Владимир Л. Евсеев[1], Антон С. Бураков[2], Виталий Г. Иваненко[3]
[1,2]Финансовый университет при Правительстве Российской Федерации (Финансовый университет),
Ленинградский пр–кт, 49, Москва, 125993, Россия
[3]Национальный исследовательский ядерный университет «МИФИ»,
Каширское ш., 31, Москва, 115409, Россия
[1]e-mail: VLEvseev@fa.ru, https://orcid.org/0000-0003-3283-3106
[2]e-mail: anton27061999@yandex.ru, https://orcid.org/0000-0003-1380-5273
[3]e-mail: VGIvanenko@mephi.ru, https://orcid.org/0000-0003-0823-5501

ИСПОЛЬЗОВАНИЕ МЕТОДОВ КЛАСТЕРНОГО АНАЛИЗА ДЛЯ ОПТИМИЗАЦИИ КАЧЕСТВЕННОЙ ОЦЕНКИ РИСКОВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

DOI: http://dx.doi.org/10.26583/bit.2021.2.07

Аннотация. Статья посвящена точности оценки рисков информационной безопасности. В статье обосновывается актуальность оценки рисков, исходя из последствий их реализации для бизнеса и вероятности их возникновения. Анализируется метод качественной оценки рисков информационной безопасности (метод экспертной оценки) на конкретном примере. Обосновывается применение методов кластерного анализа. На примерах показано использование методов кластерного анализа: метод ближайшего соседа, метод удаленного соседа, метод *k*-средних. Приводятся принципиальные недостатки первых двух методов: появление больших кластеров не имеющих сходств, отсутствие возможности у экспертов заранее задать желаемое количество кластеров. Обосновывается применение метода *k*-средних – наличие возможности у экспертов заранее задать желаемое количество кластеров с помощью задания начальных центров. Приводится сравнение результатов, полученных при обычной качественной оценке, с результатами полученными методами кластерного анализа. Обосновывается целесообразность использования методов кластерного анализа для повышения точности оценки рисков информационной безопасности.

Ключевые слова: оценка риска, методы кластерного анализа, метод ближайшего соседа, метод удаленного соседа, метод k-средних, степень реализации угрозы, степень влияния угрозы на актив, евклидово расстояние, определяющее расстояние, среднее внутрикластерное расстояние.
Для цитирования: ЕВСЕЕВ, Владимир Л.; БУРАКОВ, Антон С.; ИВАНЕНКО, Виталий Г. ИСПОЛЬЗОВАНИЕ МЕТОДОВ КЛАСТЕРНОГО АНАЛИЗА ДЛЯ ОПТИМИЗАЦИИ КАЧЕСТВЕННОЙ ОЦЕНКИ РИСКОВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ. Безопасность информационных технологий, [S.l.], v. 28, n. 2, p. 70–82, 2021. ISSN 2074-7136. Доступно на: <https://bit.mephi.ru/index.php/bit/article/view/1345>. Дата доступа: 29 apr. 2021. DOI: http://dx.doi.org/10.26583/bit.2021.2.07.

Vladimir L. Evseev[1], Anton S. Burakov[2], Vitaliy G. Ivanenko[3]
[1,2]Financial University under the Government of the Russian Federation (Financial University),
Leningradsky prospekt, 49, Moscow, 125993, Russia
[3]National Research Nuclear University MEPhI (Moscow Engineering Physics Institute),
Kashirskoe shosse, 31, Moscow, 115409, Russia
[1]e-mail: VLEvseev@fa.ru, https://orcid.org/0000–0003-3283-3106
[2]e-mail: anton27061999@yandex.ru, https://orcid.org/0000-0003-1380-5273
[3]e-mail: VGIvanenko@mephi.ru, https://orcid.org/0000-0003-0823-5501

Using cluster analysis techniques to optimize the qualitative assessment of information security risk

DOI: http://dx.doi.org/10.26583/bit.2021.2.07

Abstract. The study is devoted to the accuracy of information security risk assessment. The paper substantiates the relevance of risk assessment, based on the consequences of their implementation for business and the probability of their occurrence. The method of qualitative assessment of information security risks (the method of expert assessment) is analysed on a specific example. The application of cluster analysis methods is justified. In detail, the examples show the use of cluster analysis methods: the nearest neighbor method; the remote neighbor method; the k-means method. The principal disadvantages of the first two methods are: the appearance of large clusters that do not have similarities; the lack of the ability of experts to set the desired number of clusters in advance. The application of the k-means method is justified - the ability of experts to set the desired number of clusters in advance by setting the initial centers. The results obtained with the usual qualitative assessment are compared with the results obtained by the methods of cluster analysis. The expediency of using cluster analysis methods to improve the accuracy of information security risk assessment is justified.

Keywords: risk assessment, cluster analysis methods, nearest neighbor method, remote neighbor method, k-means method, degree of threat realization, degree of threat impact on the asset, Euclidean distance, determining distance, average intra-cluster distance.

For citation: EVSEEV, Vladimir L.; BURAKOV, Anton S.; IVANENKO, Vitaliy G. Using cluster analysis techniques to optimize the qualitative assessment of information security risk. IT Security (Russia), [S.l.], v. 28, n. 2, p. 70–82, 2021. ISSN 2074-7136. Available at: <https://bit.mephi.ru/index.php/bit/article/view/1345>. Date accessed: 29 apr. 2021. DOI: http://dx.doi.org/10.26583/bit.2021.2.07.

Введение

В настоящее время количество и сложность информационных систем стремительно растет. Вместе с этим возрастает и число угроз для этих информационных систем. Это ставит вопросы информационной безопасности (ИБ) в IT-технологиях на первое место.

Для того, чтобы реализация угроз ИБ для компании не стала фатальной, следует придерживаться систематического подхода к менеджменту риска ИБ, который позволяет предотвращать или, в случае реализации угрозы, минимизировать последствия [1].

В национальном стандарте РФ ГОСТ Р ИСО/МЭК 27005-2010[1] дано определение риска ИБ, как возможность того, что данная угроза сможет воспользоваться уязвимостью актива или группы активов и тем самым нанесет ущерб организации.

