

**Т. А. Савчук, к. т. н., доц.; С. И. Петришин**

## **ОПРЕДЕЛЕНИЕ ЕВКЛИДОВОГО РАССТОЯНИЯ МЕЖДУ ЧРЕЗВЫЧАЙНЫМИ СИТУАЦИЯМИ НА ЖЕЛЕЗНОДОРОЖНОМ ТРАНСПОРТЕ ПРИ КЛАСТЕРНОМ АНАЛИЗЕ**

*В работе проанализирована возможность применения технологий Data Mining при анализе чрезвычайных ситуаций на железнодорожном транспорте. Формализована задача кластерного анализа, выявлены основные проблемы определения расстояния между чрезвычайными ситуациями при таком анализе. Определены понятия обычного, «взвешенного» и квадрата евклидова расстояния между чрезвычайными ситуациями на железнодорожном транспорте.*

**Ключевые слова:** евклидово расстояние, чрезвычайные ситуации, железнодорожный транспорт, кластерный анализ, расстояние, степень близости, квадрат евклидова расстояния, «взвешенное» евклидово расстояние.

### **Введение**

В связи с увеличением грузоперевозок железнодорожным транспортом, учитывая вредные и опасные грузы, актуальными становятся проблемы анализа данных о чрезвычайных ситуациях, которые могут произойти при транспортировке этих грузов. При использовании стандартных математических методов для анализа таких ситуаций существует вероятность получить недостоверные решения, которые при использовании системами поддержки принятия решений или аналитиками, увеличивают риск выполнения нецелесообразных или ошибочных действий ликвидационными подразделениями. Поэтому достоверный анализ с целью идентификации таких ситуаций является актуальной проблемой при разработке систем поддержки принятия решений относительно их ликвидации / уменьшения их последствий [1].

Технологии анализа данных, основанные на применении классических статистических подходов, имеют ряд недостатков при использовании для анализа чрезвычайных ситуаций на железнодорожном транспорте. Соответствующие методы основаны на использовании усредненных показателей, на основании которых трудно выяснить реальное состояние такой ситуации. Методы математической статистики оказались полезными, прежде всего, для проверки заранее сформулированных гипотез и «грубого» разведочного анализа, являющегося основой оперативной аналитической обработки данных.

Кроме того, стандартные статистические методы не учитывают нетипичные наблюдения, недопустимые при анализе чрезвычайных ситуаций на железной дороге. Однако отдельные нетипичные значения могут быть важными для исследования, характеризуя исключительные явления. При этом сама идентификация этих наблюдений и их последующий анализ, и подробное рассмотрение полезны для понимания сущности исследуемой чрезвычайной ситуации на железнодорожном транспорте. Как показывают современные исследования, именно такие события могут стать решающими относительно будущего поведения и развития чрезвычайной ситуации [1].

Анализируя чрезвычайные ситуации, которые могут возникнуть при перевозке опасных грузов, необходимо оперировать реальными значениями, организовать поиск неявных закономерностей в данных, самостоятельное построение гипотез о взаимосвязи параметров и характеристик таких ситуаций.

## Обзор существующих методов анализа чрезвычайных ситуаций на железнодорожном транспорте

Большое количество задач, в частности, и задач анализа чрезвычайных ситуаций на железнодорожном транспорте, помогают решить технологии Data Mining в зависимости от характера решаемых задач (задачи описания и задачи прогнозирования), как показано на рис. 1. Все алгоритмы анализа данных делят на supervised learning (обучение с учителем) и unsupervised learning (обучение без учителя) (рис. 1). В первом случае задача анализа решается в несколько этапов. Сначала с помощью определенного алгоритма строится модель данных, которые подлежат анализу. Затем эта модель изучает учебные выборки до того момента, пока она не начнет работать корректно. Unsupervised learning используют, когда нет никаких предварительных знаний об анализируемых данных. Основными из задач Data Mining являются: классификация, регрессия, поиск ассоциативных правил и кластеризация (табл. 1) [2]. Рассмотрим их применение на примере анализа чрезвычайных ситуаций на железной дороге.

