

А. В. МАТВЕЕВ, канд. техн. наук, доцент кафедры прикладной математики и информационных технологий, Санкт-Петербургский университет ГПС МЧС России (Россия, 196105, г. Санкт-Петербург, Московский просп. 149; e-mail: fcvega_10@mail.ru)

А. В. МАКСИМОВ, канд. техн. наук, преподаватель кафедры прикладной математики и информационных технологий, Санкт-Петербургский университет ГПС МЧС России (Россия, 196105, г. Санкт-Петербург, Московский просп. 149; e-mail: he1nze@mail.ru)

О. В. ЩЕРБАКОВ, д-р техн. наук, профессор, заслуженный деятель наук РФ, профессор кафедры прикладной математики и информационных технологий, Санкт-Петербургский университет ГПС МЧС России (Россия, 196105, г. Санкт-Петербург, Московский просп., 149)

А. С. СМИРНОВ, д-р техн. наук, профессор, Министерство Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий (Россия, 121352, г. Москва, Давыдовская ул., 7)

УДК 621.642.88

МЕТОД ОЦЕНКИ ДОСТОВЕРНОСТИ КОЛИЧЕСТВЕННОГО АНАЛИЗА РИСКА НА ОБЪЕКТАХ НЕФТЕГАЗОВОЙ ОТРАСЛИ

Исследованы существующие практические методы оценки достоверности количественного анализа рисков на объектах нефтегазовой отрасли. Предложено использовать метод, основанный на обеспечении качества самого процесса анализа риска. Выявлены факторы, обеспечивающие качество процесса анализа риска на объектах нефтегазовой отрасли. Представлен метод оценки достоверности количественного анализа риска, основанный на применении наивного байесовского классификатора. Предложен подход к оценке качества работы самого классификатора, основанный на перекрестной проверке с последовательным исключением одного экземпляра данных обучения.

Ключевые слова: количественный анализ риска; достоверность; надежность; наивный байесовский классификатор; данные обучения; риск.

DOI: 10.18322/PVB.2018.27.01.35-49

Введение

Количественный анализ рисков для поддержки принятия решений, связанных с безопасностью [1–3], широко применяется в различных областях, например в химической промышленности [4], атомной промышленности [5, 6], на магистральном трубопроводном транспорте [7] и т. п. Особый интерес представляют вопросы анализа и оценки рисков на особо опасных объектах нефтегазовой промышленности как на стадии проектирования, так и в процессе эксплуатации [8–10]. И в российской, и в международной практике разработаны различные методы для количественного анализа рисков [11, 12].

В то же время существуют проблемы, связанные с оценкой достоверности результатов, полученных при количественном анализе рисков. Остается открытым вопрос, в какой степени лицо, принимающее решение (ЛПР), может доверять полученным результатам. Этот вопрос имеет первостепенное значение в практике проектирования и эксплуатации объектов нефтегазовой отрасли. Только достоверные и надежные результаты, полученные в процессе коли-

чественного анализа рисков, могут быть полезны для поддержки принятия решений.

Данная совокупность факторов определила актуальность проводимых исследований, а цель настоящей работы состоит в повышении объективности при оценке достоверности результатов количественного анализа риска на объектах нефтегазовой отрасли. Достижение данной цели предполагает решение задачи анализа существующих подходов к оценке, выявления их достоинств и недостатков, а также непосредственно разработки формального и количественного метода оценки достоверности результатов количественного анализа риска.

Анализ подходов к оценке достоверности результатов количественного анализа риска

Вопросы оценки достоверности результатов количественного анализа рисков были рассмотрены в работах [9, 13, 14]. Исследования показали, что существующие методы оценки достоверности коли-

чественного анализа рисков могут быть в целом разделены на четыре группы, в основе которых лежит:

- 1) сравнение с контрольными параметрами (сравнение с эталоном);
- 2) проверка в реальных условиях (эксперимент с реальным объектом нефтегазовой отрасли или сравнение со статистическими данными);
- 3) независимая экспертная оценка;
- 4) обеспечение качества при анализе рисков.

Методы первой группы основаны на сравнении нескольких результатов, полученных при проведении параллельных анализов рисков для определения достоверности. Как правило, при сопоставлении рассматриваются два критерия — надежность и достоверность [14]. Например, достоверность количественного анализа рисков на хранилище аммиака оценивается путем сравнения семи контрольных параметров [15].

Вторая группа методов проверки достоверности предполагает оценку достоверности результатов, полученных при количественном анализе риска, путем сравнения их с реальными данными либо со статистическими данными, касающимися работы исследуемой системы или процесса. Типичный пример представлен в [16], где статистические данные о реальных авариях и происшествиях сравниваются с индексами риска, рассчитанными при использовании конкретного метода количественного анализа риска.

В независимых методах экспертной оценки процесс количественного анализа риска на объектах нефтегазовой отрасли и его результаты рассматриваются независимыми экспертами, которые оценивают их, основываясь на ряде заранее определенных требований к качеству, и определяют таким образом надежность количественного анализа риска.

Методы обеспечения качества предполагают контроль качества в каждой фазе процесса количественного анализа риска, что в целом обеспечивает качество всего процесса анализа риска [14]. Именно применение данного подхода дает достоверные и надежные результаты оценки риска. Например, авторы [17] обобщили общие недостатки на каждом этапе процесса количественного анализа риска и разработали подход, основанный на контрольных списках, для обеспечения его качества.

Среди четырех рассмотренных групп методы первых двух основываются на использовании ретроспективных данных, т. е. оценки достоверности устанавливаются в основном путем сопоставления конечных результатов анализа с результатами уже апробированных и достоверных методов или имеющимися статистическими данными по объектам нефтегазовой отрасли. Методы, основанные на независимой экспертной оценке и обеспечении каче-

ства при анализе рисков, напротив, являются проактивными в том смысле, что вместо прямой оценки конечных результатов они оценивают структуру и потенциальные возможности самого процесса количественного анализа риска (с точки зрения его качества).

Первые две группы методов в некотором смысле предпочтительнее по причине того, что они более объективны, поскольку позволяют оценивать достоверность результатов анализа непосредственно путем сравнения с имеющимися объективными данными. Однако существует два основных недостатка, которые могут ограничить применимость данных методов. Во-первых, эти методы очень часто сложно реализовать из-за требований к определенному объему ретроспективной статистической информации или к наличию достоверных и надежных результатов при использовании других апробированных процедур количественного анализа риска. Во-вторых, эти методы при анализе рисков позволяют учитывать данные только по уже реализованным факторам, приводящим к авариям и нештатным ситуациям на объектах. В этих условиях ограничивается возможность совершенствования процесса количественного анализа риска в целях повышения достоверности и надежности путем включения в процесс анализа каких-то новых факторов риска и явлений, по которым фактически отсутствуют статистические данные [18]. В связи с этим предложено акцент в проводимых исследованиях сделать на так называемых проактивных (упреждающих) методах количественного анализа риска.

В данных методах достоверность оценивается (фактически прогнозируется) на основе возможностей (с точки зрения качества) процесса количественного анализа риска на объектах нефтегазовой отрасли. При разработке таких методов возникает два главных вопроса:

- как оценить потенциальные возможности (с точки зрения качества) процесса количественного анализа риска;
- как соотносить достоверность количественного анализа риска с его потенциальными возможностями?

В литературе довольно подробно изучен первый вопрос: основные факторы, влияющие на качество процесса количественного анализа риска, рассмотрены, например, в работе [12]. Второму вопросу в научной литературе уделено значительно меньше внимания. В большинстве существующих исследований (например, [19]) соотношение между достоверностью и качеством процесса анализа риска оценивается по принципу “черного ящика”, т. е. не рассматривается внутренняя структура процесса анализа, а проводится лишь анкетирование экс-

пертов, в результате которого осуществляется ранжирование значимости каждого отдельного элемента (задачи) процесса анализа. На основе полученных результатов дается оценка достоверности результатов анализа риска. Как правило, она формируется на основе простой процедуры, основанной на оценке согласованности мнений экспертов по каждому из критериев и сравнении количества согласованных критериев качества с определенным пороговым значением. Такой процесс обладает высокой степенью субъективизма и непрозрачен в плане формирования выводов об оценке достоверности результатов анализа [18]. Для повышения объективности при оценке достоверности необходим формальный и количественный метод, разработка которого и составляет цель исследований, представленных в настоящей статье.

