

Е.М. Герасименко, Д.Ю. Кравченко, Ю.А. Кравченко, Э.В. Кулиев**ПОДДЕРЖКА ПРИНЯТИЯ РЕШЕНИЙ ПО ПРЕДУПРЕЖДЕНИЮ
И ЛИКВИДАЦИИ ПОСЛЕДСТВИЙ ЧРЕЗВЫЧАЙНЫХ СИТУАЦИЙ
НА ОСНОВЕ НЕЧЕТКОГО МЕТОДА СТРУКТУРИРОВАНИЯ
ИНФОРМАЦИИ***

Статья посвящена решению научной проблемы поддержки принятия решений по предупреждению и ликвидации последствий чрезвычайных ситуаций (ЧС) на основе решения задачи структурирования информации. Актуальность данной задачи обусловлена необходимостью развития теоретических основ оптимизации риска возникновения неблагоприятных воздействий на здоровье человека и окружающую природную среду в связи с чрезвычайными ситуациями (ЧС). Авторами даны определения основным терминам исследуемой предметной области. Представлена формализованная постановка решаемой задачи. Приведена развернутая классификация чрезвычайных ситуаций с описанием особенностей представленных классов. Система правил для поддержки принятия решений при чрезвычайных ситуациях должна иметь многоуровневую иерархию, что позволяет обеспечить построение различных траекторий принятия решений по принципу сверху-вниз. Наиболее подходящей моделью для построения такого информационного пространства является онтологическая структура, которая обеспечивает создание необходимой многоуровневой иерархии с учетом всех параметров и критериев, влияющих на развитие ситуации. Основными элементами данной онтологической модели являются сущности и связи между ними, наличие которых на верхнем уровне декомпозиции будет указывать на риск возникновения чрезвычайной ситуации, а на каждом более низком уровне – будет разворачивать таксономию детального описания возможных ситуаций развития ЧС и необходимых действий для их предотвращения или ликвидации последствий. Обработка построенной онтологической модели правил реализуется в работе на основе применения предложенного авторами нечеткого метода структурирования информации при ЧС, который отличается от известных аналогов применением нового обобщенного критерия оптимизации выбора альтернатив поддержки принятия решений. Оригинальность оптимизационной постановки задачи структурирования заключается в оценке контекстной привязки элементов информации к определенному классу чрезвычайных ситуаций, междисциплинарности с учетом наличия множества связей между предметными областями, а также учета снижения уровня оперативности информации о протекании ЧС с течением времени.

Чрезвычайные ситуации; поддержка принятия решений; нечеткие правила; онтологии; классификация; структурирование информации.

E.M. Gerasimenko, D.Yu. Kravchenko, Yu.A. Kravchenko, E.V. Kuliev**DECISION SUPPORT FOR PREVENTION AND ELIMINATION
OF THE EMERGENCIES' CONSEQUENCES BASED
ON THE INFORMATION STRUCTURING FUZZY METHOD**

The article is devoted to solving the scientific problem of decision support for the prevention and elimination of emergencies' consequences based on solving the problem of structuring information. The relevance of this task is due to the need to develop theoretical foundations for optimizing the risk of adverse effects on human health and the environment in connection with emergencies. The authors give definitions to the main terms of the studied subject area. A formalized statement of the problem to be solved is presented. A detailed emergencies' classification with a description of the presented classes' features is given. The system of rules for decision support in emergencies should have a multi-level hierarchy, which allows for the construction of various

* Исследование выполнено за счет гранта Российского научного фонда № 22-71-10121, <https://rscf.ru/project/22-71-10121/> в Южном федеральном университете.

decision-making trajectories on a top-down basis. The most suitable model for building such an information space is an ontological structure that provides the creation of the necessary multi-level hierarchy, taking into account all the parameters and criteria that affect the development of the situation. The main elements of this ontological model are entities and relationships between them, the presence of which at the upper level of decomposition will indicate the risk of an emergency, and at each lower level it will expand the taxonomy of a detailed description of emergencies' possible situations and the necessary actions to prevent or eliminate them consequences. The processing of this ontological model of rules is implemented on the basis of the structuring information fuzzy method proposed by the authors in emergencies, which differs from known analogs by the use of a new generalized criterion for optimizing the choice of decision support alternatives. The originality of the optimization formulation of the structuring problem lies in the assessment of the information elements contextual binding to a certain class of emergency situations, interdisciplinary, taking into account the presence of many links between subject areas, as well as taking into account the decrease in the level of information efficiency about the course of emergencies over time.

Emergencies; decision support; fuzzy rules; ontologies; classification; information structuring.

Введение. Эффективное решение оптимизационной задачи минимизации риска возникновения неблагоприятных воздействий на здоровье человека и окружающую природную среду в связи с чрезвычайными ситуациями (ЧС) требует проведения классификации основных типов ЧС и определения базовых критериев оценки действий ответственных лиц в процессе предотвращения и/или устранения последствий ЧС.

Дадим определение понятию **чрезвычайная ситуация**. Под термином ЧС будем понимать некоторые непредвиденные условия, приведшие к возникновению инфекционных болезней, аварий, катастроф, природного стихийного или техногенного бедствия, что создает на определенной территории обстановку, которая в ближайшей перспективе приводит к ущербу здоровью людей или к человеческим жертвам, а также к значительным материальным потерям и ущербу природной среде с последующим ухудшением качества среды жизнедеятельности людей. Источник любого типа ЧС имеет два основных класса **поражающих факторов** (ПФ) – *первичный* и *вторичный* [1–5]. Первичный ПФ наступает одновременно с началом чрезвычайной ситуации, а вторичный – инициируется первичным позднее, либо не проявляется вообще. Поражающие факторы имеют **виды воздействия** представленные на рис. 1.