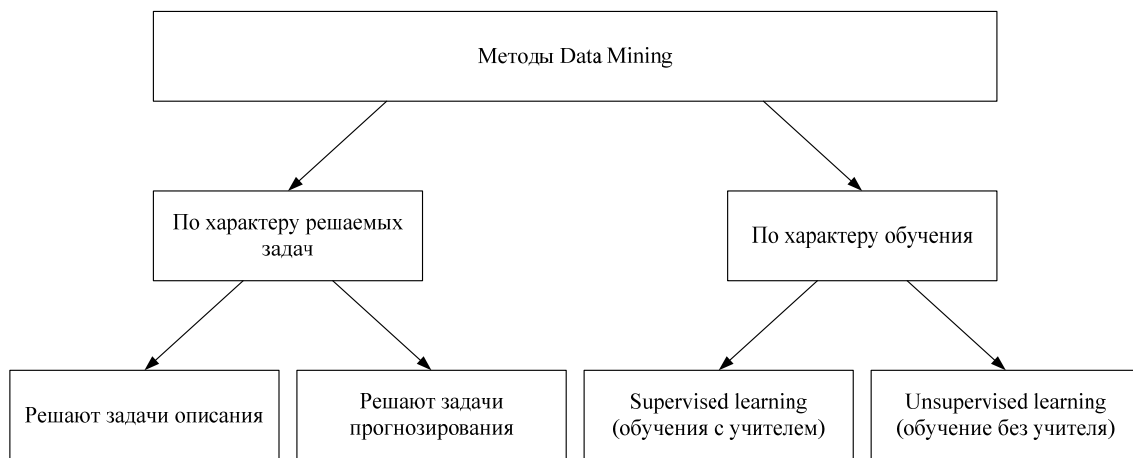


Рис. 1. Классификация методов, которые используются в технологиях Data Mining для анализа чрезвычайных ситуаций на железнодорожном транспорте

Задача классификации сводится к определению класса чрезвычайной ситуации на железнодорожном транспорте по ее характеристикам. В данной задаче множество классов чрезвычайных ситуаций, к которым может быть отнесен объект исследования, заранее известно.

Использование классификации при анализе чрезвычайных ситуаций является простым в применении и характеризуется наличием большого количества эффективных подходов к решению этой задачи, но это не является актуальным в исследуемых вопросах. Кроме того, следует отметить, что недостатки использования классификации в данном случае следующие [2]:

- 1) мощность обучающей выборки должна быть достаточно большой;
- 2) в учебную выборку должны входить чрезвычайные ситуации, представляющие все классы, что является проблемным при анализе таких ситуаций;
- 3) для каждого класса должно быть достаточно мощное множество чрезвычайных ситуаций в учебной выборке, что трудно получить при анализе чрезвычайных ситуаций на железной дороге;
- 4) проблема overfitting, сущность которой заключается в том, что классификационная функция хорошо адаптируется к данным, и, если среди них встречаются ошибки, либо аномальные значения, то функция интерпретирует их как часть внутренней структуры данных, что неприемлемо для анализа чрезвычайных ситуаций на железнодорожном транспорте;

5) проблема *underfitting*, которая заключается в том, что при проверке классификатора обнаруживается большое количество ошибок, что является неприемлемым предметной области, которая анализируется.

Задача регрессии подобно задаче классификации позволяет по известным характеристикам чрезвычайной ситуации на железнодорожном транспорте определить значение определенного ее параметра. В отличие от классификации значением этого параметра является не множество классов, а множество действительных чисел, что является неактуальным в анализе чрезвычайных ситуаций на железнодорожном транспорте.