Процесс управления рисками ИБ состоит из нескольких этапов [2, 3]. На наш взгляд самым важным этапом является оценка риска

Существует два основных способа оценки риска ИБ – количественный [4] и качественный [5, 6]. В первом – риск определяется по формуле [7]:

$$R = P(t) * S\,, \qquad (1)$$

где R – значение риска, $P(t)$ – вероятность реализации угрозы ИБ, S – степень влияния угрозы на активы (стоимость активов).

Если в выражении (1) стоимость актива S определить достаточно просто, то точно оценить вероятность реализации угрозы $P(t)$ достаточно сложно.

Поэтому, помимо количественной оценки риска (1), для анализа и оценки рисков ИБ применяется метод качественной оценки рисков. Его суть состоит в привлечении экспертов, которые ранжируют риски ИБ по степени их реализации и степени влияния на размер наносимого (нанесенного) ущерба (либо по 10-бальной шкале, либо по 5-бальной, либо другой шкале, выбранной в этом методе). Идентификационные признаки, по которым будет проводиться кластеризация: степень реализации угрозы, значение которой может быть от 0 до 10 (где степень равная 0 имеет нулевую вероятность реализации

[1]Национальный стандарт РФ ГОСТ Р ИСО/МЭК 27005-2010 "Информационная технология. Методы и средства обеспечения безопасности. Менеджмент риска информационной безопасности" (утв. приказом Федерального агентства по техническому регулированию и метрологии от 30 ноября 2010 г. N 632-ст)

угрозы, а значение степени равное 10 – вероятность реализации угрозы равна 1,0), степень влияния угрозы на актив, значение которой может быть от 0 до 10 (где степень равная 10 означает полное уничтожение актива). После выставления оценок каждым из экспертов, для каждого риска находятся средние арифметические степени его реализации и степени влияния угрозы на актив (размер наносимого ущерба) Но данный метод качественной оценки рисков ИБ не лишен недостатков. Рассмотрим это на примере. Допустим, что эксперты для оценки рисков составили матрицу, в которой выделили пять областей рисков: очень низкие, низкие, средние, высокие и очень высокие (рис. 1). Экспертами была проведена оценка пяти рисков $x_1,\ldots,x_5$ информационной безопасности по 10-бальной шкале. Для каждого риска были определены средние арифметические степени реализации угрозы и степени влияния угрозы на актив (размер наносимого ущерба): $x_1 = (5{,}6;\ 5{,}6);\ x_2 = (6{,}2;\ 6{,}4);\ x_3 = (7;\ 6{,}2);\ x_4 = (8;\ 0{,}8);\ x_5 = (0{,}6;\ 7)$, рис. 1.

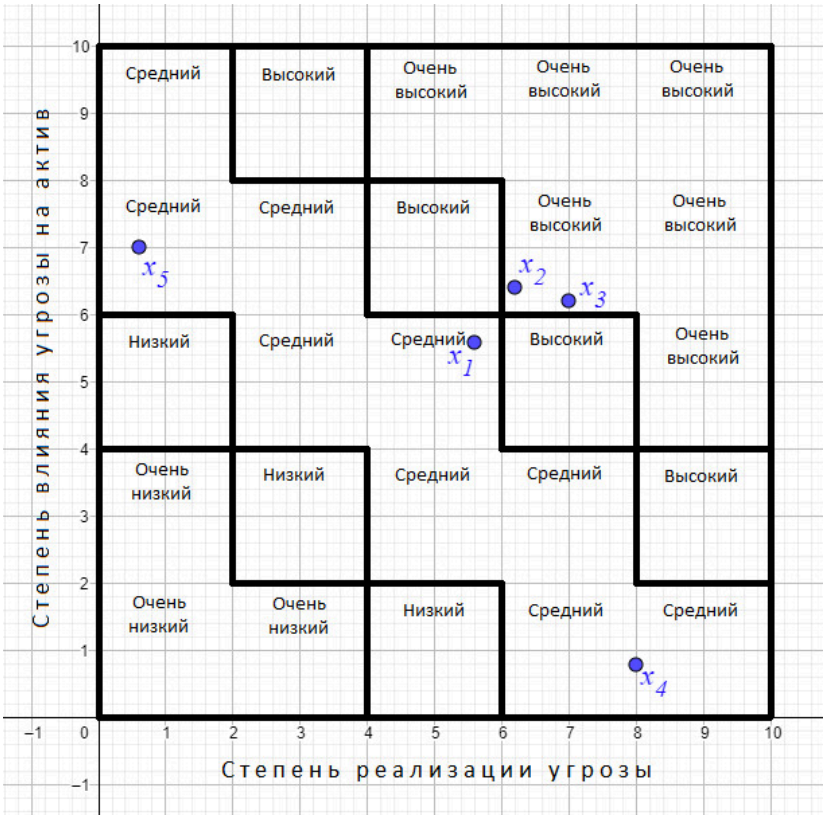

Рис. 1. Пример качественной оценки рисков
Fig. 1. Example of a qualitative risk assessment

Формально, по результатам классификации, риски x_1, x_4, x_5 должны обрабатываться специалистами по ИБ, как средние, а x_2 и x_3 – как очень высокие. Но на рис. 1 видно: риск x_1 на плоскости находится намного ближе к x_2 и x_3, поэтому его целесообразно обрабатывать, как очень высокий, или как минимум, высокий. После оценки рисков ИБ следует этап их обработки, который включает закупку и установку специалистами по ИБ средств защиты информации (СЗИ). Эффективность обработки риска зависит от результатов оценки риска.

Из-за неточности оценки рисков экспертами, они в дальнейшем могут быть обработаны некорректно, что может привести к одному из двух вариантов развития событий:

- неэффективная обработка риска, когда он относится к более низкой категории, чем реально, тогда при реализации угрозы ИБ компания понесет большие убытки;
- либо, наоборот, риск будет отнесен к более высокой категории, что приведет к значительному завышению средств на закупку и установку СЗИ, т.е. на обработку риска потратится средств больше, чем реально требуется.

Кроме того, для уменьшения общих затрат на закупку и установку СЗИ, целесообразно незначительные риски объединять в одну группу, например, области очень низкие и низкие. Для упорядочения рисков ИБ в сравнительно однородные группы целесообразно использовать математический аппарат, в частности – методы кластерного анализа.

1. Кластерные методы анализа рисков информационной безопасности

Для того, чтобы устранить недостатки, присущие качественной оценке рисков ИБ, используем кластерные методы анализа рисков [8, 9].