Центральным фактором, приводящим к возникновению неблагоприятных воздействий на здоровье человека и окружающую природную среду принято считать **опасность** [1]. Актуализация понятия «опасность» напрямую связана с появлением некоторого **источника опасности** – процесса в определенной области исследуемого пространства, воздействие которого оцениваются как вредоносные (опасные природные явления, стихийные бедствия). Техногенная сфера наиболее всего подвержена воздействию источников опасности на **потенциально опасные объекты** [1–4].

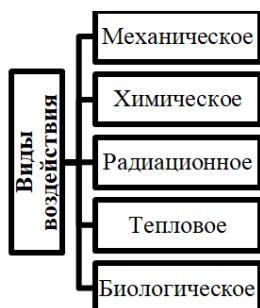


Рис. 1. Виды воздействия поражающих факторов

Основными критериями оценки и сравнения опасностей ЧС являются: условия неприемлемости (недопустимости) определенной опасности; формализация границы между опасностью и безопасностью. Данные нечеткие критерии сводятся к оценке опасности на основе определения значения показателя **риска**.

1. Постановка задачи оптимизации риска возникновения чрезвычайных ситуаций. В работе [5] под термином «риск» понимают величину, которая включает в себя вероятность возникновения опасных событий и среднее значение ущерба от них. Таким образом, показатель риска может быть представлен в виде следующего выражения:

$$R = PD \rightarrow \min, \quad (1)$$

где **P** – частота возникновения опасного события в течение года (вероятность), **D** – среднее значение ущерба от данного опасного события.

Получение численного значения показателя риска позволяет оценить опасность, а также разработать и спрогнозировать меры по ее снижению. Решение данной оптимизационной задачи подразумевает сохранение определенного баланса между основным критерием и ограничениями решения задачи. Любые мероприятия по снижению риска подразумевают появление весомой статьи расходов, что для любого вида деятельности имеет критичное значение предела величины таких расходов, именно поэтому должен соблюдаться баланс между прогнозом возникновения возможных опасных ситуаций и рентабельностью проводимой деятельности. С данным балансом напрямую связано понятие **приемлемого риска** [5]. Определение значений приемлемого риска требует комплексного рассмотрения проблемы, учитывающего технические, экономические, социальные, психологические и иные аспекты.

Подобная задача имеет высокую научную сложность из-за наличия информационной неопределенности комплексного учета разнородных характеристик, которые в ряде случаев невозможно нормировать и привести к единой шкале измерения. С учетом указанной информационной неопределенности показателей оценки опасности тех или иных событий в настоящее время защищенность человека и окружающей среды от ЧС оцениваются не по показателям качества жизни, а по предельным величинам характеристик надежности и эффективности технических систем, от которых прямо или косвенно зависит безопасность. Поэтому в данном случае процесс повышения безопасности жизнедеятельности имеет сугубо технический характер и строится на основе подхода к учету последствий возникающих ЧС. В связи с этим принята следующая **классификация безопасности** (рис. 2):

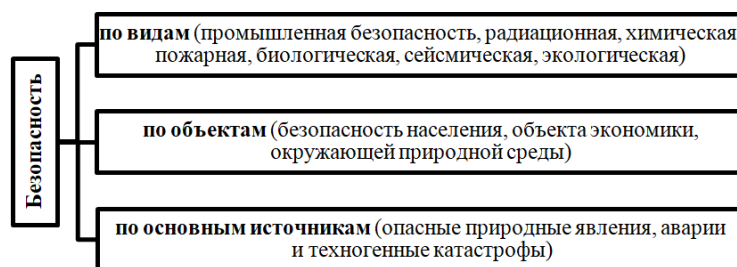


Рис. 2. Классификация безопасности

Несмотря на спорный характер такого технического подхода к оценке защищенности человека от ЧС, он в настоящее время является единственно приемлемым [1–3], позволяющим преодолеть проблему информационной неопределенности, а также контролировать и снижать опасность возникновения чрезвычайных ситуаций.

2. Классификация чрезвычайных ситуаций. В базовой терминологии [1] ЧС делят на два основных класса: **конфликтные** и **бесконфликтные**. Данные классы имеют четкую границу разделения признаков, характеристики и значения которых практически не пересекаются. Класс конфликтных ЧС имеет следующие основные подклассы (рис. 3):

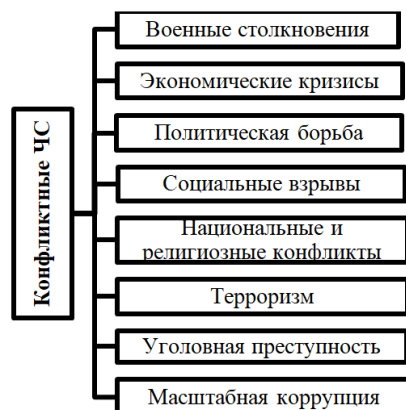


Рис. 3. Классификация конфликтных ЧС

Классификация бесконфликтных ЧС является более широкой и развернутой. Данную классификацию необходимо рассматривать по многим признакам, которые характеризуют явления с разных сторон [1]. Представим такую таксономию бесконфликтных ЧС (рис. 4).

