

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

УДК 004.89+336.713

АЛГОРИТМИ DATA MINING В АНАЛІЗІ НАДЗВИЧАЙНИХ СИТУАЦІЙ, ЩО ВИНИКАЮТЬ ПРИ ПЕРЕВЕЗЕННІ РАДІОАКТИВНО-НЕБЕЗПЕЧНИХ РЕЧОВИН

Юхимчук Сергій Васильович, Савчук Тамара Олександрівна, Вознюк Андрій Вікторович

Анотация: В роботі запропоновано підхід щодо аналізу надзвичайних техногенних ситуацій на залізниці, що виникають при наявності радіоактивних речовин. Використання алгоритмів Data Mining надає змогу підвищити швидкість аналізу аварійної ситуації. Пошук правил, що відображають аналітичні залежності між параметрами предметної області автоматично дасть змогу зменшити фактору суб'єктивності при аналізі надзвичайної ситуації.

Аннотация: В работе предложен подход к анализу чрезвычайных техногенных ситуаций на железной дороге, возникающих при наличии радиоактивных веществ. Использование алгоритмов Data Mining позволяет повысить быстроту действия анализа аварийной ситуации. Поиск правил, отражающих аналитические зависимости между параметрами предметной области автоматически, позволит уменьшить фактор субъективности при анализе чрезвычайной ситуации.

Abstract: This method proposed an approach for analyzing technological emergency situations on the railroad, which arise in the presence of radioactive substances. Using Data Mining algorithms allows to increase the performance analysis of the emergency situation. Search rules that reflect the analytical relationship between parameters of the subject field will automatically reduce the factor of subjectivity in the analysis of emergency situations.

Вступ

Висока кількість вантажоперевезень відбувається залізничним транспортом. Під час перевезення радіоактивно-небезпечних вантажів можливе виникнення надзвичайних ситуацій [1].

Для того, щоб вирішити як ліквідувати надзвичайну ситуацію, що виникає при перевезенні радіоактивно-небезпечних речовин, та її наслідки потрібно знати можливі шляхи її розвитку, що приводить до необхідності вирішити задачу аналізу надзвичайної ситуації. Завчасний і оперативний аналіз надзвичайної ситуації дозволяє істотно знижувати величину втрат.

Особливістю швидкоплинних техногенних надзвичайних ситуацій є велика кількість факторів, що можуть впливати на її подальший розвиток, неточність та неповнота інформації щодо початкового стану надзвичайної ситуації [2].

З метою протидії аварії, що виникла при перевезенні радіоактивних речовин, керівникові ліквідації аварії необхідно у найкоротший термін зібрати інформацію про таку ситуацію, визначити відповідний сценарій її розвитку та розміри зон ураження від дії небезпечних факторів, оцінити наслідки та провести аналіз використання наявних сил та засобів щодо її ліквідації.

Для отримання ефективного розв'язку цієї задачі необхідно автоматизувати процеси управління ліквідацією аварій, що виникають при перевезенні небезпечних вантажів [3].

Успішне функціонування автоматизованих систем неможливе без математичних моделей, за допомогою яких здійснюється ідентифікація небезпечної події за зовнішніми ознаками та визначення шляху розвитку надзвичайної ситуації, що дає змогу об'єктивно оцінити ситуацію, зменшити ризик втрат від аварії, а також надасть можливість в подальшому провести ефективну її ліквідацію.

До існуючих методів розв'язку поставленої задачі можна віднести такі:

- дані про надзвичайну ситуацію опрацьовує експерт (група експертів), що на основі досвіду та знань визначають шляхи її розвитку. Підхід визначається малою швидкістю опрацювання отриманої інформації, необхідністю наявності спеціально підготовлених експертів, суб'єктивністю оцінки ситуації, що виникає внаслідок впливу рішень одних експертів на інших;
- на основі заздалегідь складених інформаційно-довідкових таблиць, що визначають проміжні стани параметрів системи визначаються прогнозовані шляхи їх розвитку. Отримані за різними параметрами дані корелюються між собою та визначається шлях розвитку аварії в цілому. При використанні даного підходу ситуація розбивається на етапи розвитку аварії, що визначаються значеннями параметрів, взятих з інформаційно-довідкових таблиць, а тому можуть бути не розглянуті проміжні стани аварії, які можуть сильно впливати на її розвиток, і, як наслідок отримання помилкового результату аналізу аварії;
- визначаються аналітичні залежності між різними параметрами аварії. На основі утворених законів знаходяться правила зміни параметрів, що визначають кінцевий стан аварії в залежності від значень початкових параметрів. Метод відрізняється значною точністю прогнозу розвитку аварії, проте досить низькою швидкістю, навіть при використанні ЕОМ, внаслідок великої кількості початкових параметрів, та різного впливу кожного з них у розвиток аварії в різних ситуаціях.

Як видно з аналізу існуючих методів та підходів для розв'язання поставленої задачі, основними проблемами є низька швидкість та вплив суб'єктивного фактору на аналіз аварійних ситуацій. Це

визначає необхідність автоматизованого вирішення поставленої задачі, що можуть дати інформаційні технології.