Преимущество использования методов кластерного анализа состоит в том, что они дают возможность проводить разбиение объектов не по одному признаку, а по целому ряду признаков. Кластеры – объединение нескольких однородных элементов, которое рассматривается как самостоятельная единица, обладающая определёнными свойствами. Суть методов кластерного анализа заключается в разбиении объектов на группы по признакам таким образом, чтобы каждый объект принадлежал только одному кластеру.

Существует множество методов кластеризации для формирования групп объектов, которые в целом схожи, но имеют разные критерии объединения в группы [10]. Будем использовать следующие методы кластерного анализа: метод ближайшего соседа, метод удаленного соседа, метод k-средних.

Алгоритмы методов кластерного анализа схожи и включают следующие шаги [11]:

1. Идентификация признаков, по которым будет проводиться кластеризация.
2. Определение выборки объектов для кластеризации.
3. Определение метрики и задание значения определяющего расстояния R между объектами, с помощью которого будет определяться сходство объектов.
4. Применение выбранного метода кластерного анализа.
5. Анализ полученных результатов.

2. Пример кластеризации рисков информационной безопасности

С целью демонстрации преимуществ методов кластерного анализа выполним группировку рисков ИБ на основе данных (2) методом обычной качественной оценки и с помощью методов кластерного анализа.

Пусть дано множество рисков ИБ, которое разобьем на группы, в которых первый параметр – степень реализации угрозы, второй – степень влияния угрозы на актив (оба параметра оцениваются экспертами по 10-бальной шкале):

$$x_1 = (5,\ 7);\ x_2 = (2,\ 1);\ x_3 = (1,\ 3);\ x_4 = (9,\ 8);\ x_5 = (5,\ 4);$$
$$x_6 = (10,\ 7);\ x_7 = (1,\ 9);\ x_8 = (8,\ 2);\ x_9 = (2,\ 8);\ x_{10} = (9,\ 3) \qquad (2)$$

Выполним группировку рисков ИБ на основе данных (2) методом обычной качественной оценки.

В результате было получено четыре группы рисков: $K_{75} = \{x_7; x_5\}$; $K_{32} = \{x_3; x_2\}$; $K_{10\,8\,1\,9} = \{x_{10}; x_8; x_1; x_9\}$; $K_{46} = \{x_4; x_6\}$. При этом, если риск находился на границе двух категорий, то он относился к более высокой группе, рис. 2.

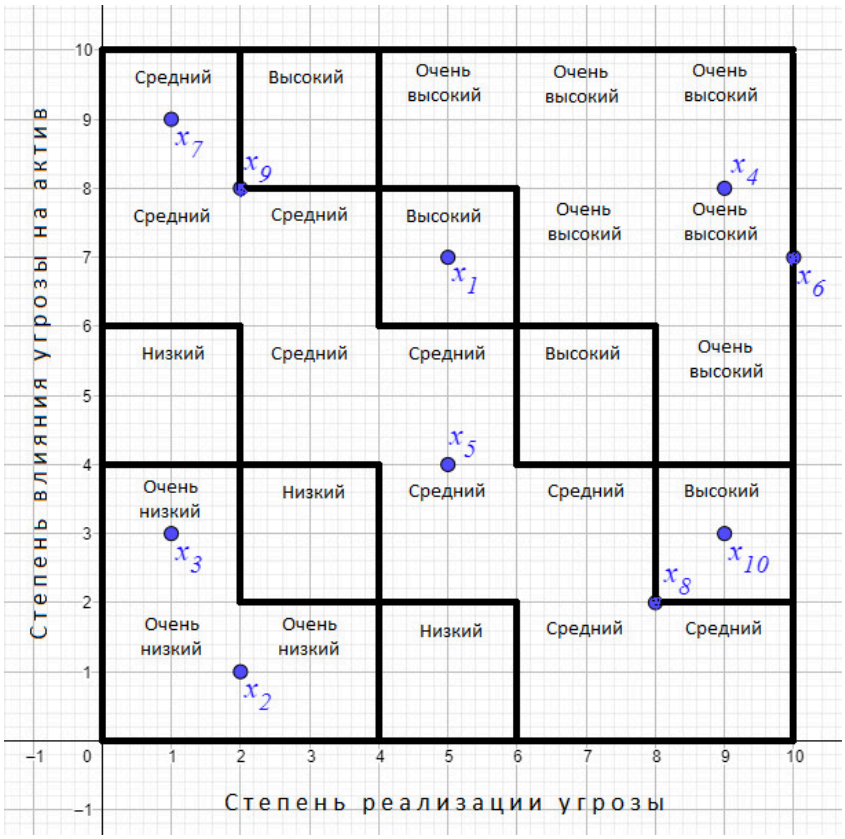

Рис. 2. Результат группирования рисков с помощью обычной качественной оценки
Fig. 2. The result of grouping risks using the usual qualitative assessment

Выполним разбиение рисков на группы с помощью методов кластерного анализа по вышеприведенному алгоритму для выбранных методов:

1. Идентификация признаков, по которым будет проводиться кластеризация.

Первый признак – степень реализации угрозы.

Второй признак – степень влияния угрозы на актив.

2. Определение выборки объектов для кластеризации.

Для сравнения с обычной качественной оценкой множество рисков ИБ берется из исходных данных (2).

Все исходные данные приводятся к единому диапазону значений (т.е. стандартизированы), что позволяет избежать некорректной кластеризации.

3. Определение метрики и задание значения определяющего расстояния R, с помощью которого будет определяться сходство объектов.

В качестве метрики выберем наиболее часто используемое для решения задач данного типа – евклидово расстояние ρ. Чем меньше это расстояние между объектами, тем они более схожи.

Задание значения определяющего расстояния R, с помощью которого будет определяться сходство объектов, выполняется лицом принимающим решение, у которого есть большой опыт в решении задач данного типа. Возьмем определяющее расстояние $R = 4$.

4. Применение выбранного метода кластерного анализа.

На данном этапе происходит применение выбранного метода кластерного анализа из трех вышеприведенных. В каждом методе свои правила формирования кластеров.

4.1. Метод ближайшего соседа [12].