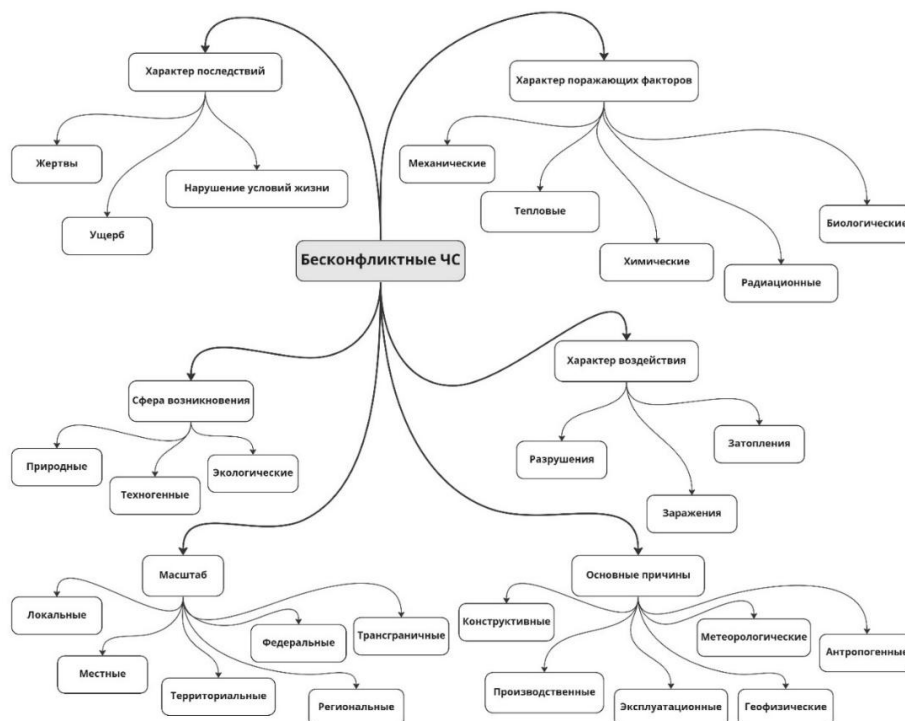


Рис. 4. Классификация бесконфликтных ЧС

Рассмотрим более подробно надкласс ЧС, связанный с определением сферы их возникновения [1–5]. Раскроем в данном классе состав **трёх базовых подклассов чрезвычайных ситуаций: ЧС природного характера; ЧС техногенного характера; ЧС экологического характера**. Составы данных подклассов ЧС опишем на примере, проиллюстрированном на рис. 5.

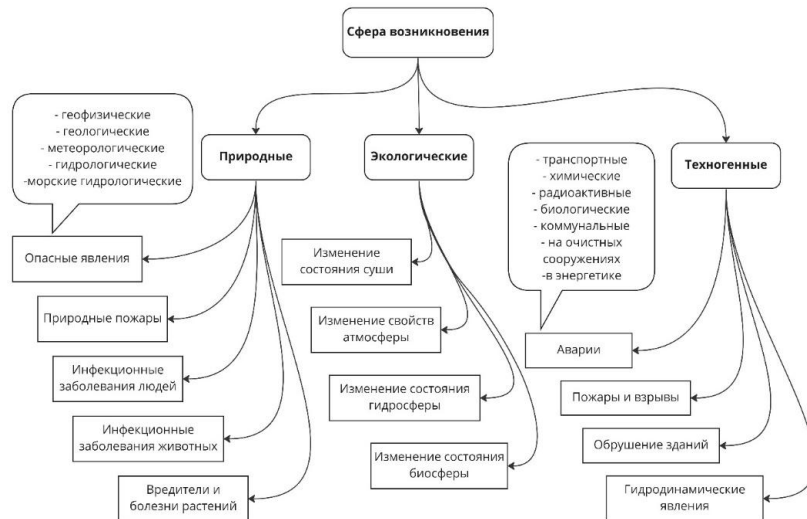


Рис. 5. Состав базовых подклассов ЧС по классу сфер возникновения

Отметим, что территория нашей страны имеет широкий спектр типов ЧС природного характера, которые могут быть вызваны различными опасными природными явлениями и процессами (землетрясения, ураганы, бури, смерчи, метели, выюги, оползни, сели, обвалы, снежные лавины, пожары, наводнения и т.д.).

3. Разработка нечеткого метода структурирования информации при поддержке принятия решений по предупреждению и ликвидации последствий чрезвычайных ситуаций. Комбинации низкоуровневых правил «если-то», построенные для конкретных прецедентов возникновения и развития ЧС, будут создавать уникальные наборы альтернатив поддержки принятия решений по устранению последствий возникновения той или иной чрезвычайной ситуации [6-11]. Каждая такая альтернатива является по своей сути траекторией, отражающей последовательность действия с множеством уточняющих вариаций, построенных на основе применения нечетких методов структурирования оперативной информации.

В большинстве случаев *оперативная информация о ЧС будет поступать в центры мониторинга в виде текстовых корпусов данных*. В подобной ситуации **нечеткий метод структурирования информации** будет состоять из последовательности операций семантической фильтрации [11–17]. Основным преимуществом данного нечеткого метода является его универсальность. Опишем разработку предложенного нечеткого метода структурирования информации.

