализ литературных источников показал, что зачастую для кластеризации (или сегментации) медицинских данных используют стандартные методы, каждый из которых имеет свои ограничения и не является универсальным [11–12]. Среди самых эксплуатируемых — метод k-means (число кластеров задается) и самоорганизующаяся карта Кохонена (число кластеров вычисляется автоматически).

Методы кластеризации, являясь только вспомогательными, интегрируются в системы информационной поддержки принятия врачебных решений, поэтому к ним предъявляются следующие обязательные

требования: возможность работы с данными смешанного типа (описываемых в общем случае взаимоувязанными качественными и количественными признаками), устойчивость решения (стабильность результатов при небольших изменениях входных данных) и простота интерпретации результатов (для возможности обоснования принимаемого решения) [5, 6].

В подавляющем большинстве методов кластеризации форма и границы кластеров определяются метрикой расстояния. Для сравнения данных, содержащих как числовые, так и категориальные признаки, наиболее распространенной (но редко используемой) является метрика Гауэра [14–16]. Ее значение вычисляется усреднением вычисленных независимо расстояний по каждому признаку.

Основными недостатками метрики для случая смешанных данных являются искажение значений количественных признаков и доминирование качественных. Также усложняет ее применение невозможность адекватной интерпретации результата (для случая смешанных данных) и высокая вычислительная сложность алгоритма его вычисления (для случая больших данных). В частных случаях (однотипных данных) наиболее близкими к метрике Гауэра являются: для количественных признаков — расстояния Евклида и Манхэттена, для качественных — Жаккара и Хэмминга.

Кластерный анализ (явно и неявно) лежит на пересечении многих разделов математики, например: статистики, эконометрики, вычислительной математики, теории принятия решений, системного анализа и машинного обучения. В настоящей работе предлагается альтернативный метод кластеризации данных, выделенный из морфологического метода экспертных оценок [17–19], который в свою очередь является развитием морфологического анализа Фрица Цвикки [20]. В нем разрешены перечисленные выше принципиальные для медицинских исследований проблемы, также он обладает рядом других преимуществ перед классическими аналогами.

1. Порядок сегментации данных

Первично морфологический анализ является формальным методом генерирования альтернатив через перечисление всех возможных сочетаний значений заданных параметров альтернативы [2]. В основе морфологического метода принятия управленческих решений [17–19] лежит аппроксимация области допустимых значений качественно однородных объектов многомерной плотностью нормального распределения. Кратко опишем порядок применения выделенного из него метода кластеризации (или сегментации в прикладном смысле).

Сначала (на первом этапе) значения каждого количественного признака исследуются на однородность (с помощью построения графика плотности вероятности или гистограммы). Каждая область сгущения значений является проявлением некоторой сущности (категории, типа) объектов. По локальным минимумам плотности распределения область допустимых значений количественного признака разбивается на непересекающиеся подмножества. Если признак качественный, его область допустимых значений является объединением одноэлементных множеств, состоящих из возможных дискретных значений признака.

После этого (на втором этапе) формируется множество многомерных областей допустимых значений совокупности признаков исследуемых объектов (путем вычисления декартова произведения полученных на первом этапе множеств). Каждый элемент такого множества описывается многомерным эллипсоидом, параметры которого определяются по «правилу трех сигм».

Затем (на третьем этапе) для возможности измерения расстояния между произвольными объектами или кластерами формируется вспомогательное линейное пространство двоичных векторов. Каждому элементу построенного на предыдущем этапе множества соответствует свой двоичный вектор. Структура каждого вектора такова, что каждой подобласти области допустимых значений каждого признака соответствует свой бит.

Для каждой многомерной области допустимых значений требуется вычислить относительную частоту попадания объектов выборки (вероятность в классическом понимании), после чего области следует упорядочить по убыванию частот и пронумеровать. Номер области понимается как случайная величина, поэтому на основании неравенства Чебышева возможно разделение ее значений на практически возможные (не менее 89%) и выбросы. Так определяются число кластеров и их границы (поскольку каждое значение случайной величины ассоциировано с конкретной многомерной областью допустимых значений объектов).