Таблица 1

**Характеристика задач Data Mining, применяемых при анализе чрезвычайных ситуаций на железнодорожном транспорте**

| Название задачи            | Сущность задачи   | Преимущества задачи  | Недостатки задачи  |
|----------------------------|---|--|--|
| Классификация              | Определение класса чрезвычайной ситуации на железнодорожном транспорте за ее известными характеристиками                                  | Простота в применении; наличие большого количества эффективных подходов к решению данной задачи  | Мощность обучающей выборки должна быть достаточно большой; в учебную выборку должны входить чрезвычайные ситуации, которые представляют все классы; для каждого класса должна быть достаточно мощная множество чрезвычайных ситуаций в учебной выборке; проблема <i>overfitting</i> ; проблема <i>underfitting</i> |
| Регрессия                  | Позволяет по известным характеристикам чрезвычайной ситуации на железнодорожном транспорте определить значение определенного их параметра | Простота в применении; наличие большого количества подходов к решению данной задачи  | Невозможность решения задачи идентификации чрезвычайных ситуаций на железнодорожном транспорте   |
| Поиск ассоциативных правил | Определение зависимостей, которые часто встречаются между чрезвычайными ситуациями  | Возможность нахождения определенных закономерностей между чрезвычайными ситуациями; легкое восприятие человеком правил; простая интерпретация языков программирования  | Правила не всегда полезны, поскольку есть три вида ассоциативных правил: полезные, тривиальные, непонятны  |
| Кластеризация              | Поиск независимых кластеров в множестве данных о анализируемых чрезвычайных ситуациях на железнодорожном транспорте                       | Итерационный поиск оптимального результата; возможность использования методов образования кластеров, выбора признаков и мер близости между двумя объектами, объектом и кластером, двумя кластерами; построение научно обоснованной классификации многомерных наблюдений на основе совокупности отобранных показателей и выявление внутренних связей между чрезвычайными ситуациями на железной дороге, подлежащими анализу | Определение на входе числа итераций при поиске решения   |

Целью поиска ассоциативных правил является определение зависимостей, которые часто повторяются среди чрезвычайных ситуаций. Найденные зависимости представляются в виде

правил и могут быть использованы как для лучшего понимания природы анализируемых данных, так и для прогнозирования возникновения определенных событий [2].

Преимуществами поиска ассоциативных правил является то, что они позволяют находить определенные закономерности между чрезвычайными ситуациями, что является актуальным для данной предметной области. А также не важно при анализе таких ситуаций легкое восприятие человеком правил и простая интерпретация языков программирования.

Недостатком поиска ассоциативных правил является то, что правила, которые находят в результате такого анализа, не всегда полезны, поскольку есть три вида ассоциативных правил: полезные, тривиальные, непонятные. Такой исход является неприемлемым для анализа чрезвычайных ситуаций на железной дороге.

Задача кластеризации заключается в поиске независимых кластеров в множестве данных о анализируемых чрезвычайных ситуациях на железнодорожном транспорте. Это позволяет понять данные. Кроме того, группировка однородных данных позволяет уменьшить их количество для упрощения анализа в дальнейшем [2].

Преимуществами кластеризации является итерационный поиск оптимального результата, что повышает вероятность нахождения оптимального решения; возможность использования методов образования кластеров и выбора признаков и мер близости между двумя объектами, объектом и кластером, двумя кластерами, что актуально при анализе чрезвычайных ситуаций на железной дороге; построение научно обоснованной классификации многомерных наблюдений на основе совокупности отобранных показателей и выявление внутренних связей между чрезвычайными ситуациями на железной дороге, подлежащими анализу.

Сложность реализации кластеризации регулируется определением на входе числа итераций при поиске решения, что важно для анализируемой предметной области, поскольку таким образом определяется точность прогнозируемых результатов работы алгоритма идентификации чрезвычайной ситуации, подлежащей анализу и состояния ее развития.

Итак, по результатам проведенного анализа можно сделать вывод о целесообразности использования кластеризации для анализа чрезвычайных ситуаций на железнодорожном транспорте. Данные свидетельствуют о том, что кластеризация характеризуется итерационным поиском оптимального решения, возможностью выбора признаков и мер близости между двумя объектами, объектом и кластером, двумя кластерами, построением научно обоснованной классификации многомерных наблюдений на основе совокупности отобранных показателей и выявление внутренних связей между чрезвычайными ситуациями на железной дороге, подлежащими анализу.

### **Постановка проблемы**

Кластерный анализ представляет собой способ группировки многомерных объектов, которыми являются чрезвычайные ситуации на железнодорожном транспорте. Данный анализ основан на представлении результатов отдельных чрезвычайных ситуаций точками определенного геометрического пространства с последующим выделением групп этих точек (кластеров, таксонов). Кластерный анализ позволяет выделять компактные, отделенные друг от друга группы чрезвычайных ситуаций, предполагая «естественное» разбиение множества на области скопления таких ситуаций, что позволяет утверждать об однородности действий при ликвидации чрезвычайных ситуаций принадлежащих одному кластеру. Кластерный анализ используется для анализа чрезвычайных ситуаций в таких случаях [3]:

– данные о чрезвычайных ситуациях представлены в виде матрицы близости или расстояний между конкретными ситуациями (1):

$$\Psi = \begin{Bmatrix} 0 & \psi_{12} & \dots & \psi_{1n} \\ \psi_{21} & 0 & \dots & \psi_{2n} \\ \dots & \dots & \dots & \dots \\ \psi_{n1} & \psi_{n2} & \dots & 0 \end{Bmatrix}, \quad (1)$$

где  $\psi_{ij}$  – расстояние между параметрами векторов  $\psi(Y_i, Y_j)$ , где  $Y_i$  и  $Y_j$  – чрезвычайные ситуации на железнодорожном транспорте;

$n$  – количество чрезвычайных ситуаций на железнодорожном транспорте, информацию о которых нужно обрабатывать (мощность базы данных чрезвычайных ситуаций на железнодорожном транспорте).

– данные о чрезвычайных ситуациях представлены в виде точек в многомерном пространстве (2):

$$Y = \{Y_1, Y_2, \dots, Y_n\} = \begin{Bmatrix} y_{11} & y_{12} & \dots & y_{1m} \\ y_{21} & y_{22} & \dots & y_{2m} \\ \dots & \dots & \dots & \dots \\ y_{n1} & y_{n2} & \dots & y_{nm} \end{Bmatrix}, \quad (2)$$

где  $Y_i$  – конкретная чрезвычайная ситуация на железнодорожном транспорте;

$y_{ij}$  – значение конкретного  $j$ -го параметра  $i$ -ой чрезвычайной ситуации;

$m$  – количество параметров чрезвычайных ситуаций, сохраненных в базе данных.

При анализе чрезвычайных ситуаций возможно использовать два приведенных способа представления данных. Для этого нужно обрабатывать мощные базы данных, поэтому в этом случае второй способ (2) является более целесообразным, поскольку матрица близости при увеличении базы данных чрезвычайных ситуаций на железнодорожном транспорте на одну такую ситуацию увеличивается на  $((l+1)^2 - l^2)$  элементов, где  $l$  – начальное количество чрезвычайных ситуаций на железнодорожном транспорте в базе данных. Также недостатком первого способа представления данных является то, что при добавлении новой чрезвычайной ситуации в базу данных нужно вычислять (используя вычислительную технику или без нее) ее степени близости или расстояния до каждой сохраненной в базе данных чрезвычайной ситуации. И лицу, принимающему решение о ликвидации чрезвычайной ситуации неизвестны координаты каждой чрезвычайной ситуации на железнодорожном транспорте. Что касается второго способа, то здесь при добавлении в базу данных новой чрезвычайной ситуации число ее элементов увеличится на  $m$ . Также преимуществом второго способа представления данных является возможность определить координаты каждой чрезвычайной ситуации как геометрического объекта.

Итак, задачу кластерного анализа чрезвычайных ситуаций на железнодорожном транспорте можно сформулировать таким образом.

Пусть существует множество чрезвычайных ситуаций на железнодорожном транспорте  $Y = \{Y_i\} (i = \overline{1, n})$ . Каждая из чрезвычайных ситуаций имеет  $m$  характеристик. Нужно разбить статический  $m$ -мерный диапазон изменения значений анализируемых признаков чрезвычайных ситуаций на интервалы группировки. То есть, множество  $Y$  разбить на  $k (k \leq n)$  кластеров таким образом, чтобы конкретная чрезвычайная  $Y_i$  ситуация принадлежала одному и только одному кластеру. Основным условием при этом является максимальное сходство чрезвычайных ситуаций, принадлежащих одному кластеру, и максимальное различие чрезвычайных ситуаций из разных кластеров [3].

### Расстояния и степени близости чрезвычайных ситуаций на железнодорожном транспорте

Сложности в формализации задачи кластерного анализа чрезвычайных ситуаций на Наукові праці ВНТУ, 2010, № 3

железнодорожном транспорте связанные с определением понятий их однородности [4] и слабой структурированностью данных.