Фактически оценка достоверности может рассматриваться как проблема распознавания образов, основанного на прецедентах [20]. Решение данной задачи предлагается осуществлять путем построения классификатора, являющегося отображением характеристик и потенциальных возможностей процесса количественного анализа риска на показатель достоверности его результатов и основанного на наборе данных обучения, которые предварительно оцениваются экспертами. Для решения задачи классификации предлагается использовать наивный байесовский классификатор (НБК) для формальной и количественной оценки достоверности и надежности количественного анализа риска.

НБК — простой, но достаточно эффективный классификатор, широко применяемый в приложениях машинного обучения [21], например для классификации текста [22]. Данный классификатор основан на использовании теоремы Байеса со строгими (наивными) предположениями о независимости признаков. В зависимости от точной природы вероятностной модели НБК может обучаться очень эффективно. Во многих практических приложениях для оценки параметров для наивных байесовых моделей используют метод максимального правдоподобия. Несмотря на наивный вид и, несомненно, очень упрощенные условия, НБК часто значительно эффективнее при решении многих сложных прикладных задач. Достоинством НБК является сравнительно небольшое количество данных для обучения, необходимых для оценки параметров, требуемых для классификации.

Следует отметить, что в решении поставленной задачи оценки качества результатов количественного анализа риска классификатор используется не напрямую для оценки качества, а скорее, как инструмент для построения критериев оценки, на осно-

ве которых определяются достоверность и надежность.

Теоретические основы оценки достоверности количественного анализа риска на объектах нефтегазовой отрасли

1. Структура оценки

Рассмотрим предлагаемый подход к оценке достоверности количественного анализа риска на основе классификации с использованием НБК. Пусть D — показатель достоверности количественного анализа риска. Предполагаем, что D определяется качеством процесса анализа риска. Типовой процесс количественного анализа риска можно условно разбить на восемь подпроцессов [23] (рис. 1). Качественное выполнение каждого из данных подпроцессов является необходимым условием для обеспечения качества всего процесса анализа риска в целом. Предлагаемая структура оценки достоверности каждого из подпроцессов количественного анализа риска представлена на рис. 2.

Достоверность и надежность количественного анализа риска формируются под воздействием следующих пяти базовых критериев (признаков) x_i ($i = 1, \dots, 5$):

- полнота учета всех опасных факторов x_1 ;
- полнота охвата потенциальных сценариев аварий x_2 ;
- целесообразность применения конкретных методов анализа x_3 ;
- достоверность и точность совокупности исходных данных x_4 ;
- точность используемых методов при расчете показателей риска x_5 .

Данные критерии отражают совокупность требований к достоверности и надежности в процессе количественного анализа риска. Значение каждого критерия количественно оценивается дискретной величиной, принимаемой из трех возможных значений, каждое из которых соответствует значению лингвистической переменной: неудовлетворительно ($x_i = 0$), приемлемо ($x_i = 1$) и удовлетворительно ($x_i = 2$) — на основе множества заранее предопределенных правил преобразования исходной экспертной информации в значения лингвистической переменной (табл. 1). Байесовский подход основывается на предположении о существовании некоторого распределения вероятностей для каждого критерия.

Показатель достоверности D также было предложено разделить на три дискретных уровня: $D \in \{0, 1, 2\}$. Качественное описание данных уровней представлено в табл. 2. Таким образом, проблема оценки достоверности была сформулирована как проблема классификации: учитывая значения пяти