Пусть имеется множество из n альтернатив поддержки принятия решений по устранению последствий возникновения ЧС. При этом, $T = \{t_1, t_2, t_3, \dots, t_n\}$ – множество значений предпочтительности построенных альтернатив по одному из критериев Q_i , $i = 1, \dots, k$, которые задаются на нечетком множестве с применением функции принадлежности μ_{Q_i} , определяющей степень соответствия альтернативы t_j , $j = 1, \dots, n$ событию, оцениваемому критерием Q_i . По итогам анализа выбирается

альтернатива, имеющая наибольший суммарный рейтинг соответствия выбранным критериям. Нечеткое множество степени соответствия альтернатив t_n критерию Q_i представлено выражением (2):

$$Q_i = \left\{ \mu_{Q_i}(t_1)/t_1, \mu_{Q_i}(t_2)/t_2, \dots, \mu_{Q_i}(t_n)/t_n \right\}. \quad (2)$$

Правило для выбора соответствующей ситуации альтернативы поддержки принятия решений представлено в выражении (3) в виде пересечения описанных выше нечетких множеств:

$$S = Q_1 \cap Q_2 \cap \dots \cap Q_k. \quad (3)$$

Операция пересечения нечетких множеств представлена как минимум функций принадлежности в выражении (4):

$$\mu_S(t_j) = \min \mu_{Q_i}(t_j), i = 1, \dots, k; j = 1, \dots, n. \quad (4)$$

Выбранной будет считаться альтернатива t^* с наибольшим значением функции принадлежности, как это показано в выражении (5):

$$\mu_S(t^*) = \max \mu_S(t_j), j = 1, \dots, n. \quad (5)$$

В предлагаемом исследовании определение значений функции принадлежности основано на получении интегральных оценок. В данном случае, в основе решения задачи структурирования информации лежит построение иерархии классов с последующей идентификацией системно значимых признаков происходящих событий и распределением последних по группам и подгруппам.

Опишем оптимизационную задачу [6, 7], в которой каждый из выбранных критериев получает квазиоптимальное значение. Авторы предлагают учитывать следующий набор критериев для решения данной задачи. Первый критерий контекстной привязки на каждой итерации отвечает за выбор элемента информации с наибольшим значением принадлежности к определенному классу чрезвычайных ситуаций [18–23]. Например, имеется P взаимосвязанных элементов, предварительно распределенных на S предметных областей, в результате необходимо упорядочить элементы в каждой предметной области для C различных траекторий поддержки принятия решений с достижением максимума значения критерия контекстной привязки. В этом случае описание критерия контекстной привязки Q_1 примет вид выражения (6):

$$Q_1 = \sum_{i=1}^P \sum_{k=1}^S \varepsilon_{ij} \delta_{ik} \rightarrow \max, \quad (6)$$

где ε_{ij} – значение коэффициента контекстной привязки элемента информации i в траектории поддержки принятия решений j , ($j = 1, \dots, C$), δ_{ik} – бинарный параметр, принимающий значение 1, если элемент i входит в предметную область k , и значение 0 – в противном случае.

Во втором критерии учитывается возможность наличия междисциплинарности элементов информации в исследуемых альтернативах траекторий поддержки принятия решений с учетом наличия множества связей между предметными областями, что позволяет идентифицировать системность в исследуемых информационных процессах. Второй критерий Q_2 определяется на основе выражения (7):

$$Q_2 = \sum_{i=1}^P \sum_{y=1}^S \sum_{k=1}^S \gamma_{iy} \delta_{ik} \rightarrow \max, y \neq k, \quad (7)$$

где γ_{iy} – матрица коэффициентов междисциплинарных связей, задающих меру отношения элемента информации i к предметной области y . Данные коэффициенты задаются на основе экспертных оценок и принимают значения в диапазоне $[0;1]$.

Третий критерий Q_3 определяет динамику снижения уровня оперативности информации о протекании ЧС с учетом избыточности при получении сведений из множества различных источников, т.е. необходим для оценки ценности полученной информации и задан выражением (8):

$$Q_3 = \sum_{k=1}^S \sum_{i=1}^P \varepsilon_{ij} \delta_{ik} \varphi_{ij} \rightarrow \max, \quad (8)$$

где φ_{ij} – коэффициент, предполагающий рассмотрение лишь наиболее значимых с учетом оперативности элементов информации, принимающий значение 1, если $\varepsilon_{ij} \geq 0,8$, и значение 0 – в противном случае.

На основе выражений (6)–(8) задан обобщенный критерий оптимизации, предполагающий построение альтернативных траекторий поддержки принятия решений по устранению последствий возникновения ЧС. Таким образом, обобщенный критерий примет вид выражения (9):

$$Q_{int} = \tau_1 Q_1 + \tau_2 Q_2 + \tau_3 Q_3 \rightarrow \max, \quad (9)$$

где τ_i – вес каждого из предложенных критериев, определенный на основе экспертных оценок.

Предложенный нечеткий метод структурирования информации при ЧС, отличается от известных применением *нового обобщенного критерия оптимизации выбора альтернатив поддержки принятия решений*. Оригинальность оптимизационной постановки задачи структурирования заключается в оценке контекстной привязки элементов информации к определенному классу чрезвычайных ситуаций, междисциплинарности с учетом наличия множества связей между предметными областями, а также учета снижения уровня оперативности информации о протекании ЧС с течением времени. В основе разработки данного метода использованы нечеткие модели структурирования знаний, предложенные авторами в работе [6]. Описанная в данном источнике интегрированная нечеткая модель структурирования знаний является универсальной и отличается от аналогичных введением оригинальной иерархии классов для семантической фильтрации элементов информации. Модель позволяет реализовать настройку весовых коэффициентов при моделировании альтернативных траекторий поддержки принятия решений.