В данном методе изначально каждый объект рассматривается как отдельный монокластер, например, объект x_1 образует монокластер $К_1$ и так далее. Между монокластерами рассчитывается евклидово расстояние ρ по выражению:

$$\rho_{(x_1;x_2)} = \left(\sum_{i=1}^{2}(x_{i1}-x_{i2})^2\right)^{1/2}, \quad (3)$$

где $x_1 -$ первый объект, $x_2 -$ второй объект, $i -$ признак объектов.

Если евклидово расстояние ρ между монокластерами меньше определяющего расстояния R, то они объединяются в новый кластер. Далее, находится евклидово расстояние ρ до ближайшего монокластера от ближайшего элемента новообразованного кластера.

Для упрощения решения задачи составим матрицу расстояний между монокластерами и уберём связи, где расстояние ρ больше определяющего расстояния R. Проведем вычисления расстояний ρ, подставив данные (2) в выражение (3). Результаты вычислений приведены в табл. 1 и табл. 2.

Таблица 1. Расстояния ρ между монокластерами

	$К_1$	$К_2$	$К_3$	$К_4$	$К_5$	$К_6$	$К_7$	$К_8$	$К_9$	$К_{10}$
$К_1$	0,0	6,7	5,7	4,1	3,0	5,0	4,5	5,8	3,2	5,7
$К_2$	6,7	0,0	2,2	9,9	4,2	10,0	8,1	7,1	7,0	7,3
$К_3$	5,7	2,2	0,0	9,4	4,1	9,8	6,0	7,1	5,1	8,0
$К_4$	4,1	9,9	9,4	0,0	5,7	1,4	8,1	6,1	7,0	5,0
$К_5$	3,0	4,2	4,1	5,7	0,0	5,8	6,4	3,6	5,0	4,1
$К_6$	5,0	10,0	9,8	1,4	5,8	0,0	9,2	5,4	8,1	4,1
$К_7$	4,5	8,1	6	8,1	6,4	9,2	0,0	9,9	1,4	10,0
$К_8$	5,8	7,1	7,1	6,1	3,6	5,4	9,9	0,0	8,5	1,4
$К_9$	3,2	7	5,1	7,0	5,0	8,1	1,4	8,5	0,0	8,6
$К_{10}$	5,7	7,3	8,0	5,0	4,1	4,1	10,0	1,4	8,6	0,0

Таблица 2. Расстояния ρ между монокластерами, где расстояние ρ меньше определяющего расстояния R

	$К_1$	$К_2$	$К_3$	$К_4$	$К_5$	$К_6$	$К_7$	$К_8$	$К_9$	$К_{10}$
$К_1$	0,0	0,0	0,0	0,0	3,0	0,0	0,0	0,0	3,2	0,0
$К_2$	0,0	0,0	2,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0
$К_3$	0,0	2,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
$К_4$	0,0	0,0	0,0	0,0	0,0	1,4	0,0	0,0	0,0	0,0
$К_5$	3,0	0,0	0,0	0,0	0,0	0,0	0,0	3,6	0,0	0,0
$К_6$	0,0	0,0	0,0	1,4	0,0	0,0	0,0	0,0	0,0	0,0
$К_7$	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,4	0,0
$К_8$	0,0	0,0	0,0	0,0	3,6	0,0	0,0	0,0	0,0	1,4
$К_9$	3,2	0,0	0,0	0,0	0,0	0,0	1,4	0,0	0,0	0,0
$К_{10}$	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,4	0,0	0,0

Результаты кластеризации методом ближайшего соседа (данные табл. 2) представлены на рис. 3.

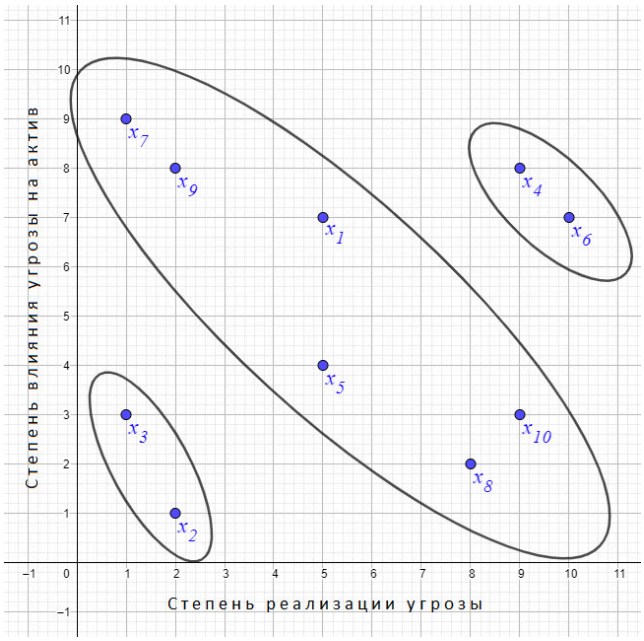

Рис. 3. Результаты кластеризации с помощью метода ближайшего соседа
Fig. 3. Results of clustering using the nearest neighbor method

4.2. Метод удаленного соседа [13].

В данном методе, по аналогии с предыдущим, каждый объект рассматривается как отдельный монокластер. Между монокластерами рассчитывается евклидово расстояние ρ по выражению (3).

Если евклидово расстояние между монокластерами ρ меньше определяющего расстояния *R*, то они объединяются в новый кластер. Разница между этим методом и предыдущим состоит в том, что на следующем этапе определяется расстояние ρ до ближайшего монокластера не от ближайшего элемента новообразованного кластера, а, наоборот, от самого дальнего.

Возьмем монокластер $К_7$ и найдем расстояние ρ до остальных монокластеров ($К_1$,…, $К_6$,) и ($К_8$, … , $К_{10}$), результаты представлены в табл. 3.

Таблица 3. Расстояние ρ от монокластера $К_7$ до остальных монокластеров

	$К_1$	$К_2$	$К_3$	$К_4$	$К_5$	$К_6$	$К_8$	$К_9$	$К_{10}$
$К_7$	4,5	8,1	6,0	8,1	6,4	9,2	9,9	1,4	10,0

Из анализа таблицы следует – наименьшее евклидово расстояние ρ, которое меньше определяющего расстояния R=4, – между седьмым и девятым монокластерами, которые и объединим в новообразованный монокластер $К_{79}$. Значит, кластер полностью сформирован.