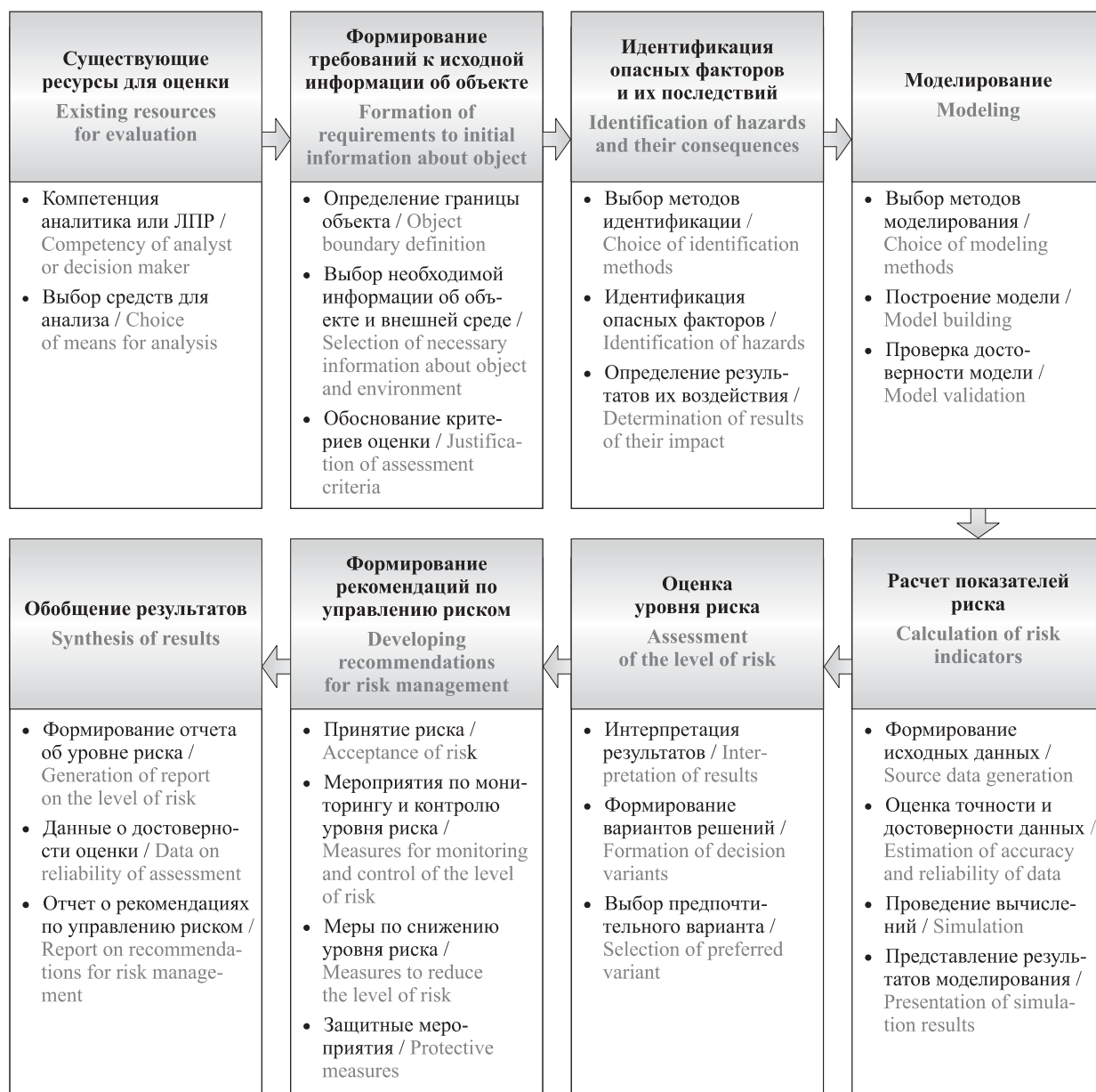


Рис. 1. Типовая схема количественного анализа риска

Fig. 1. Typical scheme for quantitative risk analysis

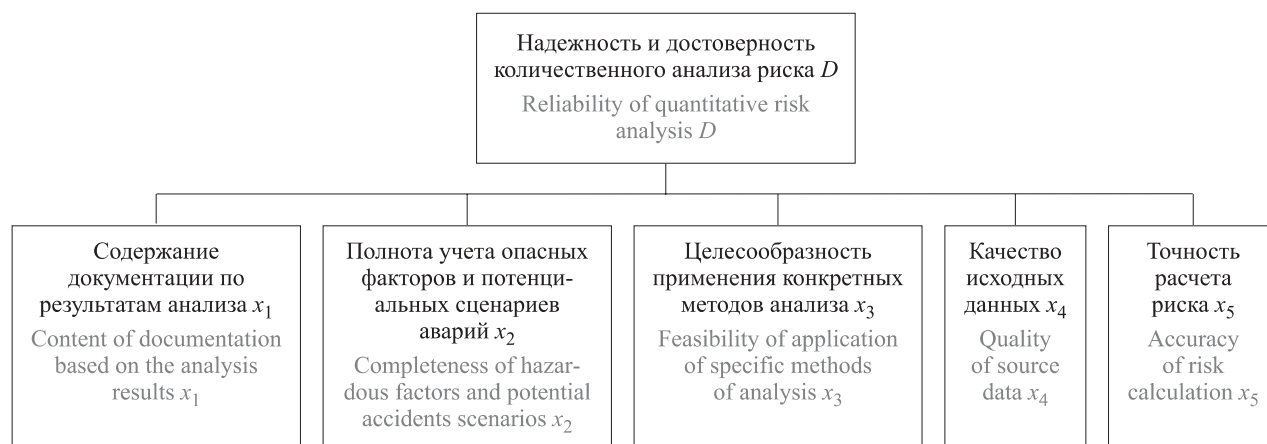


Рис. 2. Структура оценки надежности и достоверности количественного анализа риска

Fig. 2. Structure of reliability assessment and quantitative risk analysis

Таблица 1. Правила кодирования показателей базовых критериев

Table 1. Rules for coding indicators of basic criteria

Значение / Value	Описание / Description
Правила кодирования показателя x_1 / Rules for coding indicator x_1	
$x_1 = 0$	В документации по анализу риска отсутствуют следующие элементы: - цели и задачи количественного анализа риска; - полное описание исследуемого объекта; - обоснование используемых методов анализа; - исходные данные, необходимые для анализа и расчетов; - отчет о результатах анализа The documentation for risk analysis does not contain the following elements: - goals and objectives of quantitative risk analysis; - complete description of the object under study; - justification of the methods of analysis used; - initial data necessary for analysis and calculation; - report on the results of the analysis
$x_1 = 1$	В документации по анализу риска присутствует, по крайней мере, один из следующих недостатков: - описания задач являются неполными или двусмысленными; - описание исследуемого объекта не в полной мере понятно; - обоснование применяемых методов анализа расплывчато и неочевидно; - представленные результаты являются неполными (например, не учитываются факторы неопределенности) или неоднозначными The documentation for risk analysis has at least one of the following drawbacks: - descriptions of tasks are incomplete or ambiguous; - description of the investigated object is not fully understandable; - rationale for the methods of analysis used is vague and not obvious; - reported results are incomplete (for example, uncertainties are not taken into account) or ambiguous
$x_1 = 2$	Отчетная документация по анализу риска содержит полную и достаточную информацию, в том числе отсутствуют недостатки, перечисленные для уровня $x_1 = 1$ Reporting documentation on risk analysis contains complete and sufficient information, including no shortcomings listed for level $x_1 = 1$
Правила кодирования показателя x_2 / Rules for coding indicator x_2	
$x_2 = 0$	Не идентифицированы некоторые опасные факторы, а также сценарии аварий и происшествий: - проверка показала, что некоторые поражающие факторы не учитываются; - отсутствует обоснование рассмотренных аварийных и нештатных ситуаций; - проверка показала, что некоторые критические сценарии аварийных ситуаций отсутствуют Some dangerous factors have not been identified, as well as scenarios of accidents and incidents: - verification showed that some damaging factors are not taken into account; - there is no justification for the considered emergency and abnormal situations; - verification showed that some critical emergency scenarios are missing
$x_2 = 1$	Проверка показала, что все опасные поражающие факторы и критические сценарии аварий и происшествий идентифицированы, но проверка проводится на основе экспертной оценки, а не использования реальных статистических данных The audit showed that all dangerous damaging factors and critical scenarios of accidents and incidents were identified, but the verification is carried out on the basis of expert judgment, and not the use of real statistics
$x_2 = 2$	Проверка показала, что все критические сценарии аварий и происшествий идентифицированы и их достоверность подтверждается использованием реальных статистических данных The audit showed that all critical scenarios of accidents and incidents identified and their validity is confirmed by the use of real statistics
Правила кодирования показателя x_3 / Rules for coding indicator x_3	
$x_3 = 0$	Выбранные методы анализа не удовлетворяют требованиям проводимого исследования: - используемые методы не учитывают некоторых особенностей реального процесса, протекающего на объекте исследования (например, не учитывают динамику изменения во времени определенных характеристик, рассматривая их стационарными, и т. д.); - методы требуют больше ресурсов (например, количества исходных данных, вычислительной мощности и т. д.), чем может быть предоставлено

Значение / Value	Описание / Description
	The chosen methods of analysis do not meet the requirements of the research: • methods used do not take into account some features of the actual process occurring at the research object (for example, do not take into account the dynamics of the change in time of certain characteristics, considering them stationary, etc.); • methods require more resources (for example, amount of raw data, processing power, etc.) than can be provided
$x_3 = 1$	Особенности выбранного метода анализа удовлетворяют требованиям поставленной проблемы, но вывод о его использовании составлен на основе экспертного опыта The features of the chosen method of analysis satisfy the requirements of the problem posed, but conclusion about its use is compiled on the basis of expert experience
$x_3 = 2$	Выбранные методы анализа удовлетворяют требованиям проводимого исследования и уже были успешно апробированы при анализе риска на подобных объектах The selected methods of analysis satisfy the requirements of the conducted research and have already been successfully tested in the risk analysis of similar objects
Правила кодирования показателя x_4 / Rules for coding indicator x_4	
$x_4 = 0$	• Отсутствует достаточная база статистических данных (недостаток репрезентативности данных), и исходные данные определены исключительно на основе экспертных оценок; • в вводимых экспертами исходных данных не учитывается их эпистемическая неопределенность [24, 25]; • есть существенные погрешности в измерении исходных данных • There is no sufficient statistical database (lack of data representativeness), and the baseline data are determined exclusively on the basis of expert assessments; • the initial data input by experts does not take into account their epistemic uncertainty [24, 25]; • there are significant errors in the measurement of the original data
$x_4 = 1$	• Отсутствует достаточная база статистических данных (недостаток репрезентативности данных); • исходные данные основаны на экспертных оценках с полным учетом эпистемической неопределенности; • есть незначительные погрешности в измерении исходных данных • There is no sufficient statistical database (lack of data representativeness); • baseline data is based on expert estimates with full consideration of epistemic uncertainty; • there are insignificant errors in the measurement of the initial data
$x_4 = 2$	• Для анализа рисков использованы исходные данные, сформированные на основе достаточно большого объема эмпирической информации; • отсутствуют погрешности в измерении исходных данных • For the analysis of risks, the initial data generated based on a sufficiently large amount of empirical information are used; • there are no errors in the measurement of the initial data
Правила кодирования показателя x_5 / Rules for coding indicator x_5	
$x_5 = 0$	• Процесс расчета риска содержит серьезные ошибки; • возникла рассогласованность в результатах расчетов с результатами известных значений (отклонения результатов от эталона) • The process of calculating the risk contains serious errors; • there was a mismatch in the results of calculations with the results of known values (deviation of results from the standard)
$x_5 = 1$	• Процесс расчета риска не содержит серьезных ошибок; • могут существовать только погрешности самого процесса расчета (например, точность моделирования по методу Монте-Карло или использования каких-либо численных методов); • неопределенности, вызванные погрешностями, не моделируются • The process of calculating the risk does not contain serious errors; • there can only exist errors in the calculation process itself (for example, the accuracy of the Monte Carlo simulation or the use of any numerical methods); • uncertainties due to errors are not modeled
$x_5 = 2$	• Могут существовать только погрешности самого процесса расчета (например, точность моделирования по методу Монте-Карло или использования каких-либо численных методов); • неопределенности, вызванные погрешностями, корректно смоделированы • There can only exist errors in the calculation process itself (for example, the accuracy of the Monte Carlo simulation or the use of any numerical methods); • uncertainties caused by errors, correctly modeled