4. Результаты вычислительного эксперимента. Предложенный нечеткий метод структурирования информации при ЧС протестирован на обработке семантической сети, используемой в качестве информационной модели. Семантическая сеть построена в специально созданном программном приложении на основе декомпозиции информации о различных вариантах возникновения предпосылок чрезвычайных ситуаций, а также действий по ликвидации их возможных последствий.

Исследователь задает произвольное количество уровней декомпозиции, оптимальным для подобных информационных моделей считается количество структурных единиц в одном уровне равное шести. Такая организация сети позволяет выделить уровни детализации анализируемой информации. Каждый узел сети имеет привязку к определенному элементу информации (рис. 6).

Первому уровню принадлежит только один понятийный узел сети. Он является корнем сети и обобщает все ее содержимое, являясь мета-информацией. Каждый новый уровень раскрывает содержание понятий предыдущего уровня, тем самым раскрывая для исследователя всю предметную область и обнаруживая наличие междисциплинарных связей с другими узлами сети.

Сеть включает в себя множество предметных областей $F = \{F_1, F_2, F_3 \dots F_z\}$, где $z = \overline{1, m}$, m – integer constant (количество предметных областей), причем, $F_z = \{O_{z1}, O_{z2}, O_{z3}, \dots, O_{zk_z}\}$, где k_z – integer constant (количество элементов (объектов) знаний в F_z), причем, в случае задания отношений между объектами одной предметной области получим, что $\forall i, j \rightarrow i = \overline{1, k_z}; j = \overline{1, k_z}; i \neq j \exists C_z[i][j] = n$,

где n – количество связей между O_{zi} и O_{zj} , z – integer constant (номер исследуемой предметной области). В случае задания всех, в т.ч. междисциплинарных, отношений между объектами предметных областей получим, что $\forall a, b \rightarrow a = \overline{1, \sum_{z=1}^m k_z}$; $b = \overline{1, \sum_{z=1}^m k_z}$; $a \neq b \exists C_{interdisc}[a][b] = p$, где p – количество всех, в т.ч. междисциплинарных, связей между объектами сети.

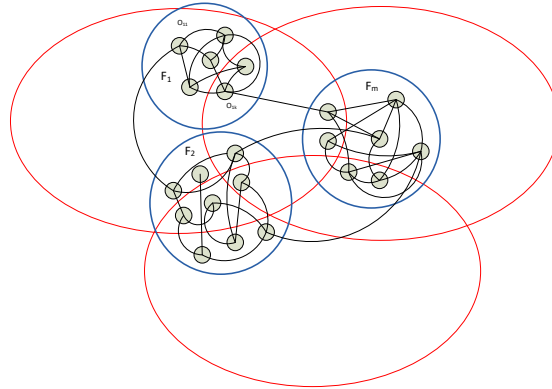


Рис. 6. Абстрактная модель семантической сети с несколькими уровнями декомпозиции

В результате проведенных исследований получена временная сложность предложенного модифицированного алгоритма (рис. 7).



Рис. 7. Зависимость времени работы метода от числа вершин семантической сети

Временная сложность алгоритма в представленном примере составила $O(n^2)$, где n – количество анализируемых алгоритмом входных данных. По мнению авторов в большинстве реальных случаев не понадобится сеть такой большой размерности. Более вероятным диапазоном используемого количества вершин в семантической сети будет $n \in [5000: 50000]$, в этом случае ВСА предложенного нечеткого метода структурирования информации при поддержке принятия решений по предупреждению и ликвидации последствий чрезвычайных ситуаций составит $O(n)$.

Заключение. В данной статье представлена разработка нечеткого метода структурирования информации. Основной целью работы являлось повышение эффективности средств поддержки принятия решений по предупреждению и ликвидации последствий чрезвычайных ситуаций на основе методов нечеткой логики.

Основными результатами проведенного исследования являются следующие:

1. Даны определения основным терминам исследуемой предметной области. Представлена формализованная постановка решаемой задачи;
2. Приведена развернутая классификация чрезвычайных ситуаций с описанием особенностей представленных классов;
3. Разработан нечеткий метод структурирования информации, состоящий из последовательности операций семантической фильтрации. Основным преимуществом данного нечеткого метода является его универсальность. Предложенный нечеткий метод структурирования информации при ЧС, отличается от известных применением нового обобщенного критерия оптимизации выбора альтернатив поддержки принятия решений и позволяет проводить оценку контекстной привязки элементов информации к определенному классу чрезвычайных ситуаций, междисциплинарности с учетом наличия множества связей между предметными областями, а также учета снижения уровня оперативности информации о протекании ЧС с течением времени.