Аналогично формируются остальные кластеры. Результаты кластеризации методом удаленного соседа представлены на рис. 4.

В двух примененных методах кластерного анализа нельзя с уверенностью утверждать, какое количество групп получится в итоге. При применении метода обычной качественной оценки было сформировано четыре группы, а в двух рассмотренных методах кластерного анализа – соответственно три и пять групп. Сравнению же подлежат методы, позволяющие получать одинаковое количество групп

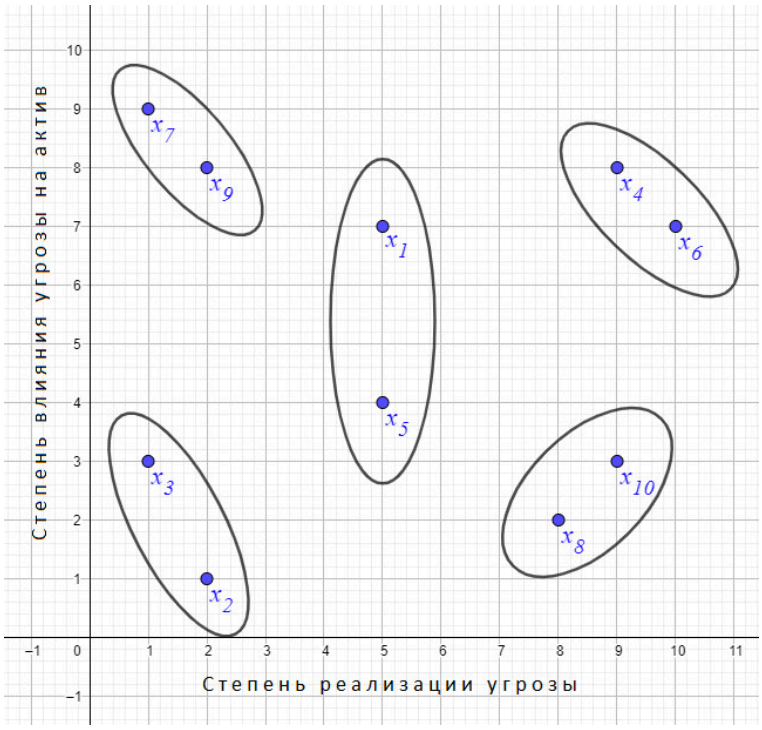

Рис. 4. Результаты кластеризации с помощью метода удаленного соседа
Fig. 4. Results of clustering using the remote neighbor method

4.3. Метод k-средних [14].

В начале данного метода задаются первоначальные центры для формирования кластеров. После этого, для каждого монокластера находится ближайший центр, который затем рассчитывается, как среднее значение параметров объектов, которые оказались ближе к нему. Процесс повторяется до тех пор, пока не будет изменений в распределении до и после.

Зададим 4 центра: $Z_1 = (2;2)$; $Z_2 = (2;8)$; $Z_3 = (8;2)$; $Z_4 = (8;8)$. Составим матрицу с расстояниями ρ от монокластеров до выбранных центров. Результаты представлены в табл. 4.

Таблица 4. Расстояния ρ между монокластерами и выбранными центрами

	$К_1$	$К_2$	$К_3$	$К_4$	$К_5$	$К_6$	$К_7$	$К_8$	$К_9$	$К_{10}$
Z_1	5,8	1,0	1,4	9,2	3,6	9,4	7,1	6,0	6,0	7,1
Z_2	3,2	7,0	5,1	7,0	5,0	8,1	1,4	8,5	0,0	8,6
Z_3	5,8	6,1	7,1	6,1	3,6	5,4	9,9	0,0	8,5	1,4
Z_4	3,2	9,2	8,6	1,0	5,0	2,2	7,1	6,0	6,0	5,1

В табл. 5 оставлены только наименьшие значения расстояний ρ для каждого из монокластеров. У кластеров $К_1$ и $К_5$ есть два равноудаленных центра, выберем тот, у которого значения показателей больше.

Таблица 5. Минимальные расстояния ρ между монокластерами и выбранными центрами

	$К_1$	$К_2$	$К_3$	$К_4$	$К_5$	$К_6$	$К_7$	$К_8$	$К_9$	$К_{10}$
Z_1		1,0	1,4		3,6					
Z_2	3,2						1,4		0,0	
Z_3					3,6			0,0		1,4
Z_4	*3,2*			1,0		2,2				

Получаем четыре новообразованных кластера:

$К_{79} = \{x_7; x_9\}$; $К_{32} = \{x_3; x_2\}$; $К_{10\,8\,5} = \{x_{10}; x_8; x_5\}$; $К_{461} = \{x_4; x_6; x_1\}$.

Затем рассчитаем новые центры:

$$Z_1 = \left(\frac{1+2}{2}; \frac{1+3}{2}\right) = (1{,}5; 2); \qquad Z_2 = \left(\frac{1+2}{2}; \frac{9+8}{2}\right) = (1{,}5; 8{,}5);$$
$$Z_3 = \left(\frac{5+8+9}{3}; \frac{4+2+3}{3}\right) = (7{,}3; 3); \quad Z_4 = \left(\frac{5+9+10}{3}; \frac{7+8+7}{3}\right) = (8; 7{,}3).$$

Снова найдем расстояния ρ до новообразованных центров Результаты представлены в табл. 6.

Таблица 6. Расстояния ρ от объектов до новообразованных центров

	$К_1$	$К_2$	$К_3$	$К_4$	$К_5$	$К_6$	$К_7$	$К_8$	$К_9$	$К_{10}$
Z_1	6,1	1,1	1,1	9,6	4,0	9,9	7,0	6,5	6,0	7,6
Z_2	3,8	7,5	5,5	7,5	5,7	8,6	0,7	9,2	0,7	9,3
Z_3	4,6	5,7	6,3	5,3	2,5	4,8	8,7	1,2	7,3	1,7
Z_4	3,0	8,7	8,2	1,2	4,5	2,0	7,2	5,3	6,0	4,4

Оставим в табл. 6 только минимальные расстояния ρ для каждого объекта. Результаты представлены в табл. 7.