За метрику расстояния для векторов введенного на третьем этапе линейного пространства принимается половина расстояния Хемминга, равного количеству позиций, в которых соответствующие биты различны. Значение метрики трактуется как количество категориально различных одноименных признаков двух сравниваемых через двоичные векторы объектов или кластеров.

Для возможности работы с большими массивами данных потребовалась модификация данного метода, сводящаяся к иерархическому укрупнению кластеров (с помощью описанной функции расстояния). На нижнем уровне в одну группу объединяются те области, между

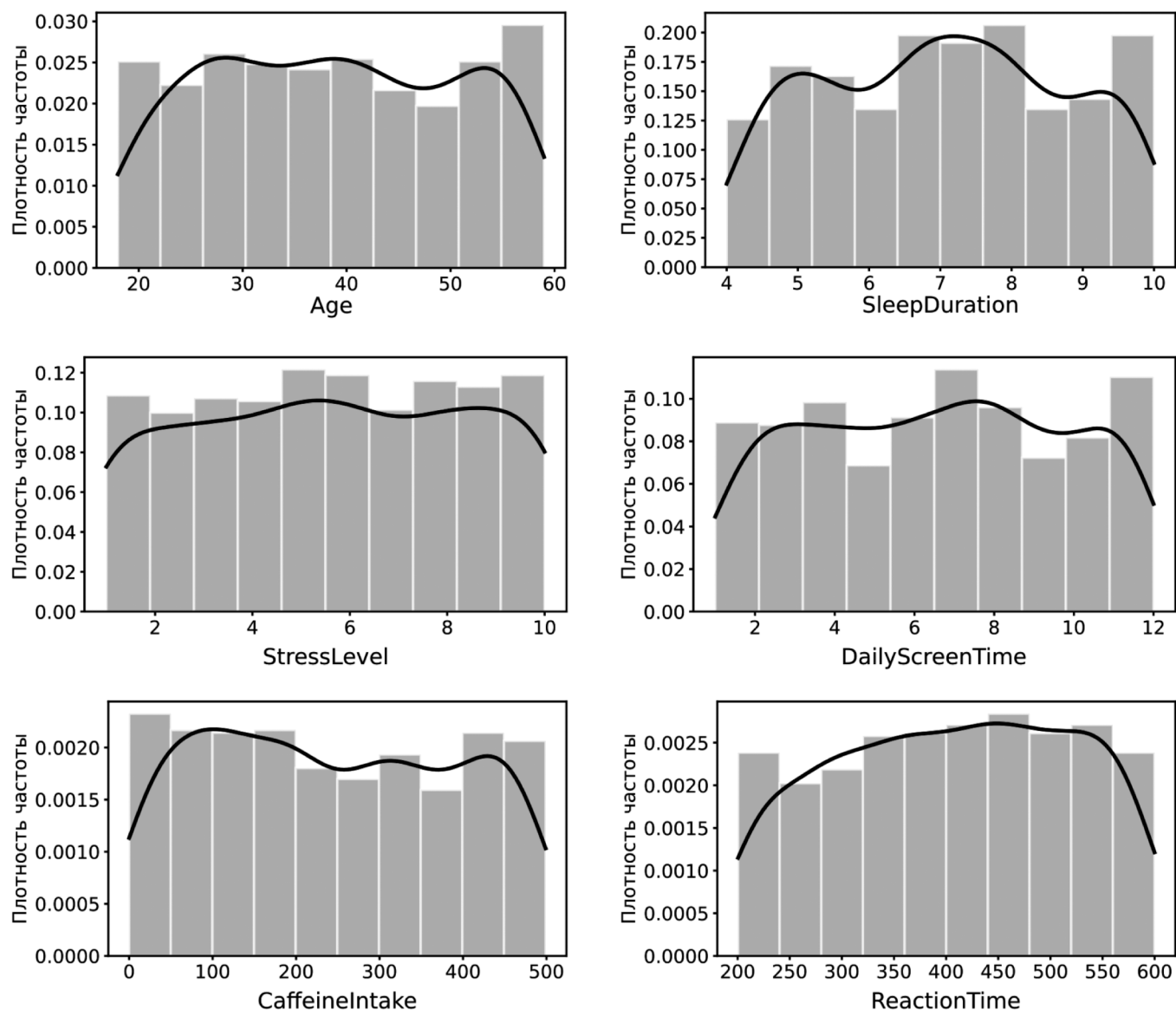


Рис. 1. Распределения значений количественных признаков

Источник: составлено авторами на основании реализации алгоритма

которыми расстояние минимальное и одинаковое для любой пары областей внутри этой группы (причем укрупненные кластеры не должны пересекаться). После этого требуется перестроить двоичные векторы (ввести новое линейное пространство), поскольку областей сгущения у признака может стать меньше или могут изменяться их границы. Укрупнения можно итерационно продолжить до образования одной области, кластера.