Таблица 2. Уровни достоверности и надежности количественного анализа риска**Table 2.** Levels of reliability and quantitative risk analysis

Уровень достоверности Level of reliability	Описание / Description
$D = 0$: ненадежный $D = 0$: unreliable	- Результат количественного анализа риска объекта некачественный; - никакие дальнейшие суждения о достоверности количественного анализа риска невозможны; - результаты количественного анализа риска не должны использоваться для поддержки принятия решений - The result of a quantitative risk analysis of the object is substandard; - no further judgment on the reliability of quantitative risk analysis is possible; - the results of a quantitative risk analysis should not be used to support decision making
$D = 1$: надежный, но не в полной мере достоверный $D = 1$: reliable but not fully reliable	- Результат количественного анализа риска качественный, но: - некоторые критические факторы и явления не идентифицируются и не анализируются; - или некоторые важные риски неточно определены. - Результаты такого количественного анализа риска могут использоваться для поддержки принятия некоторых решений, но не для принятия критически важных для безопасности объектов решений - The result of a quantitative risk analysis is qualitative, but: - some critical factors and phenomena are not identified and analyzed; - or some important risks are not accurately determined. - The results of such a quantitative risk analysis can be used to support the adoption of certain decisions, but not to make decisions critical to the security of objects
$D = 2$: надежный и достоверный $D = 2$: reliable and veracious	- Результат количественного анализа риска качественный: - все критические факторы и явления идентифицированы и учитываются при анализе риска; - все важные риски (и их неопределенности) точно определены. - Результаты такого количественного анализа риска могут использоваться для поддержки принятия важных решений - The result of quantitative risk analysis is qualitative: - all critical factors and phenomena are identified and taken into account in risk analysis; - all important risks (and their uncertainties) are precisely defined. - The results of this quantitative risk analysis can be used to support the adoption of important decisions

критериев $x_1, x_2, \dots, x_5$, определить соответствующий уровень для надежности и достоверности D .

Следует отметить, что и структура оценки, представленная на рис. 2, и правила кодирования (см. табл. 1 и 2) построены в общем виде для демонстрации возможностей предлагаемого подхода. Безусловно, они могут быть адаптированы экспертами в практическом применении при решении каких-либо специфических частных задач.

2. Оценка надежности, основанная на наивном байесовском классификаторе

Сначала рассмотрим общий подход к классификации на основе НБК, а затем опишем метод на основе НБК, предлагаемый непосредственно для оценки достоверности и надежности количественного анализа риска объектов нефтегазовой отрасли.

2.1. Наивный байесовский классификатор

Определим $x = [x_1, x_2, \dots, x_n] \in X$ как входной вектор признаков функции классификации, где X — пространство признаков. Соответственно, классификатор (НБК) — это функция $f_{\text{НБК}}$, которая отобра-

жает входной вектор признаков $x \in X$ в выходное значение класса $D \in \{0, 1, \dots, K\}$. Обычно вектор признаков также принимает дискретные значения, т. е. $x_i \in \{0, 1, \dots, n_i\}$, где $i = 1, 2, \dots, n$. Учитывая характеристики вектора x , НБК классифицирует принадлежность объекта к соответствующему классу с максимальной апостериорной вероятностью [26]:

$$D = \arg \max_D P(D|x), \quad (1)$$

где $P(D|x)$ — апостериорная вероятность данного класса D (т. е. данного значения достоверности и надежности количественного анализа риска) при данном значении признака x .

Значение вероятности в (1) рассчитывается на основе применения теоремы Байеса со строгими (наивными) предположениями о независимости [27]:

$$P(D|x) = \frac{P(x, D)}{P(x)} = \frac{P(x|D)P(D)}{\sum_{D=0}^K P(x|D)P(D)}, \quad (2)$$

где $P(D)$ — априорная вероятность данного класса;

$P(x|D)$ — правдоподобие, т. е. вероятность данного значения признака при данном классе;

$P(x)$ — априорная вероятность данного значения признака;

$P(x, D)$ — апостериорная вероятность данного значения признака при данном классе.

С условием независимости признаков для классификации действует апостериорное правило принятия решения. В этом случае числитель в (2) преобразуется к виду:

$$P(x|D)P(D) = P(D) \prod_{i=1}^n P(x_i|D). \quad (3)$$

Заметим, что знаменатель в (2) одинаковый для всех возможных значений D . Соответственно, классификатор — это функция, определенная следующим образом:

$$D = \arg \max_D P(D) \prod_{i=1}^n P(x_i|D). \quad (4)$$

Для применения НБК вероятности $P(D)$ и $P(x_i|D)$ в (4) необходимо оценить на основе данных обучения. Данные обучения представляют собой наборы образцов, для которых известны значения признаков и соответствие при этом D определенному классу. Предположим, что у нас есть $N_{\text{обуч}}$ данных для обучения, обозначенных $(x^{(q)}, D^{(q)})$, $q = 1, 2, \dots, N_{\text{обуч}}$. В этом случае требуемые вероятности аппроксимируются относительными частотами из набора данных обучения:

$$P(D = k) = \frac{\sum_{q=1}^{N_{\text{обуч}}} I(D^{(q)} = k)}{N_{\text{обуч}}}; \quad (5)$$

$$P(x_i = j|D = k) = \frac{\sum_{q=1}^{N_{\text{обуч}}} I(x_i^{(q)} = j, D^{(q)} = k)}{\sum_{q=1}^{N_{\text{обуч}}} I(D^{(q)} = k)}, \quad (6)$$

где $I(\dots)$ — индикаторная (характеристическая) функция; $i = 1, 2, \dots, n$; $j = 0, 1, \dots, n_i$; $k = 0, 1, \dots, K$.

Существует одна потенциальная проблема при оценке вероятностей с использованием выражений (5) и (6). Предположим, что данный класс k и значения какого-либо признака j никогда не встречаются вместе в наборе данных обучения, тогда оценка, основанная на вероятностях, будет равна нулю, т. е.

$$\sum_{q=1}^{N_{\text{обуч}}} I(x_i^{(q)} = j, D^{(q)} = k) = 0. \text{ В этом случае } P(x_i = j|D = k) = 0, \text{ что согласно (3) приводит к } P(x|D) = 0$$

независимо от значений вероятностей для других признаков. Таким образом, при перемножении нулевая оценка приведет к потере информации о других вероятностях. Данная проблема достаточно часто

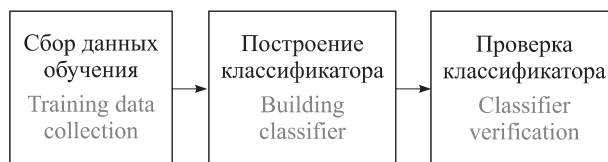


Рис. 3. Схема наивного байесовского классификатора для оценки достоверности количественного анализа риска

Fig. 3. The scheme of a naive Bayesian classifier for assessing the reliability of a quantitative risk analysis

встречается в практических задачах классификации, поэтому предпочтительно вводить небольшие поправки во все оценки вероятностей так, чтобы никакая вероятность не была строго равна нулю. Чтобы избежать данной проблемы, при оценке $P(D = k)$ и $P(x_i = j|D = k)$ часто применяется метод сглаживания по Лапласу [27]:

$$P(D = k) = \frac{\sum_{q=1}^{N_{\text{обуч}}} I(D^{(q)} = k) + \alpha}{N_{\text{обуч}} + (K + 1)\alpha}; \quad (7)$$

$$P(x_i = j|D = k) = \frac{\sum_{q=1}^{N_{\text{обуч}}} I(x_i^{(q)} = j, D^{(q)} = k) + \alpha}{\sum_{q=1}^{N_{\text{обуч}}} I(D^{(q)} = k) + (n_i + 1)\alpha}, \quad (8)$$

где $\alpha \in (0, 1]$ — коэффициент корректировки, введенный для компенсации возможных нулевых значений вероятностей;

$(K + 1), (n_i + 1)$ — число возможных значений соответственно для D и признака x_i .