Таблица 7. Минимальные расстояния ρ между монокластерами и новообразованными центрами

	$К_1$	$К_2$	$К_3$	$К_4$	$К_5$	$К_6$	$К_7$	$К_8$	$К_9$	$К_{10}$
Z_1		1,1	1,1							
Z_2							0,7		0,7	
Z_3					2,5			1,2		1,7
Z_4	3,0			1,2		2,0				

Кластеры остались теми же, что и были получены по данным табл. 5, значит процесс кластеризации завершен. Итоговые результаты кластеризации с помощью метода *k*-средних представлены на рис. 5.

5. Анализ полученных результатов.

Количество кластеров и их состав, полученные разными методами кластеризации, различны. На данном этапе проводится анализ полученных кластеров, трактовка специфики отдельно взятого кластера.

5.1. С помощью кластеризации методом *ближайшего соседа* было получено три кластера: $К_{32} = \{x_3; x_2\}$; $К_{1\,5\,7\,8\,9\,10} = \{x_1, x_5, x_7, x_8, x_9, x_{10}\}$; $К_{46} = \{x_4; x_6\}$. Риски из кластера $К_{32}$ можно охарактеризовать, как риски с низким уровнем вероятности возникновения угрозы и низким потенциалом ущерба. Напротив, риски из кластера $К_{46}$ имеют высокую степень реализации угрозы, а значит обладают большим потенциалом ущерба. Самый же большой кластер $К_{1\,5\,7\,8\,9\,10}$ включает в себя риски среднего уровня.

Особенностью данного метода является близкое расположение друг к другу монокластеров, из-за чего могут появляться большие кластеры, некоторые элементы которых могут иметь мало сходств.

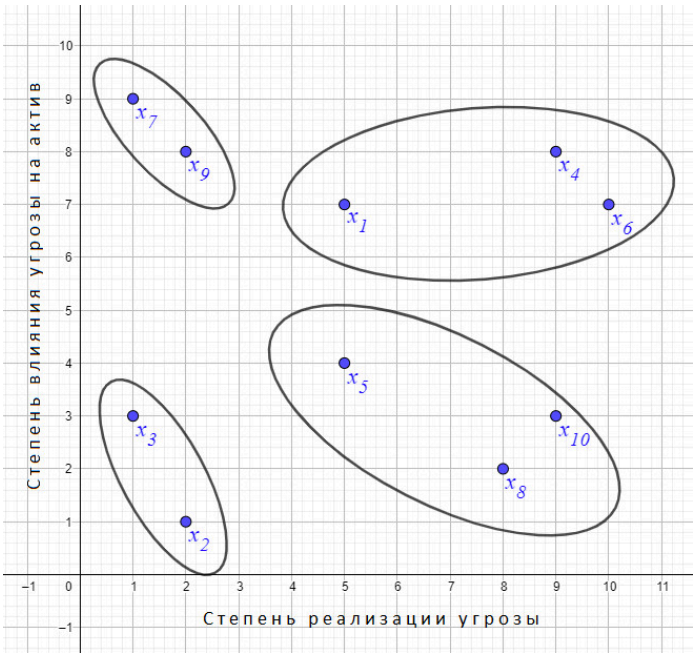

Рис. 5 Результаты кластеризации с помощью метода K-внутригрупповых средних
Fig. 5 Results of clustering using the K-intragroup mean method

5.2. С помощью кластеризации методом *удаленного соседа* было получено пять кластеров: $К_{79} = \{x_7; x_9\}$; $К_{32} = \{x_3; x_2\}$; $К_{10\,8} = \{x_{10}x_8\}$; $К_{51} = \{x_5; x_1\}$; $К_{46} = \{x_4; x_6\}$. Риски из кластера $К_{32}$ можно охарактеризовать, как очень низкие риски. Кластер $К_{10\,8}$, несмотря на высокую вероятность возникновения угрозы, включает в себя низкие риски и потенциальный ущерб небольшой. Кластер $К_{51}$ содержит средние риски. Кластер $К_{79}$ включает в себя высокие риски, у которых не очень большая степень реализации угрозы, но большой потенциальный ущерб. Кластер $К_{46}$ состоит из очень высоких рисков.

Данный метод является антиподом метода ближайшего соседа, в котором монокластеры располагаются близко друг к другу.

5.3. С помощью кластеризации методом *k-средних* были получены четыре кластера: $K_{79} = \{x_7; x_9\}$; $K_{32} = \{x_3; x_2\}$; $K_{10\,8\,5} = \{x_{10}; x_8; x_5\}$; $K_{461} = \{x_4; x_6; x_1\}$. Риски из кластера $К_{32}$ можно охарактеризовать как низкие риски. Кластер $К_{10\,8\,5}$ включает в себя средние риски, но, несмотря на высокую вероятность, потенциальный ущерб не велик. Кластер $К_{79}$ содержит высокие риски. Кластер $К_{461}$ состоит из очень высоких рисков.

Преимущество данного метода заключается в наличии возможности у экспертов заранее задать желаемое количество кластеров с помощью задания начальных центров.

Для сравнения результатов, полученных при обычной качественной оценке, с результатами, полученными методами кластерного анализа, находим среднее внутрикластерное расстояние для каждого метода, используя выражение для нахождения среднего внутрикластерного расстояния [15]:

$$d = \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho(x_i;\ x_j), \tag{4}$$

где $\rho(x_i;\ x_j)$ – евклидово расстояние между объектами x_i и x_j, n – количество объектов в кластере.

Затем находим среднее арифметическое средних внутрикластерных расстояний для всех кластеров в каждом методе. Результаты представлены в табл. 8.

Таблица 8. Сравнение результатов оценки рисков информационной безопасности с помощью различных методов

Название метода	Количество кластеров	Среднее арифметическое средних внутрикластерных расстояний
Метод ближайшего соседа	3	13,2
Метод удаленного соседа	5	3,8
Метод *k*-средних	4	6,7
Обычная качественная оценка	4	10,6

Анализ данных табл. 8 показывает, что среднее арифметическое средних внутрикластерных расстояний значительно различаются в методах с различным количеством кластеров.

Сравнению же подлежат только те данные средних арифметических средних внутрикластерных расстояний используемых методов, у которых формируется одинаковое количество кластеров. В данном случае необходимо сравнить средние арифметические средних внутрикластерных расстояний (чем оно меньше, тем точнее определены риски), полученных при обычной качественной оценке и методом *k*-средних, так как количество получившихся кластеров у них одинаково.