Традиционным же графическим способом представления результатов агломеративной иерархической кластеризации является построение дендрограммы. Она демонстрирует степень близости отдельных кластеров, а также последовательность их объединения.

Принципиальным преимуществом иерархической кластеризации является возможность проведения

структурного анализа многомерных данных с разной степенью детализации.

2. Пример сегментации данных

Для иллюстрации преимуществ представленного в предыдущем разделе метода, приведем результаты кластеризации околomedicalных (но интуитивно понятных) данных, описывающих когнитивные способности человека¹. Датасет содержит порядка 80000 записей со значениями девяти признаков (без учета привнесенных агрегирующих признаков). Как правило, результаты кластеризации выражаются в разметке (маркировке) записей датасета (путем добавления целевых признаков с техническими значениями).

¹ <https://www.kaggle.com/datasets/samxsam/human-cognitive-performance-analysis>

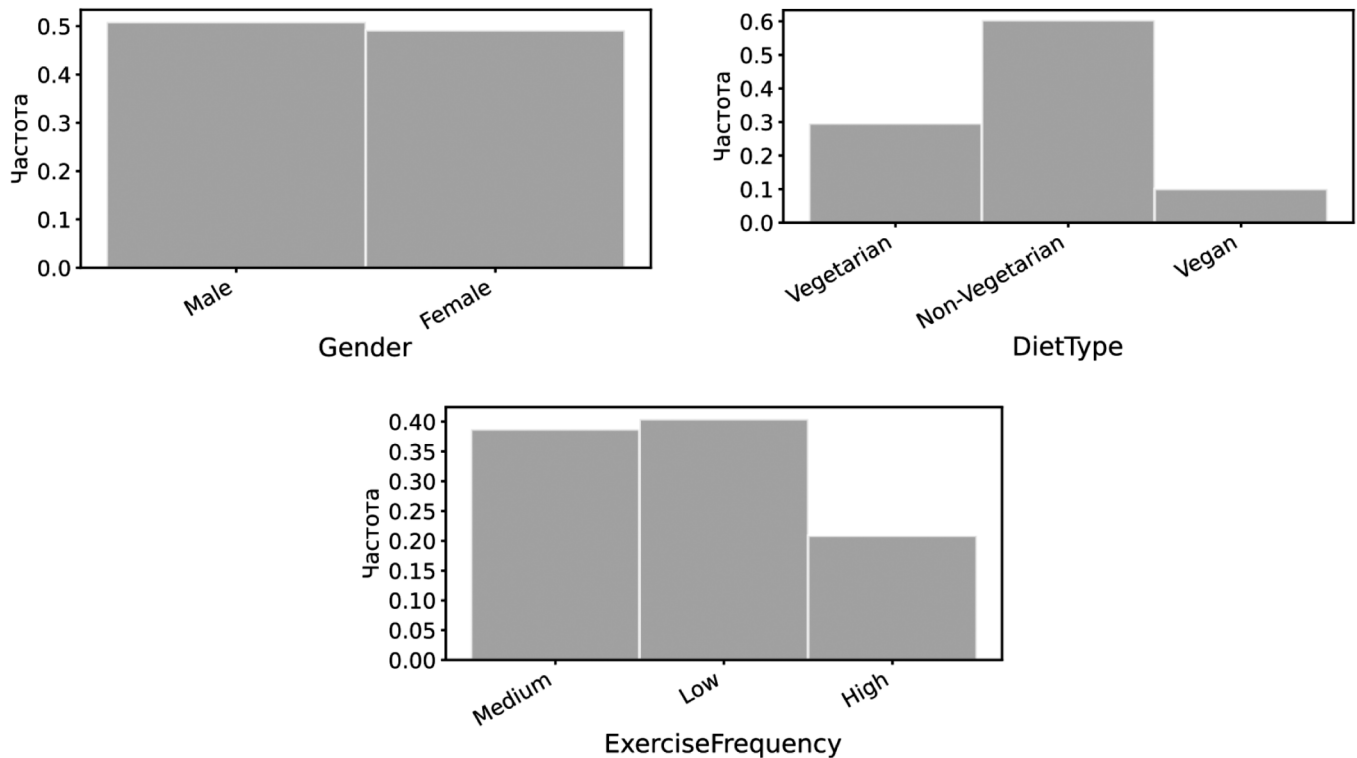


Рис. 2. Распределения значений качественных признаков

Источник: составлено авторами на основании реализации алгоритма

Шесть из признаков количественные (распределения их значений приведены на рис. 1): Age (возраст), SleepDuration (среднесуточное время сна), StressLevel (самооценка уровня стресса), DailyScreenTime (среднесуточное время, проводимое за экраном монитора), CaffeineIntake (среднесуточное потребление кофеина) и ReactionTime (среднее время когнитивной реакции при прохождении тестов). Три являются качественными (рис. 2): Gender (пол), DietType (пищевые предпочтения) и ExerciseFrequency (уровень физической активности).

Заметим, что метод не требует специальной предобработки данных (не видоизменяет данные) и может работать с пропусками значений (как с вариантами значений).

В таблице 1 приведены вычисленные по локальным минимумам гистограмм (или плотностей распределения) области допустимых значений количественных признаков. Так, например, отдельное значение признака SleepDuration (среднесуточное время сна, в часах) может принадлежать только одной из четырех качественно однородных подобластей (интервалов): [1.96; 6), [6; 7), [7; 8) или [8; 13.12). Каждая из них представляется средним значением и границами доверительного интервала, определяемых по «правилу трех сигм» (распределение полагается нормальным).