Для оценки эффективности предложенного нечеткого метода разработано программное приложение и проведен вычислительный эксперимент. Полученные результаты проведенных экспериментальных исследований подтверждают эффективность предложенного нечеткого метода структурирования информации при поддержке принятия решений по предупреждению и ликвидации последствий чрезвычайных ситуаций. Временная сложность представленного алгоритма является полиномиальной. При выполнении исследования решены все поставленные задачи, цель данной работы достигнута.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Государственный доклад о состоянии защиты населения и территорий Российской Федерации от чрезвычайных ситуаций природного и техногенного характера в 2014 году. – М.: МЧС России, 2015. – 318 с.
2. Современные системы мониторинга и прогнозирования чрезвычайных ситуаций / под общ. ред. В.А. Пучкова. – М.: ФКУ ЦСИ ГЗ МЧС России, 2013. – 352 с.
3. *Болов В.Р.* Применение современных технологий, методов мониторинга и прогнозирования в обеспечении системы управления в кризисных ситуациях // Журнал-каталог «Средства спасения. Противопожарная защита. Российские инновационные системы». – 2010. – № 10.
4. *Исаев, В.С., Макиев, Ю.Д., Малышев, В.П., Таранов, А.А., Камзолкин, В.Л.* Методика оценки эффективности мероприятий по повышению устойчивости функционирования критически важных объектов и объектов жизнеобеспечения в условиях угроз террористического характера // Информационный сборник. – М.: ЦСИ ГЗ МЧС России, 2010. – № 42. – С. 52-68.
5. *Горбунов, С.В., Макиев, Ю.Д., Малышев, В.П.* Анализ технологий прогнозирования чрезвычайных ситуаций природного и техногенного характера // Стратегия гражданской безопасности, проблемы и решения: Науч.-аналит. сб. – М.: 2011. – Т. 1. – № 1 (1). – С. 43-53.
6. *Кравченко Ю.А.* Модель структуризации знаний на основе оценки наличия существенных признаков // Тр. Конгресса по интеллектуальным системам и информационным технологиям «IS&IT'15». Научное издание в 3-х т. Т. 1. – 2015. – С. 293-298.
7. *Варшавский П.Р., А.П. Еремеев, Куриленко И.Е.* Моделирование временных зависимостей в интеллектуальных системах поддержки принятия решений на основе прецедентов // Information technologies and knowledge. – 2012. – Vol. 6, No. 3. – P. 227-239.
8. *Бова, В.В., Курейчик В.В., Лецанов Д.В.* Модель семантического поиска в системах управления знаниями на основе генетических процедур // Информационные технологии. – 2017. – Т. 23, № 12. – С. 876-883.
9. *Chen C., Koufaris M.* The Impact of Decision Support System Features on User Overconfidence and Risky Behavior // Nature Publishing Group. – 2014. – No. 4. – P. 1-17.
10. *Ларичев О.И.* Вербальный анализ решений. – М.: Наука, 2006. – 181 с.
11. *Капунов Т.Г., Курейчик В.М.* Адаптивный генетический алгоритм на основе нечетких правил // Известия ЮФУ. Технические науки. – 2018. – № 5 (199). – С. 26-34.

12. Arenas M., Botoeva E., Calvanese D., Ryzhikov V. Exchanging OWL 2 QL Knowledge Bases // 23rd International Joint Conference on Artificial Intelligence. – 2013. – P. 703-710.
13. Федоров Д.Ю. Применение структуризации знаний для обеспечения информационной безопасности личности // Национальная безопасность и стратегическое планирование. – 2013. – № 2. – С. 40-43.
14. Кравченко Ю.А. Технологии управления знаниями как основа поддержки принятия решений // Тр. Конгресса по интеллектуальным системам и информационным технологиям «IS&IT'14». Научное издание в 4-х т. Т. 1. – 2014. – С. 151-157.
15. Hecker A. Knowledge Beyond the Individual. Making Sense of a Notion of Collective Knowledge in Organization Theory // Organization Studies. – 2012. – Vol. 33. – P. 423-445.
16. Bova V.V., Kureichik V.V., Nuzhnov E.V. The Combined Method of Semantic Similarity Estimation of Problem Oriented Knowledge on the Basis of Evolutionary Procedures // Artificial Intelligence Trends in Intelligent Systems. – 2015. – P. 74-83.
17. Rodzin S., Rodzina L. Theory of Bioinspired Search for Optimal Solutions and its Application for the Processing of Problem-Oriented Knowledge // 8th IEEE International Conference on Application of Information and Communication Technologies. – 2014. – P. 142-147.
18. Dolinina O., Suchkova N. Formal Models of the Structural Errors in the Knowledge Bases of Intellectual Decision Making Systems // Artificial Intelligence Trends in Intelligent Systems. – 2015. – P. 156-167.
19. Jensen J. A Systematic Literature Review of the Use of Semantic Web Technologies in Formal Education // Br. J. Edu. Technol. – 2019. – Vol. 50. – P. 505-517.
20. Гулякина Н.А., Давыденко И.Т. Семантические модели и метод согласованной разработки баз знаний // Программные продукты и системы. – 2020. – № 3. – С. 420-429.
21. Шалимов П.Ю. Модели количественной шкалы оценки семантической информации // Автоматизация и моделирование в проектировании и управлении. – 2020. – № 2 (8). – С. 24-32.
22. Грегер С.Э., Поршнев С.В. Метод объектно-ориентированного представления многоуровневых семантических моделей // Cloud of Science. – 2016. – Т. 3, № 4. – С. 548-561.
23. Шевченко Е.Л. Метрика близости семантических аннотаций для задачи сравнения профилей процессов // Восточно-Европейский журнал передовых технологий. – 2013. – Т. 3, № 2 (63). – С. 48-52.