В общем случае однородность двух  $i$ -ой и  $j$ -ой чрезвычайной ситуаций на железнодорожном транспорте определяется заданием правила расчета величины  $\psi_{ij}$ , характеризующей или расстояние  $a(Y_i, Y_j)$  между объектами  $Y_i$  и  $Y_j$  из исследуемого множества чрезвычайных ситуаций на железнодорожном транспорте  $Y = \{Y_i\} (i = \overline{1, n})$ , или степень близости  $\omega(Y_i, Y_j)$  между теми же ситуациями. Если заданная функция  $a(Y_i, Y_j)$ , то близкие по значению этой метрики чрезвычайные ситуации считаются однородными, то есть принадлежащими одному кластеру. Но при этом необходимо сравнивать  $a(Y_i, Y_j)$  с определенными пороговыми значениями, которые определяются в каждом случае. Обозначенный подход целесообразно использовать для определения меры близости  $\omega(Y_i, Y_j)$  при формировании однородных кластеров чрезвычайных ситуаций на железнодорожном транспорте. При этом должны быть соблюдены следующие требования:

- требование симметрии ( $\omega(Y_i, Y_j) = \omega(Y_j, Y_i)$ );
- требование максимального сходства чрезвычайных ситуаций ( $\omega(Y_i, Y_i) = \max(\omega(Y_i, Y_j))$ );
- требование соответствия расстояния между чрезвычайными ситуациями на железнодорожном транспорте и степени близости между ними (если  $a(Y_1, Y_2) \geq a(Y_2, Y_3)$  то  $\omega(Y_1, Y_2) \leq \omega(Y_2, Y_3)$ ).

#### Измерение расстояния между чрезвычайными ситуациями

Расстоянием между чрезвычайными ситуациями  $Y_i$  и  $Y_j$  или метрикой называется неотъемлемая действительная функция  $a(Y_i, Y_j)$ , если [5]:

- $a(Y_i, Y_j) \geq 0$  для всех  $Y_i$  и  $Y_j$  с множества  $Y = \{Y_i\} (i = \overline{1, n})$ ;
- $a(Y_i, Y_j) = 0$  тогда и только тогда, когда  $Y_i = Y_j$ ;
- $a(Y_i, Y_j) = a(Y_j, Y_i)$ ;
- $a(Y_i, Y_j) \leq a(Y_i, Y_k) + a(Y_k, Y_j)$ , где  $Y_i, Y_j$  и  $Y_k$  – любые три чрезвычайные ситуации на железнодорожном транспорте с множества  $Y = \{Y_i\} (i = \overline{1, n})$ .

При кластерном анализе чрезвычайных ситуаций возникает проблема измерения расстояния между отдельными ситуациями. Основные трудности, возникающие при этом такие [5]:

- неоднозначность выбора способа нормирования;
- неоднозначность определения расстояния между объектами.

Если рассмотреть результаты исследований некоторых чрезвычайных ситуаций на железнодорожном транспорте, то на рис. 2 по результатам этих исследований построим корреляционное поле. Масштабность за осями избирается произвольно. На рис. 2 (а) выделены определенные классы А, Б, В, при изменении масштаба оси «Температура» (рис. 2 (б)) меняются – формируются класс А<sub>1</sub>, который идентичен предыдущей визуализации, и класс Б<sub>1</sub>, который объединяет классы Б и В, которые являются недопустимым при анализе чрезвычайных ситуаций на железной дороге. Определить расстояние между чрезвычайными ситуациями в данном случае невозможно, поскольку признаки измерены в разных единицах. Нужно нормировать показатели, то есть перевести их в безразмерные величины с целью корректного измерения расстояния между чрезвычайными ситуациями на железнодорожном транспорте. Нормирование представляет собой переход к определенному одинаковому описанию всех признаков, до введения новой условной единицы измерения, что позволяет

допускать формальное сравнение чрезвычайных ситуаций на железнодорожном транспорте.