2.2. Оценка достоверности результатов анализа риска

Рассмотрим возможность применения НБК для разработки классификатора при решении задачи оценки достоверности, структура которой представлена на рис. 2. В этом случае у нас есть пять признаков, т. е. $x = [x_1, x_2, \dots, x_5]$. Каждый признак имеет три дискретных уровня: $x_i \in \{0, 1, 2\}$, $i = 1, 2, \dots, 5$. Следовательно, $X = \{0, 1, 2\} \times \dots \times \{0, 1, 2\} = \{0, 1, 2\}^5$. Достоверность D может принадлежать одному из трех возможных классов (см. табл. 2), т. е. $D \in \{0, 1, 2\}$. Процесс классификации осуществляется в три этапа (рис. 3).

Сбор данных обучения

Байесовский подход основан на теореме, утверждающей, что если плотности распределения значений каждого из признаков известны, то искомый алгоритм можно выписать в явном аналитическом виде. Более того, этот алгоритм оптимален, т. е. обладает минимальной вероятностью ошибок. На практике плотности распределения чаще всего неизвестны.

Их необходимо оценивать (восстанавливать) по обучающей выборке. В результате байесовский алгоритм перестает быть оптимальным, так как восстановить плотность по выборке можно лишь с некоторой погрешностью.

Поскольку $X = \{0, 1, 2\}^5$, то вектор признаков x может принимать $3^5 = 243$ разных значения. Часть из них, обозначаемая $x^{(q)}$, $q = 1, 2, \dots, N_{\text{обуч}}$, выбирается в качестве образцов обучения. Достоверность этих учебных образцов, обозначенных $D^{(q)}$, $q = 1, 2, \dots, N_{\text{обуч}}$, оценивается экспертами согласно описанию, представленному в табл. 1. Затем данные обучения используются для построения НБК, который далее заменяет эксперта при оценке достоверности количественного анализа риска на объектах нефтегазовой отрасли.

Поскольку НБК обучается на данных обучения на основе экспертной оценки, важно, чтобы используемые данные обучения в полной мере охватывали все пространство признаков. С другой стороны, хочется максимально сократить количество данных обучения, поскольку зачастую процесс сбора данных для обучения является весьма трудозатратным и требует много времени. План обучения должен быть составлен таким образом, чтобы собранные данные обучения были одинаково распределены и представляли все пространство X .

Другой вопрос, который необходимо учитывать при разработке схемы сбора данных обучения, — это размер выборки $N_{\text{обуч}}$. Большое значение $N_{\text{обуч}}$ повысило бы эффективность разработанного классификатора с точки зрения его точности. С другой стороны, большое количество данных обучения $N_{\text{обуч}}$ трудозатратно при сборе данных (экспертам придется оценивать слишком много сценариев). Следовательно, необходимо найти какой-то компромисс при определении количества данных обучения $N_{\text{обуч}}$.

Построение классификатора

Структурная схема построения НБК представлена на рис. 4. На этапе подготовки размер выборки данных обучения и схемы их сбора определяются с использованием методов, описанных выше. Затем данные обучения ($x^{(q)}$, $D^{(q)}$), $q = 1, 2, \dots, N_{\text{обуч}}$ обрабатываются экспертами по правилам кодирования (см. табл. 2). На этапе обучения классификатор строится путем оценки $P(D)$ и $P(x_i|D)$ из данных обучения с использованием соответственно выражений (7) и (8). На этапе оценки построенный НБК применяется для замены роли экспертов и определения достоверности результатов нового количественного анализа риска. При рассмотрении структуры оценки надежности и достоверности анализа риска значения признаков определяются сначала на основе правил кодирования, приведенных в табл. 2. В ко-



Рис. 4. Схема построения НБК для оценки достоверности количественного анализа риска

Fig. 4. Scheme of NBC construction for assessing the reliability of quantitative risk analysis

нечном счете достоверность и надежность результатов анализа риска устанавливаются на основе построенного классификатора согласно выражению (4).

Проверка классификатора

На практике размер выборки данных обучения, доступных для построения НБК, может оказаться незначительным, что приведет к снижению точности классификатора. Если НБК недостаточно точен, то, по всей вероятности, он не сможет корректно “имитировать” поведение эксперта при оценке достоверности. Поэтому нам необходимо рассмотреть вопрос о проверке разработанного НБК, т. е. убедиться в том, что классификатор может адекватно отражать мнение экспертов.

Тестирование классификатора можно проводить на основе перекрестной проверки (кросс-валидация, cross-validation) метода оценки качества классификатора и его поведения на независимых данных [28–30]. При оценке классификатора имеющиеся в наличии данные разбиваются на k частей. Затем на $(k - 1)$ частях данных производится обучение классификатора, а оставшаяся часть данных участвует в тестировании. Процедура повторяется k раз; в итоге каждая из k частей данных используется для тестирования. В результате получается оценка качества разработанного классификатора с наиболее равномерным использованием имеющихся данных.

Частным случаем является “скользящий контроль”, или перекрестная проверка, с последовательным исключением одного экземпляра данных (leave-

oneout CV, LOOCV), т. е. $k = n$. При этом строится n моделей по $(n - 1)$ выборочным значениям, а исключенный экземпляр данных каждый раз используется для расчета ошибки при классификации. В [31, 32] теоретически обосновано применение “скользящего контроля” и показано, что если исходные выборки независимы, то средняя ошибка перекрестной проверки дает несмещенную оценку ошибки классификации.

Точность результатов классификации измеряется вероятностью (P_D) того, что классификатор правильно определил достоверность и надежность результатов анализа риска. “Скользящий контроль” используется в работе для оценки P_D , где один экземпляр данных обучения остается для проверки классификатора, а остальные используются для обучения. Алгоритм реализации данного метода изображен на рис. 5, где $(x^{(q)}, D^{(q)})$, $q = 1, 2, \dots, N_{\text{обуч}}$ являются данными обучения, $f_{\text{НБК}, i}$ — классификатор, полученный после удаления данных обучения $(x^{(i)}, D^{(i)})$, P_D — оценка точности классификации.

Заключение

1. Несмотря на существующий арсенал различных методов количественного анализа рисков как в российской, так и в международной практике, зачастую остается открытым вопрос оценки достоверности, а также возможности доверия со стороны лиц, принимающих управленческие решения, полученным результатам.

2. Предлагается подход к оценке надежности и достоверности количественного анализа риска на объектах нефтегазовой отрасли, основанный на системе классификации, а именно на применении наивного байесовского классификатора, который позволяет самостоятельно давать экспертную оценку достоверности и надежности результатов анализа риска, заменяя при этом самого эксперта.

3. К достоинствам применения НБК для оценки достоверности количественного анализа риска можно отнести то, что классификация выполняется достаточно легко и быстро, превосходит многие другие алгоритмы и при этом требует меньшего объема обучающих данных. НБК очень хорошо работает с категориальными признаками, что как раз и нашло свое отражение в настоящей работе.