В таблице 2 описаны области допустимых значений качественных признаков. Для примера, признак

ExerciseFrequency (уровень физической активности) может принимать одно из трех значений — «Medium» (средний), «Low» (низкий), «High» (высокий).

Применение базового варианта алгоритма привело к образованию 751 кластера, 84 из которых — это области выбросов. Мощность множества многомерных областей допустимых значений совокупности признаков равна произведению мощностей доменов признаков (т.е. 46080). Подавляющее большинство образованных областей не соответствует реальной действительности (их частота равна нулю). Размерность вспомогательного линейного пространства двоичных векторов равна 32.

Для решения проблемы переобучения применена итерационная процедура укрупнения кластеров — построено агломеративное дерево, состоящее из 14 уровней. На втором уровне дерева (после первой итерации укрупнения) получены 746 кластеров, 84 из которых — выбросы. На 11-ом уровне (после 10-ой итерации укрупнения) размещаются 18 кластеров (выбросами являются 7, что составляет 10.8 % данных). На верхнем уровне (в корне дерева) содержится всего один кластер (выбросов нет).

В таблице 3 приведены двоичные векторы (ассоциированные с кластерами), соответствующие кластерам 10-ой итерации укрупнения (с 12-го по 18-ый — выбросы). Разъясним структуру двоичных векторов этого этапа

Таблица 1.

Области допустимых значений количественных признаков

№	Признак	Область
1	Age (возраст)	{Age_1: [2.98; 24), Age_2: [24; 36), Age_3: [36; 48), Age_4: [48; 75.46]}
2	SleepDuration (среднесуточное время сна)	{SleepDuration_1: [1.96; 6), SleepDuration_2: [6; 7), SleepDuration_3: [7; 8), SleepDuration_4: [8; 13.12]}
3	StressLevel (уровень стресса)	{StressLevel_1: [0; 2), StressLevel_2: [2; 4), StressLevel_3: [4; 6), StressLevel_4: [6; 8), StressLevel_5: [8; 15.2]}
4	DailyScreenTime (среднесуточное время за экраном монитора)	{DailyScreenTime_1: [0; 2), DailyScreenTime_2: [2; 4), DailyScreenTime_3: [4; 9), DailyScreenTime_4: [9; 17.15]}
5	CaffeineIntake (среднесуточное потребление кофеина, мг)	{CaffeineIntake_1: [0.0; 124), CaffeineIntake_2: [124; 274), CaffeineIntake_3: [274; 374), CaffeineIntake_4: [374; 686.17]}
6	ReactionTime (среднее время когнитивной реакции, мс)	{ReactionTime_1: [67.87; 260), ReactionTime_2: [260; 499), ReactionTime_3: [499; 745.56]}

Источник: составлено авторами на основании реализации алгоритма.