REFERENCES

1. Gosudarstvennyy доклад o sostoyanii zashchity naseleniya i territoriy Rossiyskoy Federatsii ot chrezvychaynykh situatsiy prirodnogo i tekhnogennogo kharaktera v 2014 godu [State report on the state of protection of the population and territories of the Russian Federation from natural and man-made emergencies in 2014]. Moscow: MChS Rossii, 2015, 318 p.
2. Sovremennyye sistemy monitoringa i prognozirovaniya chrezvychaynykh situatsiy [Modern systems of monitoring and forecasting of emergency situations], under the general ed. of V.A. Puchkova. Moscow: FKU TsSI GZ MChS Rossii, 2013, 352 p.
3. Bolov V.R. Primenenie sovremennykh tekhnologiy, metodov monitoringa i prognozirovaniya v obespechenii sistemy upravleniya v krizisnykh situatsiyakh [Application of modern technologies, methods of monitoring and forecasting in providing a management system in crisis situations], Zhurnal-katalog «Sredstva spaseniya. Protivopozharnaya zashchita. Rossiyskie innovatsionnye sistemy» [Journal-catalog "Means of rescue. Fire protection. Russian Innovation Systems"], 2010, No. 10.
4. Isaev, V.S., Makiev, Yu.D., Malyshev, V.P., Taranov, A.A., Kamzolkin, V.L. Metodika otsenki effektivnosti meropriyatiy po povysheniyu ustoychivosti funktsionirovaniya kriticheski vazhnykh ob"ektov i ob"ektov zhizneobespecheniya v usloviyakh ugroz terroristicheskogo kharaktera [Methodology for assessing the effectiveness of measures to improve the stability of the functioning of critical facilities and life support facilities in the face of terrorist threats], Informatsionnyy sbornik [Information collection]. Moscow: TsSI GZ MChS Rossii, 2010, No. 42, pp. 52-68.
5. Gorbunov, S.V., Makiev, Yu.D., Malyshev, V.P. Analiz tekhnologiy prognozirovaniya chrezvychaynykh situatsiy prirodnogo i tekhnogennogo kharaktera [Analysis of technologies for forecasting natural and man-made emergencies], Strategiya grazhdanskoy bezopasnosti, problemy i resheniya: Nauch.-analit. sb. [Strategy of civil security, problems and solutions: Scientific and analytical collection]. Moscow: 2011, Vol. 1, No. 1 (1), pp. 43-53.

6. Kravchenko Yu.A. Model' strukturizatsii znaniy na osnove otsenki nalichiya sushchestvennykh priznakov [A model of knowledge structuring based on the assessment of the presence of essential features], *Tr. Kongressa po intellektual'nym sistemam i informatsionnym tekhnologiyam «IS&IT'15»* [Proceedings of the Congress on Intelligent Systems and Information Technologies "IS&IT'15"]. Scientific publication in 3 vol. Vol. 1, 2015, pp. 293-298.
7. Varshavskiy P.R., A.P. Ereemeev, Kurilenko I.E. Modelirovanie vremennykh zavisimostey v intellektual'nykh sistemakh podderzhki prinyatiya resheniy na osnove pretsedentov [Modeling of time dependencies in intelligent decision support systems based on precedents], *Information technologies and knowledge* [information technologies and knowledge], 2012, Vol. 6, No. 3, pp. 227-239.
8. Bova, V.V., Kureychik V.V., Leshchanov D.V. Model' semanticheskogo poiska v sistemakh upravleniya znaniyami na osnove geneticheskikh protsedur [Semantic search model in knowledge management systems based on genetic procedures], *Informatsionnye tekhnologii* [Information Technologies], 2017, Vol. 23, No. 12, pp. 876-883.
9. Chen C., Koufaris M. The Impact of Decision Support System Features on User Overconfidence and Risky Behavior, *Nature Publishing Group*, 2014, No. 4, pp. 1-17.
10. Larichev O.I. Verbal'nyy analiz resheniy [Verbal analysis of decisions]. Moscow: Nauka, 2006, 181 p.
11. Kaplunov T.G., Kureychik V.M. Adaptivnyy geneticheskiy algoritm na osnove nechetkikh pravil [Adaptive genetic algorithm based on fuzzy rules], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2018, No. 5 (199), pp. 26-34.
12. Arenas M., Botoeva E., Calvanese D., Ryzhikov V. Exchanging OWL 2 QL Knowledge Bases, *23rd International Joint Conference on Artificial Intelligence*, 2013, pp. 703-710.
13. Fedorov D.Yu. Primenenie strukturizatsii znaniy dlya obespecheniya informatsionnoy bezopasnosti lichnosti [Application of knowledge structuring to ensure the information security of the individual], *Natsional'naya bezopasnost' i strategicheskoe planirovanie* [National security and strategic planning], 2013, No. 2, pp. 40-43.
14. Kravchenko Yu.A. Tekhnologii upravleniya znaniyami kak osnova podderzhki prinyatiya resheniy [Технологии управления знаниями как основа поддержки принятия решений], *Tr. Kongressa po intellektual'nym sistemam i informatsionnym tekhnologiyam «IS&IT'14»* [Proceedings of the Congress on Intelligent Systems and Information Technologies "IS&IT'14"]. Scientific publication in 4 vol. Vol. 1, 2014, pp. 151-157.
15. Hecker A. Knowledge Beyond the Individual. Making Sense of a Notion of Collective Knowledge in Organization Theory, *Organization Studies*, 2012, Vol. 33, pp. 423-445.
16. Bova V.V., Kureichik V.V., Nuzhnov E.V. The Combined Method of Semantic Similarity Estimation of Problem Oriented Knowledge on the Basis of Evolutionary Procedures, *Artificial Intelligence Trends in Intelligent Systems*, 2015, pp. 74-83.
17. Rodzin S., Rodzina L. Theory of Bioinspired Search for Optimal Solutions and its Application for the Processing of Problem-Oriented Knowledge, *8th IEEE International Conference on Application of Information and Communication Technologies*, 2014, pp. 142-147.
18. Dolinina O., Suchkova N. Formal Models of the Structural Errors in the Knowledge Bases of Intellectual Decision Making Systems, *Artificial Intelligence Trends in Intelligent Systems*, 2015, pp. 156-167.
19. Jensen J. A Systematic Literature Review of the Use of Semantic Web Technologies in Formal Education, *Br. J. Edu. Technol.*, 2019, Vol. 50, pp. 505-517.
20. Gulyakina N.A., Davydenko I.T. Semanticheskie modeli i metod soglasovannoy razrabotki baz znaniy [Semantic models and the method of coordinated development of knowledge bases], *Programmnye produkty i sistemy* [Software products and systems], 2020, No. 3, pp. 420-429.
21. Shalimov P.Yu. Modeli kolichestvennoy shkaly otsenki semanticheskoy informatsii [Models of the quantitative scale of semantic information evaluation], *Avtomatizatsiya i modelirovanie v proektirovanii i upravlenii* [Automation and modeling in design and management], 2020, No. 2 (8), pp. 24-32.
22. Greger S.E., Porshnev S.V. Metod ob"ektno-orientirovannogo predstavleniya mnogourovnevnykh semanticheskikh modeley [The method of object-oriented representation of multilevel semantic models], *Cloud of Science*, 2016, Vol. 3, No. 4, pp. 548-561.
23. Shevchenko E.L. Metrika blizosti semanticheskikh annotatsiy dlya zadachi sravneniya profiley protsessov [The proximity metric of semantic annotations for the task of comparing process profiles], *Vostochno-Evropeyskiy zhurnal peredovykh tekhnologiy* [Eastern European Journal of Advanced Technologies], 2013, Vol. 3, No. 2 (63), pp. 48-52.