4. Основным ограничением применения НБК является предположение о том, что все критерии при классификации независимы друг от друга. На практике, однако, между критериями могут существовать различные зависимости. Учет данных зависимостей будет рассмотрен в следующих исследованиях. Кроме того, оценки критериев $x_1, x_2, \dots, x_5$ связаны со множеством субъективных суждений,

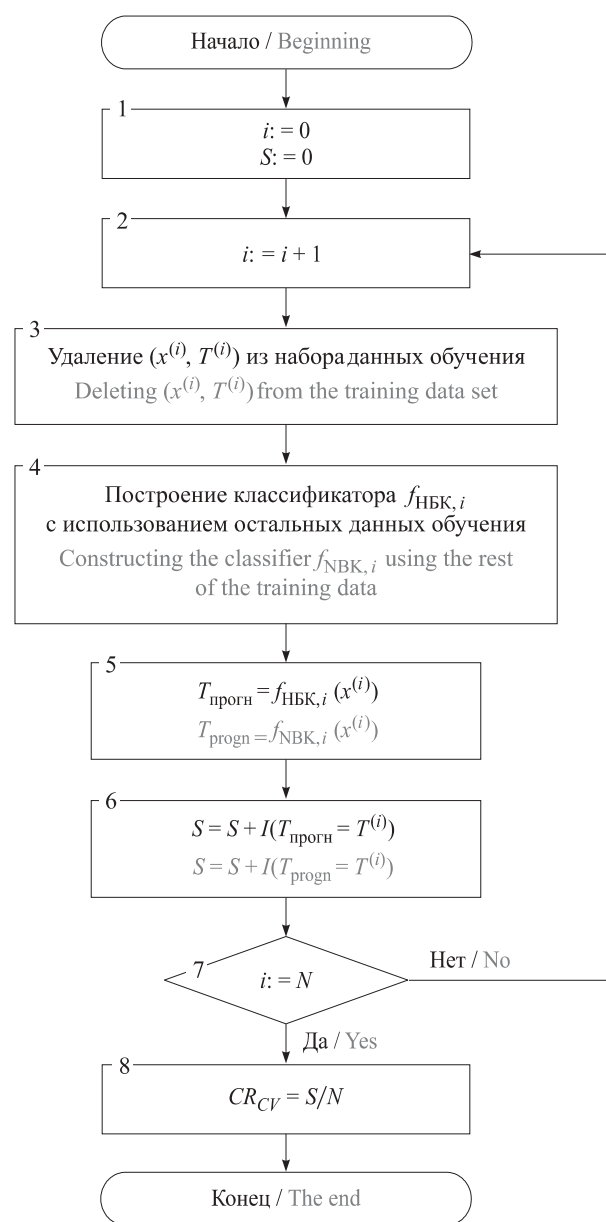


Рис. 5. Алгоритм перекрестной проверки классификатора
Fig. 5. Algorithm for cross checking the classifier

а значит существуют неопределенности при их оценке. Учет данных неопределенностей также будет рассмотрен в будущих исследованиях.

Исследование проведено при финансовой поддержке РФФИ в рамках научного проекта № 16-08-01085.

СПИСОК ЛИТЕРАТУРЫ

1. Грановский Э. А. Техническое регулирование безопасности промышленных объектов: анализ и количественная оценка риска // Безопасность в техносфере. — 2016. — Т. 5, № 5. — С. 54–63. DOI: 10.12737/24152.
2. Можяев А. С., Демидов Ю. Ф. Алгоритмические основы технологии автоматизированного структурно-логического моделирования в задачах системного анализа надежности, безопасности и риска // Моделирование и анализ безопасности и риска в сложных системах : труды Международной научной школы, 2002. — СПб. : Бизнес-пресса, 2002. — 12 с.
3. Белов П. Г. Стратегическое планирование развития и обеспечения национальной безопасности России: прогнозирование и снижение риска чрезвычайных ситуаций // Национальная безопасность и стратегическое планирование. — 2015. — № 1(9). — С. 47–58.
4. Егоров А. Ф., Савицкая Т. В., Михайлова П. Г., Курбатова М. Г. Модели оценки риска возникновения аварий на технологическом оборудовании с опасными химическими веществами. Ч. 1. Теоретические основы // Безопасность в техносфере. — 2008. — № 5. — С. 4–13.
5. Острейковский В. А., Смолин Д. И. Количественная оценка вероятностей исходных событий при анализе риска от эксплуатации атомных станций // Вестник кибернетики. — 2013. — № 12. — С. 130–137.
6. Матвеев А. В., Иванов М. В., Шевченко А. Б. Аналитическая модель системы управления пожарной безопасностью АЭС // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика. Телекоммуникации. Управление. — 2010. — Т. 6. — № 113. — С. 91–95.
7. Серенков П. С., Воронин А. Н., Литский В. К. Анализ отечественных и зарубежных методик оценки риска в магистральном трубопроводном транспорте // Вестник Полоцкого государственного университета. Серия F: Строительство. Прикладные науки. — 2013. — № 8. — С. 83–88.
8. Абдрахманов Н. Х., Абдрахманова К. Н., Ворохобко В. В., Абдрахманов Р. Н., Басырова А. Р. Моделирование сценариев развития аварийных ситуаций для нестационарных опасных производственных объектов нефтегазового комплекса // Нефтегазовое дело : электронный научный журнал. — 2015. — № 5. — С. 516–531. DOI: 10.17122/ogbus-2015-5-516-531.
9. Малиев Е. М., Шиянов А. В. Исследование достоверности количественного метода анализа рисков для оценки состояния уровня промышленной безопасности объектов нефтеперерабатывающей промышленности // Актуальные вопросы современной науки : материалы II Международной научно-практической конференции. — Чистополь : Бриг, 2015. — С. 34–39.
10. Сидорова М. Н., Хасбутдинова Е. В. Современные методы качественного и количественного анализа рисков в нефтегазовой отрасли // Фундаментальная математика и ее приложения в естествознании : тезисы докладов IX Международной школы-конференции для студентов, аспирантов и молодых ученых (3–7 октября 2016 г., г. Уфа) / Отв. ред. Б. Н. Хабибуллин, Е. Г. Екомасов, Р. М. Ахметханов. — Уфа : РИЦ БашГУ, 2016. — С. 350–352.
11. Лесных В. В., Алексеева В. А., Литвин Ю. В. Современное состояние проблемы анализа организационных рисков: терминология, классификация, методы качественной и количественной оценки // Управление риском. — 2015. — № 1(73). — С. 14–25.
12. Абдрахманов Н. Х., Шайбаков Р. А., Марков А. Г. Анализ современного уровня развития методологии системных рисков при проектировании и эксплуатации опасных производственных объектов // Горный информационно-аналитический бюллетень (научно-технический журнал). — 2015. — № 57. — С. 789–795.
13. Aven T. Risk assessment and risk management: Review of recent advances on their foundation // European Journal of Operational Research. — 2016. — Vol. 253, Issue 1. — P. 1–13. DOI: 10.1016/j.ejor.2015.12.023.
14. Aven T., Heide B. Reliability and validity of risk analysis // Reliability Engineering and System Safety. — 2009. — Vol. 94, Issue 11. — P. 1862–1868. DOI: 10.1016/j.ress.2009.06.003.