Из сравнения следует, что при одинаковом количестве кластеров метод *k*-средних дает более точное решение (оценка рисков) по сравнению с обычной качественной оценкой, так как среднее арифметическое средних внутрикластерных расстояний в этом случае меньше.

Заключение

В работе проведено исследование применения в менеджменте рисков ИБ методов кластерного анализа и показано, что их использование позволяет повысить точность оценки рисков. В результате, при дальнейшей обработке рисков это позволит избежать, с одной стороны, использования недостаточного количества мер (закупка и установка СЗИ) для обработки рисков (а это может привести к реализации угроз), а с другой стороны – перерасходования средств на закупку и установку СЗИ, когда на обработку рисков затраты компании окажутся больше, чем необходимо. Кроме того, использование методов кластерного анализа позволяет структурировать угрозы ИБ по степени влияния на реализацию угроз и степени влияния угроз на активы компаний. Эффективность применения методов кластерного анализа в оценке рисков ИБ очевидна.

СПИСОК ЛИТЕРАТУРЫ:

1. Bodin L.D., Gordon L.A., Loeb M.P. Information security and risk management // Communications of the ACM. 2008. P. 64–68. URL: https://www.researchgate.net/publication/220425249_Information_security_and_risk_management (дата обращения: 15.03.2021). DOI: https://doi.org/ 10.1145/1330311.1330325.
2. Campbell T. The Information Security Manager // Practical Information Security Management. 2016. P. 31–42. URL: https://www.researchgate.net/publication/311318229_Practical_Information_Security_Management (дата обращения: 15.03.2021). DOI: https://doi.org/10.1007/978-1-4842-1685-9.
3. Козунова С.С., Кравец А.Г. Формализованное описание процедуры управления рисками информационной системы // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. 2018 № 2. С. 61–70. URL: https://cyberleninka.ru/article/n/formalizovannoe-opisanie-protsedury-upravleniya-riskami-informatsionnoy-sistemy (дата обращения: 15.03.2021). DOI: https://doi.org/10.24143/2072-9502-2018-2-61-70.
4. Баранова Е., Мальцева А. Анализ рисков информационной безопасности для малого и среднего бизнеса // Директор по безопасности. 2015 № 9. С. 58–63. URL: https://publications.hse.ru/articles/157681360 (дата обращения: 15.03.2021).

5. Wangen G. Information Security Risk Assessment: A Method Comparison // JOURNAL OF LATEX CLASS FILES, VOL. 6, NO. 1, JANUARY 2007. P. 1–7. URL: https://ieeexplore.ieee.org/document/7912273 (дата обращения: 15.03.2021). DOI: https://doi.org/10.1109/MC.2017.107.
6. Куркина Е.П., Шувалова Д.Г. Оценка рисков: экспертный метод // Проблемы науки. 2017 № 1 (14). С. 63–39. URL: https://cyberleninka.ru/article/n/otsenka-riska-ekspertnyy-metod (дата обращения: 15.03.2021).
7. Винокур И.Р. Методика анализа и управления рисками. Количественная оценка рисков // Вестник Пермского национального исследовательского политехнического университета. Социально-экономические науки. 2020 № 1. С. 204–217. URL: https://cyberleninka.ru/article/n/metodika-analiza-i-upravleniya-riskami-kolichestvennaya-otsenka-riskov (дата обращения: 15.03.2021). DOI: https://doi.org/10.15593/2224-9354/2020.1.16.
8. Махрусе Н. Современные тенденции методов интеллектуального анализа данных: метод кластеризации // Московский экономический журнал. 2019 № 6. С. 359–377. URL: https://cyberleninka.ru/article/n/sovremennye-tendentsii-metodov-intellektualnogo-analiza-dannyh-metod-klasterizatsii (дата обращения: 15.03.2021). DOI: https://doi.org/10.24411/2413-046X-2019-16034.
9. Kettenring J.R. The practice of cluster analysis. Journal of Classification. 2006, 23. P. 3–30. URL: https://link.springer.com/article/10.1007/s00357-006-0002-6 (дата обращения: 15.03.2021). DOI: https://doi.org/10.1007/s00357-006-0002-6.
10. Тюрин А.Г., Зуев И.О. Кластерный анализ, методы и алгоритмы кластеризации // Вестник МГТУ МИРЭА. 2014 № 2 июнь 2014 выпуск 3. С. 86-97. URL: https://rtj.mirea.ru/upload/medialibrary/fba/09-tyurin.pdf (дата обращения: 15.03.2021).
11. Лось А.Б., Кабаов А.С., Трунцев В.И. Особенности использования кластерного анализа в системе менеджмента информационной безопасности // Промышленные контроллеры АСУ. 2013 № 8. С. 67–71. URL: https://publications.hse.ru/articles/145281528 (дата обращения: 15.03.2021).
12. Алексеева В.А., Калимуллина В.А. Применение метода ближайших соседей при моделировании кредитных рисков // Вестник Ульяновского государственного технического университета. 2014 № 3 (67). С. 54–56. URL: https://cyberleninka.ru/article/n/primenenie-metoda-blizhayshih-sosedey-pri-modelirovanii-kreditnyh-riskov (дата обращения: 15.03.2021).
13. Якимов А.И., Борчик Е.М., Башаримов В.В. Совместном использовании методов кластерного анализа многомерных данных // Доклады Белорусского государственного университета информатики и радиоэлектроники. 2011 № 5 (59). С. 95–102. URL: https://cyberleninka.ru/article/n/sovmestnom-ispolzovanii-metodov-klasternogo-analiza-mnogomernyh-dannyh (дата обращения: 15.03.2021).
14. Осипова Ю.А., Лавров Д.Н. Применение кластерного анализа методом k-средних для классификации текстов научной направленности // Математические структуры и моделирование. 2017 № 3 (43). С. 108–121. URL: https://cyberleninka.ru/article/n/primenenie-klasternogo-analiza-metodom-k-srednih-dlya-klassifikatsii-tekstov-nauchnoy-napravlennosti (дата обращения: 15.03.2021).
15. Герасимова Н.И. Метод кластеризации многомерных данных на основе модифицированного алгоритма функционирования карт Кохонена / Н.И. Герасимова; науч. рук. С. В. Аксёнов // Молодежь и современные информационные технологии : сборник трудов XIII Международной научно-практической конференции студентов, аспирантов и молодых ученых, г. Томск, 9-13 ноября 2015 г. : в 2 т. Томск : Изд-во ТПУ, 2016. Т. 1. С. 136–137. URL: http://earchive.tpu.ru/bitstream/11683/17107/1/conference_tpu-2015-C04-v1-059.pdf (дата обращения: 15.03.2021).