Статью рекомендовал к опубликованию д.т.н., профессор Л.С. Лисицына.

Герасименко Евгения Михайловна – Южный федеральный университет; e-mail: egerasimenko@sfedu.ru; г. Таганрог, Россия; тел.: 88634371651; кафедра систем автоматизированного проектирования; доцент.

Кравченко Даниил Юрьевич – e-mail: dkravchenko@sfedu.ru; тел.: 88634371651; кафедра систем автоматизированного проектирования; студент.

Кравченко Юрий Алексеевич – e-mail: yakravchenko@sfedu.ru; тел.: 88634371651; кафедра систем автоматизированного проектирования, д.т.н.; доцент.

Кулиев Эльмар Валерьевич – e-mail: ekuliev@sfedu.ru; тел.: 88634371651; кафедра систем автоматизированного проектирования; доцент.

Gerasimenko Evgeniya Mihailovna – Southern Federal University; e-mail: egerasimenko@sfedu.ru; Taganrog, Russia; phone: +78634371651; the department of computer aided design; associate professor.

Kravchenko Daniil Yurievich – e-mail: kravchenkodanil122@gmail.com; phone: +78634371651; the department of computer aided design; student.

Kravchenko Yury Alekseevich – e-mail: yakravchenko@sfedu.ru; phone: +78634371651; the department of computer aided design; dr. of eng. sc.; associate professor.

Kuliev Elmar Valerievich – e-mail: ekuliev@sfedu.ru; phone: +78634371651; the department of computer aided design, associate professor.

УДК 004.9

DOI 10.18522/2311-3103-2023-2-212-226

Ф.С. Булыга, В.М. Курейчик

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ВЕКТОРИЗАЦИИ ТЕКСТОВЫХ ДАННЫХ БОЛЬШОЙ РАЗМЕРНОСТИ

Представленная публикация посвящена обзору проблемы представления текстовой информации для последующего осуществления кластерного анализа в рамках обработки и управления информацией большой размерности. Современные требования предъявляемые к аналитическим, поисковым и рекомендательным информационным системам демонстрируют слабую сформированность целостного решения, способного обеспечить достаточный уровень быстродействия и качества получаемых результатов в рамках функционирования текущего рынка информационных технологий. Поиск решения представленной проблемы влечет за собой необходимость в проведении объективного анализа существующих решений представления текстовой информации в векторном пространстве, с целью формирования целостного представления о достоинствах и недостатках анализируемых подходов, а также формированием критериев, позволяющих реализовать собственный подход, лишенный выявленных слабостей. Представленная работа является аналитической, и позволяет получить представление о современном состоянии и проработанности выявленной проблемы в рамках ограниченной предметной области. Кластеризация текстовых данных – автоматическое формирование подмножеств, элементами которых выступают экземпляры документов некоторой исследуемой, неструктурированной выборки фиксированной размерности. Данный процесс можно классифицировать как обучения без учителя, предполагающее, отсутствие эксперта, собственноручно присваивающего исходной выборке документов индексы классов. Однако, осуществление кластерного анализа текстовых данных без какой-либо предварительной обработки – невозможно. Для этого необходимо обеспечить стандартизацию и приведение входных данных к единому формату и виду. В рамках данного этапа осуществления кластерного анализа, в представленной публикации рассматриваются методы предварительной обработки текстовых данных. Новизна представленной публикации заключается в формировании теоретического базиса основных методов векторизации текстовых данных, путем систематизации и объектив-