Таблица 2.

Области допустимых значений качественных признаков

№	Признак	Область
1	Gender (пол)	{Gender_1: {'Male'}, Gender_2: {'Female'}}
2	DietType (пищевые предпочтения)	{DietType_1: {'Vegetarian'}, DietType_2: {'Non-Vegetarian'}, DietType_3: {'Vegan'}}
3	ExerciseFrequency (физическая активность)	{ExerciseFrequency_1: {'Medium'}, ExerciseFrequency_2: {'Low'}, ExerciseFrequency_3: {'High'}}

Источник: составлено авторами на основании реализации алгоритма.

кластеризации на примере второго из них. Каждый вектор имеет девять смысловых частей, каждая из которых ассоциирована с областью допустимых значений конкретного признака и содержит определенное число бит (последовательно слева направо): Age, SleepDuration,

Таблица 3.

Двоичные векторы для 11-го уровня дерева кластеров

№ области	Вектор	Частота
1	(0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1)	0.211
2	(0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1) Домен для Age: {Age_1: {[2.98; 24)}, Age_2: {[24; 36)}, Age_3_4: {[36, 75.46)}} Домен для ReactionTime: {ReactionT_2: {[260; 499)}, ReactionT_1_3: {[67.87, 260), [499, 745.56)}}}	0.133
3	(0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1)	0.128
4	(0, 1, 0, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1)	0.072
5	(0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1)	0.072
6	(0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1)	0.069
7	(0, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1)	0.060
8	(1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1)	0.048
9	(0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1)	0.040
10	(0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1)	0.033
11	(0, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1)	0.026
12	(0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1)	0.023
13	(1, 0, 0, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1)	0.023
14	(1, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1)	0.020
15	(0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1)	0.013
16	(1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1)	0.012
17	(1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1)	0.012
18	(1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1)	0.005

Источник: составлено авторами на основании реализации алгоритма.

StressLevel, DailyScreenTime, CaffeineIntake, ReactionTime, Gender, DietType, ExerciseFrequency.

Таким образом, первые три бита (0, 0 и 1) относятся к описанию составляющих области допустимых значений признака Age. Единица уточняет возможный возраст для объектов данного кластера (он не может быть меньше 36). При объединении областей объединяются (и сокращаются) и их наименования: область Age_3_4 является результатом объединения атомарных областей Age_3 и Age_4. Шестая группа битов (0, 1) относится к области значений признака ReactionTime — скорость когнитивной реакции для объектов этого кластера принадлежит только второму интервалу.

Для примера сравним кластеры, которым соответствуют второй и третий двоичные векторы. Расстояние между ними вычисляется как половина расстояния Хэмминга и равно 2 — кластеры качественно различны по двум признакам (по возрасту и скорости когнитивной реакции). Соответствующие области допустимых значений признаков для каждого кластера приведены в таблице 4.

Таблица 4.

Области допустимых значений для двух кластеров

№ кластера	Область допустимых значений
2	{Age_3_4: {[36, 75.46]}, SleepD_1_2_3_4: {[1.96, 13.12]}, StressL_1_3_5: {[0, 2], [4, 6], [8, 15.2]}, DailyST_1_2_3_4: {[0, 17.15]}, Caffeinel_1_2_4_3: {[0.0, 686.17]}, ReactionT_1_3: {[67.87, 260], [499, 745.56]}, Gender_1_2: {'Female', 'Male'}, DietT_1_2_3: {'Non-Vegetarian', 'Vegan', 'Vegetarian'}, ExerciseF_1_2_3: {'High', 'Low', 'Medium'}}
3	{Age_2: {[24, 36]}, SleepD_1_2_3_4: {[1.96, 13.12]}, StressL_1_3_5: {[0, 2], [4, 6], [8, 15.2]}, DailyST_1_2_3_4: {[0, 17.15]}, Caffeinel_1_2_4_3: {[0.0, 686.17]}, ReactionT_2: {[260, 499]}, Gender_1_2: {'Female', 'Male'}, DietT_1_2_3: {'Non-Vegetarian', 'Vegan', 'Vegetarian'}, ExerciseF_1_2_3: {'High', 'Low', 'Medium'}}

Источник: составлено авторами на основании реализации алгоритма.