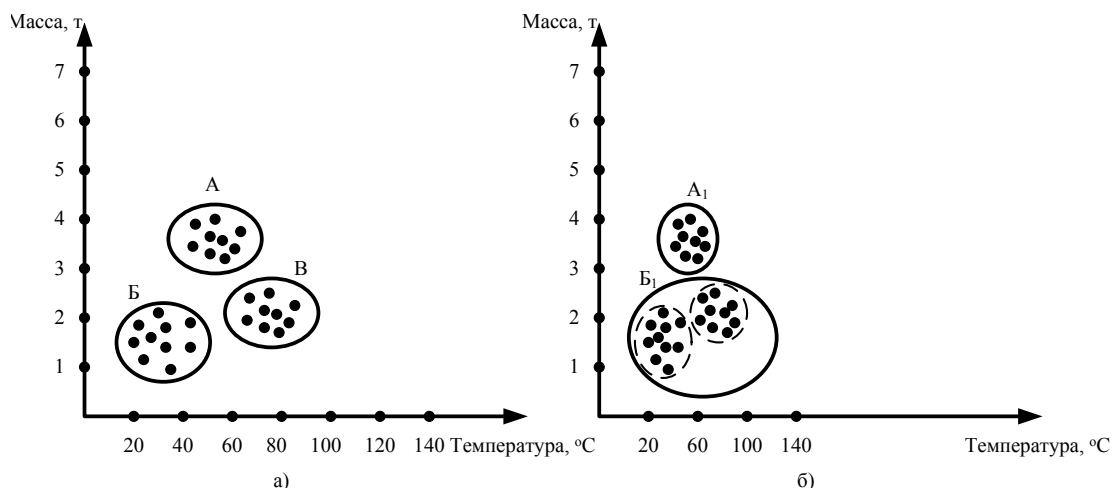


Рис. 2. Разбивка совокупности чрезвычайных ситуаций на железнодорожном транспорте на кластеры в зависимости от масштабов измерения переменных

### Использование евклидова расстояния при анализе чрезвычайных ситуаций на железнодорожном транспорте

Евклидово расстояние [6] является одним из наиболее используемых метрик в кластерном анализе, поскольку оно соответствует интуитивным представлениям о близости и своей квадратичной формой соответствует классическим статистическим конструкциям. Геометрически данную метрику целесообразно использовать для объединения объектов в скоплениях, которые являются типичными для слабо коррелированных множеств.

Формула общего евклидова расстояния имеет вид (3):

$$a_E(Y_i, Y_j) = \sqrt{(y_{i1} - y_{j1})^2 + (y_{i2} - y_{j2})^2 + \dots + (y_{im} - y_{jm})^2}, \quad (3)$$

где  $a_E(Y_i, Y_j)$  – евклидово расстояния между двумя чрезвычайными ситуациями на железнодорожном транспорте  $Y_i$  и  $Y_j$ ;

$y_{i1}, y_{i2}, \dots, y_{im}$  – вектор значений характеристик, описывающий  $i$ -ую чрезвычайную ситуацию на железнодорожном транспорте;

$y_{j1}, y_{j2}, \dots, y_{jm}$  – вектор значений характеристик, описывающий  $j$ -ую чрезвычайную ситуацию на железнодорожном транспорте.

Данную метрику целесообразно применять в следующих случаях:

- значения параметров  $y_{i1}, y_{i2}, \dots, y_{im}$  однородные по своему физическому смыслу, и если установлено, что все они одинаково важны с точки зрения решения задачи об отнесении чрезвычайной ситуации на железнодорожном транспорте до определенного кластера;

- пространство признаков совпадает с геометрическим пространством действительности и понятия близости чрезвычайных ситуаций совпадает с понятием геометрической близости в этом пространстве.

Следовательно, данную метрику можно применять при анализе чрезвычайных ситуаций на железнодорожном транспорте в случаях, когда будут анализироваться такие характеристики, которые будут близки по своему физическому содержанию, что является неприемлемым для данной предметной области, поскольку нужно анализировать все факторы для получения достоверного результата.

Чтобы лучше различать отдаленные объекты иногда используют квадрат евклидова расстояния [6] (4):

$$a_E(Y_i, Y_j)^2 = (y_{i1} - y_{j1})^2 + (y_{i2} - y_{j2})^2 + \dots + (y_{im} - y_{jm})^2. \quad (4)$$

При необходимости определения «веса» каждой характеристики  $\lambda_l$  чрезвычайной ситуации на железнодорожном транспорте (например, температура в цистерне при перевозке легковоспламеняющихся веществ и относительная влажность воздуха), которая будет пропорциональной степени его важности с точки зрения отнесения определенной чрезвычайной ситуации к конкретному кластеру, целесообразно использовать «взвешенное» евклидово расстояние [4] (5):

$$a_{3E}(Y_i, Y_j) = \sqrt{\lambda_1 \cdot (y_{i1} - y_{j1})^2 + \lambda_2 \cdot (y_{i2} - y_{j2})^2 + \dots + \lambda_m \cdot (y_{im} - y_{jm})^2}, \quad (5)$$