REFERENCES:

[1] Bodin L.D., Gordon L.A., Loeb M.P. Information security and risk management. Communications of the ACM. 2008. P. 64–68.
URL: https://www.researchgate.net/publication/220425249_Information_security_and_risk_management (accessed: 15.03.2021). DOI: https://doi.org/ 10.1145/1330311.1330325.
[2] Campbell T. The Information Security Manager. Practical Information Security Management. 2016. P. 31–42. URL:https://www.researchgate.net/publication/311318229_Practical_Information_Security_Management (accessed: 15.03.2021). DOI: https://doi.org/10.1007/978-1-4842-1685-9.
[3] Kozunova S.S., Kravets A.G. Formalized description of the information system risk management procedure. Bulletin of the Astrakhan State Technical University. Series: Management, Computer Engineering and Computer Science. 2018, no. 2. P. 61–70. URL: https://cyberleninka.ru/article/n/formalizovannoe-opisanie-protsedury-upravleniya-riskami-informatsionnoy-sistemy (дата обращения: 15.03.2021). DOI: https://doi.org/10.24143/2072-9502-2018-2-61-70 (in Russian).

[4] Baranova E., Maltseva A. Analysis of information security risks for small and medium-sized businesses. Director of Security. 2015, no. 9. P. 58–63. URL: https://publications.hse.ru/articles/157681360 (accessed: 15.03.2021) (in Russian).

[5] Wangen G. Information Security Risk Assessment: A Method Comparison JOURNAL OF LATEX CLASS FILES. Vol. 6, no. 1, JANUARY 2007. P. 1–7. URL: https://ieeexplore.ieee.org/document/7912273 (accessed: 15.03.2021). DOI: https://doi.org/10.1109/MC.2017.107.

[6] Kurkina E.P., Shuvalova D.G. Risk assessment: expert method. Problems of science. 2017, no. 1 (14). P. 63–39. URL: https://cyberleninka.ru/article/n/otsenka-riska-ekspertnyy-metod (accessed: 15.03.2021) (in Russian).

[7] Vinokur I.R. Methods of analysis and risk management. Quantitative risk assessment. Bulletin of the Perm National Research Polytechnic University. Socio-economic sciences. 2020, no. 1. P. 204–217. URL: https://cyberleninka.ru/article/n/metodika-analiza-i-upravleniya-riskami-kolichestvennaya-otsenka-riskov (accessed: 15.03.2021). DOI: https://doi.org/10.15593/2224-9354/2020.1.16 (in Russian).

[8] Mahruse N. Modern trends in data mining methods: clusterization method. Moscow Economic Journal. 2019, no. 6. P. 359–377. URL: https://cyberleninka.ru/article/n/sovremennye-tendentsii-metodov-intellektualnogo-analiza-dannyh-metod-klasterizatsii (accessed: 15.03.2021). DOI: https://doi.org/10.24411/2413-046X-2019-16034 (in Russian).

[9] Kettenring J.R. The practice of cluster analysis. J. Classif. 2006, 23. P. 3–30. URL: https://link.springer.com/article/10.1007/s00357-006-0002-6 (accessed: 15.03.2021). DOI: https://doi.org/10.1007/s00357-006-0002-6.

[10] Tyurin A.G., Zuev I.O. Cluster analysis, methods and clustering algorithms. Vestnik MGTU MIREA. 2014, no. 2, June 2014, issue 3. P. 86–97. URL: https://rtj.mirea.ru/upload/medialibrary/fba/09-tyurin.pdf (accessed: 15.03.2021) (in Russian).

[11] Los A.B., Kabov A.S., Trunci V.I. Features of using cluster analysis in the system of information security management. Industrial controllers ASU. 2013, no. 8. P. 67–71. URL: https://publications.hse.ru/articles/145281528 (accessed: 15.03.2021) (in Russian).

[12] Alekseev V.A., Kalimullina V.A. Application of the method of nearest neighbors in the modeling of credit risk. Vestnik of Ulyanovsk state technical University. 2014, no. 3 (67). P. 54–56. URL: https://cyberleninka.ru/article/n/primenenie-metoda-blizhayshih-sosedey-pri-modelirovanii-kreditnyh-riskov (accessed: 15.03.2021) (in Russian).

[13] Yakimov A.I., Borchik E.M., Basharimov V.V. Joint use of methods of cluster analysis of multidimensional data. Reports of the Belarusian State University of Informatics and Radioelectronics. 2011, no. 5 (59). P. 95–102. URL: https://cyberleninka.ru/article/n/sovmestnom-ispolzovanii-metodov-klasternogo-analiza-mnogomernyh-dannyh (accessed: 15.03.2021) (in Russian).

[14] Osipova Yu.A., Lavrov D.N. Application of cluster analysis by the k-means method for classification of scientific texts. Mathematical structures and modeling. 2017, no. 3 (43). P. 108–121. URL: https://cyberleninka.ru/article/n/primenenie-klasternogo-analiza-metodom-k-srednih-dlya-klassifikatsii-tekstov-nauchnoy-napravlennosti (accessed: 15.03.2021) (in Russian).

[15] Gerasimova N.I. Method of clusterization of multidimensional data on the basis of a modified algorithm for the functioning of Kohonen maps N.I. Gerasimova; scientific hands. S.V. Aksenov Youth and modern information technologies: proceedings of the XIII International Scientific and Practical Conference of Students, postgraduates and young scientists, Tomsk, November 9-13, 2015: in 2 vols. Tomsk: Publishing House of TPU, 2016. Vol. 1. P. 136–137. URL: http://earchive.tpu.ru/bitstream/11683/17107/1/conference_tpu-2015-C04-v1-059.pdf (accessed: 15.03.2021) (in Russian).

Поступила в редакцию – 19 марта 2021 г. Окончательный вариант – 22 апреля 2021г.
Received – March 19, 2021. The final version – April 22, 2021.