Аналогично осуществляется сравнение и отдельных элементов выделенных кластеров.

Приведен сведенный к кластеризации частный случай сегментации исследуемых объектов — по всем признакам, без контекста исследования. В общем и прикладном смысле сегментация позволяет делать срезы по интересующим (некоторым) признакам внутри сформированной однородной группы объектов (классе, кластере). Отдельная процедура сегментации может осуществляться посредством методов кластеризации.

Если кластеризации (или сегментации) подлежат все объекты и выборку, представленную датасетом, возможно уменьшить с сохранением свойств репрезентативности, то дерево кластеров может быть построено по редуцированному датасету, а принадлежность к выявленным кластерам объектов, не участвовавших в построении дерева, можно определить на основании введенной метрики расстояния (с учетом крайних доверительных интервалов количественных признаков).

Описанный подход к кластеризации многомерных данных показывает пригодный для практического при-

менения и достоверный результат, подтверждаемый обработкой разнообразных размеченных датасетов: получаемые кластеры оказываются преимущественно однородными по исходным целевым признакам.

Заключение

Для проверки точности и корректности предлагаемого метода кластеризации (с целью сегментации) многомерных данных был разработан специальный встраиваемый программный модуль. В качестве языка программирования был использован Python (из-за наличия у него большого количества современных библиотек для обработки, визуализации и анализа данных). Чтобы модуль было потенциально возможно интегрировать в любую профессиональную систему поддержки принятия решений, для него с помощью фреймворка FastAPI был дополнительно разработан прикладной программный интерфейс (с облегченным механизмом эксплуатации).

Математический аппарат, лежащий в основе работы модуля, носит универсальный характер и может найти прикладное применение в различных предметных областях. Особую практическую значимость программное решение может иметь при проведении медицинских исследований, поскольку удовлетворяет заявленным медицинским сообществом требованиям. А именно: 1) предусмотрена возможность одновременной работы с качественными и количественными признаками; 2) число кластеров вычисляется автоматически (без необходимости подбора параметров кластеризации); 3) результаты кластеризации легко поддаются интерпретации (что крайне важно для обоснования принимаемых на их основе решений). Последнее обеспечивается тем, что: 1) данные сохраняют исходный вид и единицы измерения; 2) значение метрики расстояния информативно; 3) используются базовые математические понятия.

Легко преодолимым недостатком программного решения в настоящее время является грубая аппроксимация границ областей допустимых значений кластеров (посредством разлиновывающей признаковое пространство прямоугольной сетки), из-за чего объекты, лежащие на периферии истинных доверительных интервалов, оказываются в числе выбросов. Существенно повысить точность предложенной модели возможно, если проверять объекты, маркируемые как выбросы, на принадлежность многомерным эллиптическим доверительным областям кластеров. При этом кластеры объектов, лежащих на пересечениях таких областей, потребуется дополнительно уточнять с помощью функции правдоподобия.