где  $a_{3E}(Y_i, Y_j)$  – «взвешенное» евклидово расстояния между двумя чрезвычайными ситуациями на железнодорожном транспорте  $Y_i$  и  $Y_j$ ;

$\lambda_1, \lambda_2, \dots, \lambda_m$  ( $0 \leq \lambda_l \leq 1 (l = \overline{1, m})$ ) – вектор значений весовых коэффициентов, соответствующих характеристикам  $y_1, y_2, \dots, y_m$  чрезвычайных ситуаций на железнодорожном транспорте;

$y_{i1}, y_{i2}, \dots, y_{im}$  – вектор значений характеристик, описывающий  $i$ -ую чрезвычайную ситуацию на железнодорожном транспорте;

$y_{j1}, y_{j2}, \dots, y_{jm}$  – вектор значений характеристик, описывающий  $j$ -ую чрезвычайную ситуацию на железнодорожном транспорте.

Для определения вектора значений весовых коэффициентов  $\lambda_1, \lambda_2, \dots, \lambda_m$  используют учебные выборки чрезвычайных ситуаций на железнодорожном транспорте или опыт экспертов. Попытки определить весовые коэффициенты  $\lambda_1, \lambda_2, \dots, \lambda_m$  только по информации, содержащейся в исходных данных, не дают нужного результата и могут увеличить погрешность полученного результата.

Итак, данная метрика является приемлемой для проведения кластерного анализа чрезвычайных ситуаций на железнодорожном транспорте, поскольку она учитывает значимость каждой характеристики чрезвычайной ситуации на железнодорожном транспорте, что повышает достоверность результата.

### Вывод

Таким образом, для анализа чрезвычайных ситуаций на железнодорожном транспорте целесообразно использовать кластеризацию, которая характеризуется итерационным поиском оптимального решения; возможностью выбора признаков и мер близости между двумя объектами, объектом и кластером, двумя кластерами; построением научно обоснованной классификации многомерных наблюдений на основании совокупности отобранных показателей и выявлением внутренних связей между чрезвычайными ситуациями на железной дороге подлежащими анализу.

Среди рассмотренных метрик целесообразным для использования при анализе чрезвычайных ситуаций на железной дороге является «взвешенное» евклидово расстояние, учитывающее значимость каждой характеристики чрезвычайной ситуации на железнодорожном транспорте, что способствует получению достоверного результата.

### СПИСОК ЛИТЕРАТУРЫ

1. Савчук Т. О. Використання ієрархічних методів кластеризації для аналізу надзвичайних ситуацій на залізничному транспорті / Т. О. Савчук, С. І. Петришин // Міжнародний науково-технічний журнал «Вимірювальна та обчислювальна техніка в технологічних процесах» (м. Хмельницький) – 2009. – №1 – С.193 – 198.
2. Барсегян А. А. Методы и модели анализа данных: OLAP и Data Mining. / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод – СПб.: БХВ-Петербург, 2004. – 336 с.



3. Савчук Т. О. Порівняльний аналіз використання методів кластеризації для ідентифікації надзвичайних ситуацій на залізничному транспорті / Т. О. Савчук, С. І. Петришин // Наукові праці Донецького національного технічного університету. – Серія «Інформатика, кібернетика і обчислювальна техніка». – 2010. – Випуск 11(134). – С. 135 – 141.

4. Айвазян С. А. Прикладная статистика: Классификация и снижение размерности. / С. А. Айвазян, В. М. Бухштабер, И. С. Енюков. – М.: Финансы и статистика, 1989. – 607 с.

5. Мандель И. Д. Кластерный анализ. / И. Д. Мандель. – М.: Финансы и статистика, 1988. – 176с.

6. Дюран Б. Кластерный анализ / Б. Дюран, П. Одел.; пер. с англ. – М.: Статистика, 1977. – 128 с.

**Савчук Тамара Александровна** – к. т. н., доцент кафедры компьютерных наук.

**Петришин Сергей Иванович** – студент 5 курса кафедры компьютерных наук.

Винницкий национальный технический университет.