15. Lauridsen K., Christou M., Amendola A., Markert F., Kozine I., Fiori M. Assessing the uncertainties in the process of risk analysis of chemical establishments: Part 1 and 2 // Proceedings of European Conference on Safety and Reliability (ESREL). — Torino, Italy : Politecnico di Torino, 2001. — P. 592–606.
16. Sornette D., Maillart T., Kröger W. Exploring the limits of safety analysis in complex technological systems // International Journal of Disaster Risk Reduction. — 2013. — Vol. 6. — P. 59–66. DOI: 10.1016/j.ijdr.2013.04.002.
17. Rouhiainen V. QUASA: A method for assessing the quality of safety analysis // Safety Science. — 1992. — Vol. 15, Issue 3. — P. 155–172. DOI: 10.1016/0925-7535(92)90002-h.
18. Kirchsteiger C. On the use of probabilistic and deterministic methods in risk analysis // Journal of Loss Prevention in the Process Industries. — 1999. — Vol. 12, Issue 5. — P. 399–419. DOI: 10.1016/S0950-4230(99)00012-1.
19. Pinto A., Ribeiro R. A., Nunes I. L. Ensuring the quality of occupational safety risk assessment // Risk Analysis. — 2013. — Vol. 33, Issue 3. — P. 409–419. DOI: 10.1111/j.1539-6924.2012.01898.x.
20. Горелик А. Л., Скрипкин В. А. Методы распознавания. — 4-е изд. — М. : Высшая школа, 2004. — 262 с.
21. Асминг В. Э., Кременецкая Е. О., Виноградов Ю. А., Федоров А. В. О применении наивных байесовских классификаторов в сейсмологии // Сейсмические приборы. — 2015. — Т. 51, № 4. — С. 29–40.
22. Гулин В. В. Сравнительный анализ методов классификации текстовых документов // Вестник Московского энергетического института. — 2011. — № 6. — С. 100–108.
23. Матвеев А. В. Системное моделирование управления риском возникновения чрезвычайных ситуаций : дис. ... канд. техн. наук. — СПб. : СПб УГПС МЧС России, 2007. — 150 с.
24. Попова О. А. Анализ новых подходов к представлению неопределенности в данных для крупномасштабных систем // Управление развитием крупномасштабных систем : материалы Восьмой международной конференции (29 сентября – 1 октября 2015 г., г. Москва) : в 2 т. / Под общ. ред. С. Н. Васильева, А. Д. Цвиркуна. — М. : Институт проблем управления им. В. А. Трапезникова РАН, 2015. — Т. 2. — С. 385–387.
25. Ferson S. RAMAS Risk Calc 4.0 Software: Risk assessment with uncertain numbers. — Boca Raton, Florida : Lewis Publishers, 2002.
26. Rish I. An empirical study of the naive Bayes classifier // Workshop on empirical methods in artificial intelligence, IJCAI. — 2001. — Vol. 3. — P. 41–46. URL: <https://www.cc.gatech.edu/~isbell/reading/papers/Rish.pdf> (дата обращения: 10.11.2017).
27. Narayanan V., Arora I., Bhatia A. Fast and accurate sentiment classification using an enhanced Naive Bayes model // Intelligent Data Engineering and Automated Learning — IDEAL 2013. — Berlin, Heidelberg : Springer, 2013. — P. 194–201. DOI: 10.1007/978-3-642-41278-3_24.
28. Шитиков В. К., Мاستицкий С. Э. Классификация, регрессия и другие алгоритмы Data Mining с использованием R. — Тольятти, Лондон, 2017. — 351 с. URL: <http://www.ievbras.ru/ecostat/Kiril/R/DM/DM.R.pdf> (дата обращения: 10.11.2017).
29. Krstajic D., Buturovic L., Leahy D., Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification methods // Journal of Cheminformatics. — 2014. — Vol. 6, Issue 1. — 15 p. DOI: 10.1186/1758-2946-6-10.
30. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning. Data Mining, Inference, and Prediction. — 2nd ed. — New York : Springer, 2009. — 745 p. DOI: 10.1007/978-0-387-84858-7.
31. Алгоритмы и программы восстановления зависимостей / Под ред. В. Н. Вапника. — М. : Наука, 1984. — 816 с.
32. Raghuraj R., Lakshminarayanan S. Variable predictive models — A new multivariate classification approach for pattern recognition applications // Pattern Recognition. — 2009. — Vol. 42, No. 1. — P. 7–16. DOI: 10.1016/j.patcog.2008.07.005.

Материал поступил в редакцию 21 ноября 2017.

Для цитирования: Матвеев А. В., Максимов А. В., Щербаков О. В., Смирнов А. С. Метод оценки достоверности количественного анализа риска на объектах нефтегазовой отрасли // Пожаровзрывобезопасность / Fire and Explosion Safety. — 2018. — Т. 27, № 1. — С. 35–49. DOI: 10.18322/PVB.2018.27.01.35-49.

METHOD OF ESTIMATION FOR THE RELIABILITY OF QUANTITATIVE RISK ANALYSIS ON OBJECTS OF OIL AND GAS INDUSTRY

MATVEEV A. V., Candidate of Technical Sciences, Docent of Department of Applied Mathematics and Information Technology, Saint Petersburg University of State Fire Service of Emercom of Russia (Moskovskiy Avenue, 149, Saint Petersburg, 196105, Russian Federation; e-mail: fcvega_10@mail.ru)

MAKSIMOV A. V., Candidate of Technical Sciences, Lecturer of Department of Applied Mathematics and Information Technology, Saint Petersburg University of State Fire Service of Emercom of Russia (Moskovskiy Avenue, 149, Saint Petersburg, 196105, Russian Federation; e-mail: he1nze@mail.ru)

SHCHERBAKOV O. V., Doctor of Technical Sciences, Professor, Professor of Department of Applied Mathematics and Information Technology, Saint Petersburg University of State Fire Service of Emercom of Russia (Moskovskiy Avenue, 149, Saint Petersburg, 196105, Russian Federation)

SMIRNOV A. S., Doctor of Technical Sciences, Professor, Emercom of Russia (Davydkovskaya St., 7, Moscow, 121352, Russian Federation)

ABSTRACT

Despite the existing arsenal of various methods of quantitative risk analysis both in Russian and international practice, problems arise related to the assessment of the reliability of the results obtained during their practical application, including at oil and gas facilities. The question remains to what extent the decision-maker can trust the results. The article considers the problem of assessing the reliability of quantitative risk analysis at oil and gas facilities. Existing methods for estimating reliability are investigated. It is proposed to use an approach based on ensuring the quality of the process of risk analysis itself. To increase the objectivity in assessing the reliability of the results of a risk analysis, a formal quantitative method was proposed. The article introduces 5 criteria that ensure the quality of the process of risk analysis at oil and gas facilities. A system of rules for coding the values of each of the basic criteria into three discrete qualitative levels was developed. The solution of the task was accomplished by constructing a classifier in which the reliability index of a quantitative risk analysis of oil and gas industry objects is a function of the values of the basic criteria. The reliability of the risk analysis was evaluated on the basis of a naive Bayesian classifier that takes into account the values of the five basic criteria in the evaluation framework. The results of the classifier work are based on a variety of training data that were previously evaluated by experts. The article suggests an approach to the assessment of the quality of the classifier itself, based on a cross-checking with successive exclusion of one copy of the training data. The merits of using the naive Bayesian classifier for assessing the reliability of quantitative risk analysis in oil and gas industry objects include the fact that the classification is carried out quite easily and quickly, surpasses many other algorithms, and requires a smaller amount of training data. A naive Bayesian classifier works very well with categorical features, which is exactly what is reflected in this article.

Keywords: quantitative risk analysis; reliability; trustworthiness; naive Bayesian classifier; training data; risk.

REFERENCES

1. Granovskiy E. A. Technical regulation of safety of industrial facilities: analyses and risk quantitative assessment. *Bezopasnost v tekhnosfere / Safety in Technosphere*, 2016, vol. 5, no. 5, pp. 54–63 (in Russian). DOI: 10.12737/24152.
2. Mozhaev A. S., Demidov Yu. F. Algorithmic foundations of automated structural-logical modeling technology in problems of system analysis of reliability, safety and risk. In: *Modelirovaniye i analiz bezopasnosti i riska v slozhnykh sistemakh* [Modeling and analysis of safety and risk in complex systems]. Proceedings of the International Scientific School. Saint Peterburg, Biznes-pressa Publ., 2002. 12 p. (in Russian).