ЛИТЕРАТУРА

1. Руководство аналитика Deductor 5.3. URL: <https://basegroup.ru/deductor/manual/guide-analyst-530> (дата обращения: 21.01.2025).
2. Перегудов, Ф.И. Введение в системный анализ / Ф.И. Перегудов, Ф.П. Тарасенко. — М.: Высшая школа, 1989. — 367 с.
3. Казанцев, В.С. К вопросу оценки полноты и достоверности таблиц медицинских данных / В.С. Казанцев, С.А. Никифоров, Е.В. Ползик // Врачи и информационные технологии. — 2007. — № 2. — С. 45–49.
4. Касюк, С.Т. Кластерный анализ данных медицинских исследований с использованием языка R / С.Т. Касюк, Г.А. Диденко, О.А. Степанова // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и Технические Науки. — 2021. — № 04/2. — С. 23–32. doi:10.37882/2223-2966.2021.04-2.11.
5. Альсова, О.К. Программная система кластерного анализа данных смешанного типа / О.К. Альсова, К.С. Ускова // Автоматика и программная инженерия. — 2013. — № 1(3). — С. 75–81.
6. Альсова, О.К. Алгоритмы кластеризации разнотипных данных на примере решения медицинской задачи / О.К. Альсова // Труды СПИИРАН. — 2014. — № 37. — С. 156–169. doi:10.15622/sp.37.10.
7. Берестнева, О.Г. Методы исследования структуры медицинских данных / О.Г. Берестнева, И.А. Осадчая, Е.В. Немеров // Вестник науки Сибири. — 2012. — № 1(2). — С. 333–338.
8. Савченко, Т.Н. Применение методов кластерного анализа для обработки данных психологических исследований / Т.Н. Савченко // Экспериментальная психология. — 2010. — Т. 3 — № 2. — С. 67–86.
9. Судаков, С.А. Кластерный анализ в психиатрии и клинической психологии: Руководство / С.А. Судаков — М.: ООО «Медицинское информационное агентство», 2010. — 160 с.
10. Мезенцев, Ю.А. Задачи и алгоритмы оптимальной кластеризации многомерных объектов по множеству разнородных показателей и их приложения в медицине / Ю.А. Мезенцев, О.М. Разумникова, И.В. Эстрайх, И.В. Тарасова, О.А. Трубникова // Компьютерные исследования и моделирование. — 2024. — Т. 4 — № 3. — С. 673–693. doi:10.20537/2076-7633-2024-16-3-673-693.
11. Дюк, В. Информационные технологии в медико-биологических исследованиях / В. Дюк, В. Эмануэль — СПб.: Питер, 2003. — 528 с.
12. Айвазян, С.А. Прикладная статистика. Классификация и снижение размерности / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин — М.: Финансы и статистика, 1989. — 471 с.
13. Бериков, В.Б. Современные тенденции в кластерном анализе / В.Б. Бериков, Г.С. Лбов // Всероссийский конкурсный отбор обзорно-аналитических статей по приоритетному направлению «Информационно-телекоммуникационные системы». — 2008. — 26 с.
14. Gower, J.C. A general coefficient of similarity and some of its properties. *Biometrics*, 1971, vol. 27, no. 4, pp. 857–871.
15. Годяев, А.А. Программа построения самоорганизующихся карт Кохонена при категориальных и смешанных данных / А.А. Годяев, А.В. Гиголаев, Н.И. Цуканова // Вестник РГТУ. Серия: Прикладная математика. — 2017. — № 61. — С. 78–87. doi:10.21667/1995-4565-2017-61-3-78-87.
16. Уткин, В.А. Классификация и диагностика в современной медицине с применением дискриминантного и кластерного анализа / В.А. Уткин // Медицинский вестник Северного Кавказа. — 2009. — № 4. — С. 70–73.
17. Харченко, Е.А. Морфологический подход к принятию обоснованных решений по экспертным суждениям / Е.А. Харченко // Вестник ТвГУ. Серия: Прикладная математика. — 2019. — № 2. — С. 42–56. doi:10.26456/vtpmk531.
18. Харченко, Е.А. Алгоритм морфологического метода экспертных оценок для решения задачи прогнозирования / Е.А. Харченко // Компьютерные инструменты в образовании. — 2023. — № 2. — С. 5–20. doi:10.32603/2071-2340-2023-2-5-20.
19. Харченко, Е.А. Механическая модель морфологического метода принятия решений в условиях неопределенности / Е.А. Харченко // Компьютерные инструменты в образовании. — 2024. — № 3. — С. 14–32. doi:10.32603/2071-2340-2024-3-14-32.
20. Одрин, В.М. Метод морфологического анализа технических систем / В.М. Одрин — М.: ВНИИПИ, 1989. — 312 с.

© Крайнов Кирилл Алексеевич (krainov.2003@yandex.ru); Поляков Евгений Алексеевич (polevg2015@yandex.ru);

Харченко Елена Алексеевна (elenakhaa@yandex.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»