3. Belov P. G. Strategic planning of development and ensuring national security of Russia: forecasting and decrease in risk of emergency situations. *Natsionalnaya bezopasnost i strategicheskoye planirovaniye / National Security and Strategic Planning*, 2015, no. 1(9), pp. 47–58 (in Russian).
4. Egorov A. F., Savitskaya T. V., Mikhaylova P. G., Kurbatova M. G. Models for assessing the risk of accidents on process equipment with hazardous chemicals. Part 1. Theoretical basis. *Bezopasnost v tekhnosfere / Safety in Technosphere*, 2008, no. 5, pp. 4–13 (in Russian).
5. Ostrejkovsky V. A., Smolin D. I. The quantification for probabilities of the initial events under consideration of operational risk at nuclear power stations. *Vestnik kibernetiki / Proceedings in Cybernetics*, 2013, no. 12, pp. 130–137 (in Russian).
6. Matveev A. V., Ivanov M. V., Shevchenko A. B. Analytical model of a control system of fire safety on the atomic power stations. *Nauchno-tekhnicheskiye vedomosti Sankt-Peterburgskogo gosudarstvennogo politekhnicheskogo universiteta. Informatika. Telekommunikatsii. Upravleniye / St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems*, 2010, vol. 6, no. 113, pp. 91–95 (in Russian).
7. Serenkov P. S., Voronin A. N., Lipsky V. K. The analysis of domestic and foreign methods of risk assessment in main pipeline transport. *Vestnik Polotskogo gosudarstvennogo universiteta. Seriya F: Stroitelstvo. Prikladnyye nauki / Vestnik of Polotsk State University. Part F. Constructions. Applied Sciences*, 2013, no. 8, pp. 83–88 (in Russian).
8. Abdrakhmanov N. Kh., Abdrakhmanova K. N., Vorohobko V. V., Abdrakhmanov R. N., Basyrova A. R. Modeling of scenarios of development of emergencies for non-stationary hazardous production facilities of an oil and gas complex. *Elektronnyy nauchnyy zhurnal "Neftegazovoye delo" / Electronic scientific Journal "Oil and Gas Business"*, 2015, no. 5, pp. 516–531 (in Russian). DOI: 10.17122/ogbus-2015-5-516-531.
9. Maliev E. M., Shiyarov A. V. Investigation of the reliability of a quantitative risk analysis method for assessing the state of the level of industrial safety of oil refining facilities. In: *Aktualnyye voprosy sovremennoy nauki* [Actual problems of modern science]. Proceedings of II International Scientific and Practical Conference. Chistopol, Brig Publishing House, 2015, pp. 34–39 (in Russian).
10. Sidorova M. N., Khasbutdinova E. V. Modern methods of qualitative and quantitative analysis of risks in the oil and gas industry. In: Khabibullin B. N., Ekomasov E. G., Akhmetkhanov R. M. (resp. eds.). *Fundamentalnaya matematika i yeye prilozheniya v yestestvoznaniy* [Fundamental mathematics and its applications in natural science]. Abstracts of IX International School-Conference for students, graduate students and young scientists (3–7 October 2016, Ufa). Ufa, BSU Publishing center, 2016, pp. 350–352 (in Russian).
11. Lesnykh V. V., Alekseeva V. A., Litvin Yu. V. Current state of organizational risk analysis problem: terms, classification, methods of qualitative and quantitative assessment. *Upravleniye riskom / Risk Management*, 2015, no. 1(73), pp. 14–25 (in Russian).
12. Abdrakhmanov N. Kh., Shaybakov R. A., Markov A. G. The analysis of a modern level of development of methodology of system risks at design and operation of hazardous production facilities. *Gornyy informatsionno-analiticheskiy byulleten (nauchno-tekhnicheskiy zhurnal) / Mining Informational and Analytical Bulletin (Scientific and Technical Journal)*, 2015, no. 57, pp. 789–795 (in Russian).
13. Aven T. Risk assessment and risk management: Review of recent advances on their foundation. *European Journal of Operational Research*, 2016, vol. 253, issue 1, pp. 1–13. DOI: 10.1016/j.ejor.2015.12.023.
14. Aven T., Heide B. Reliability and validity of risk analysis. *Reliability Engineering and System Safety*, 2009, vol. 94, issue 11, pp. 1862–1868. DOI: 10.1016/j.ress.2009.06.003.
15. Lauridsen K., Christou M., Amendola A., Markert F., Kozine I., Fiori M. Assessing the uncertainties in the process of risk analysis of chemical establishments: Part 1 and 2. In: *Proceedings of European Conference on Safety and Reliability (ESREL)*. Torino, Italy, Politecnico di Torino, 2001, pp. 592–606.
16. Sornette D., Maillart T., Kröger W. Exploring the limits of safety analysis in complex technological systems. *International Journal of Disaster Risk Reduction*, 2013, vol. 6, pp. 59–66. DOI: 10.1016/j.ijdr.2013.04.002.
17. Rouhiainen V. QUASA: A method for assessing the quality of safety analysis. *Safety Science*, 1992, vol. 15, issue 3, pp. 155–172. DOI: 10.1016/0925-7535(92)90002-h.
18. Kirchsteiger C. On the use of probabilistic and deterministic methods in risk analysis. *Journal of Loss Prevention in the Process Industries*, 1999, vol. 12, issue 5, pp. 399–419. DOI: 10.1016/S0950-4230(99)00012-1.
19. Pinto A., Ribeiro R. A., Nunes I. L. Ensuring the quality of occupational safety risk assessment. *Risk Analysis*, 2013, vol. 33, issue 3, pp. 409–419. DOI: 10.1111/j.1539-6924.2012.01898.x.

20. Gorelik A. L., Skripkin V. A. *Metody raspoznavaniya. 4-ye izd.* [Recognition methods. 4th ed.]. Moscow, Vysshaya shkola Publ., 2004. 262 p. (in Russian).
21. Asming V. E., Kremenetskaya E. O., Vinogradov Yu. A., Fedorov A. V. On usage of naive Bayesian classifiers in seismology. *Seismicheskiye pribory / Seismic Instruments*, 2015, vol. 51, no. 4, pp. 29–40 (in Russian).
22. Gulin V. V. Comparative analysis of methods for text document classification. *Vestnik Moskovskogo energeticheskogo instituta / MPEI Vestnik*, 2011, no. 6, pp. 100–108 (in Russian).
23. Matveev A. V. *System modeling of emergency risk management*. Cand. tech. sci. diss. Saint Petersburg, 2007. 150 p. (in Russian).
24. Popova O. A. Analysis of new approaches to presenting uncertainty in data for large-scale systems. In: Vasilyev S. N., Tsvirkun A. D. (eds.). *Upravleniye razvitiyem krupnomasshtabnykh sistem* [Management of the development of large-scale systems]. Proceedings of 8th International Conference. Moscow, V. A. Trapeznikov Institute of Control Sciences RAS, 2015, vol. 2, pp. 385–387 (in Russian).
25. Ferson S. *RAMAS Risk Calc 4. 0 Software: Risk assessment with uncertain numbers*. Boca Raton, Florida, Lewis Publishers, 2002.
26. Rish I. An empirical study of the naive Bayes classifier. In: *Workshop on empirical methods in artificial intelligence, IJCAI*, 2001, vol. 3, pp. 41–46. Available at: <https://www.cc.gatech.edu/~isbell/reading/papers/Rish.pdf> (Accessed 10 November 2017).
27. Narayanan V., Arora I., Bhatia A. Fast and accurate sentiment classification using an enhanced Naive Bayes model. In: *Intelligent Data Engineering and Automated Learning — IDEAL 2013*. Berlin, Heidelberg, Springer, 2013, pp. 194–201. DOI: 10.1007/978-3-642-41278-3_24.
28. Shitikov V. K., Mastitskiy S. E. *Klassifikatsiya, regressiya i drugiye algoritmy Data Mining s ispolzovaniyem R* [Classification, regression and other Data Mining algorithms using R]. Tolyatti, London, 2017. 351 p. (in Russian). Available at: <http://www.ievbras.ru/ecostat/Kiril/R/DM/DM R.pdf> (Accessed 10 November 2017).
29. Krstajic D., Buturovic L., Leahy D., Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification methods. *Journal of Cheminformatics*, 2014, vol. 6, issue 1. 15 p. DOI: 10.1186/1758-2946-6-10.
30. Hastie T., Tibshirani R., Friedman J. *The elements of statistical learning. Data Mining, Inference, and Prediction*. 2nd ed. New York, Springer, 2009. 745 p. DOI: 10.1007/978-0-387-84858-7.
31. Vapnik V. N. (ed.). *Algoritmy i programmy vosstanovleniya zavisimostey* [Algorithms and programs for recovering dependencies]. Moscow, Nauka Publ., 1984. 816 p. (in Russian).
32. Raghuraj R., Lakshminarayanan S. Variable predictive models — A new multivariate classification approach for pattern recognition applications. *Pattern Recognition*, 2009, vol. 42, no. 1, pp. 7–16. DOI: 10.1016/j.patcog.2008.07.005.

For citation: Matveev A. V., Maksimov A. V., Shcherbakov O. V., Smirnov A. S. Method of estimation for the reliability of quantitative risk analysis on objects of oil and gas industry. *Pozharovzryvobezopasnost / Fire and Explosion Safety*, 2018, vol. 27, no. 1, pp. 35–49 (in Russian). DOI: 10.18322/PVB.2018.27.01.35-49.