Зважаючи на суб'єктивність оцінки в існуючих методах та з метою зменшення її впливу, актуальним є використання інформаційних технологій для розв'язку поставленої задачі.

Використання елементів традиційних підходів залишає точність результатів прогнозу та аналізу.

Мета дослідження

Метою дослідження є аналіз застосування Data Mining для аналізу радіоактивно-небезпечних надзвичайних ситуацій на залізниці. На основі початкових, можливо неповних, даних про радіоактивно-небезпечну надзвичайну ситуацію, провести її аналіз та спрогнозувати можливі шляхи її розвитку.

Розв'язання проблеми

Параметри предметної області можуть визначатися числовими значеннями, бути лінгвістичними змінними чи такими, що визначатимуться в процесі аналізу надзвичайної ситуації за табличними даними на основі інших. Для зберігання усіх параметрів про надзвичайну ситуацію, окрім тих, що є початковими, тобто задаються на початку аналізу, використовуються бази даних.

Отже, є доцільним використання таких алгоритмів аналізу надзвичайних ситуацій, що взаємодіють безпосередньо з базою даних. При визначенні характеристик, пов'язаних з небезпечною подією, використовуються багатокритеріальні запити до бази даних, що можуть бути й багаторівневими. При використанні реляційних баз даних ефективність пошуку буде зменшуватись експоненційно з ростом кількості критеріїв запиту. Тому, для усунення цього недоліку необхідно використовувати багатовимірні бази даних [4].

Для аналізу аварії необхідні знання, що нададуть інформацію про те, які аналітичні залежності є між параметрами, що описують надзвичайну ситуацію. Для виділення цих знань та використання їх для подальшого аналізу, слід використовувати алгоритми Data Mining [5].

Алгоритми Data Mining дозволять знайти в базі даних не тільки знання про правила розвитку радіоактивно-небезпечної надзвичайної ситуації, що вже є відомими чи тривіальними, а й такі, що були невідомими до початку аналізу радіоактивно-небезпечної надзвичайної ситуації.

Процес аналізу радіоактивно-небезпечної надзвичайної ситуації включає такі основні кроки:

- на основі набору даних про аварію з радіоактивними речовинами будується модель радіоактивно-небезпечної надзвичайної ситуації, що відобразить аналітичні залежності між параметрами аварії та правилами, що їх визначатимуть;
- на основі побудованої моделі проводитиметься аналіз аварійної ситуації.

При аналізі радіоактивно-небезпечної надзвичайної ситуації слід розв'язати такі інтелектуальні задачі Data Mining [6]: класифікація (визначення класу об'єкту радіоактивно-небезпечної аварії за його характеристиками), пошук асоціативних правил (для аварій з однаковими фінальними станами і діями, необхідними для їх ліквідації визначаються фактори, що є спільними, на основі яких виводяться нові правила розвитку аварій), кластеризація (розбиття множини можливих аварій на деякі, спорідненні за ознакою фінального стану аварії, підмножини за параметрами, що не є наперед відомими, а визначаються у процесі обробки ситуації).

Задача класифікації зводиться до визначення приналежності об'єкту аварії до певної групи значень параметрів, що поділяються за типом зони радіаційного забруднення за такими параметрами, як початкові властивості радіоактивно-небезпечної аварії (агрегатний стан, активність, тип радіоактивної речовини, причина виникнення аварії, параметри зовнішнього середовища). Кожний клас об'єкту визначає особливий шлях розвитку надзвичайної ситуації.

Проведемо класифікацію за таким алгоритмом:

- введемо ймовірність $P(y = c_r)$ того, що радіоактивно-небезпечна аварія відноситься до класу c_r , що включає в себе такі можливі наслідки: вибух, пожежа, викид радіоактивної речовини у атмосферу, вихід радіоактивної речовини, відсутність шляху розвитку аварії;
- введемо подію рівності атрибутів об'єкту радіоактивно-небезпечної аварії деяким значенням, E та ймовірність настання події $P(E)$;
- сформуємо правила, в умовних частинах яких порівнюються всі незалежні зміни з допустимими значеннями.

Тоді для об'єкта, що додається до бази даних, ймовірність належності об'єкту класу c_r при умові рівності його параметру x_h значенню c_d^h визначається, так [5]:

$$P(x_h = c_d^h | y = c_r) = \frac{P(x_h = c_d^h \text{ і } y = c_r)}{P(y = c_r)}, \quad (1)$$

що дозволяє розв'язати задачу класифікації.

Даний метод можливо використовувати при незалежності множини параметрів надзвичайної ситуації, що опрацьовуються. Для задоволення даному правилу усі параметри про надзвичайну ситуацію розбиваються на декілька множин $\lambda_1, \lambda_2, \dots, \lambda_n$. При цьому,

λ_1 – множина вхідних параметрів про надзвичайну ситуацію;

λ_n – множина кінцевих параметрів, що описує проведений аналіз надзвичайної ситуації;

$\lambda_i, i = 2, n-1$ – множини проміжних параметрів, причому таких, що кожна наступна множина є множиною декларативних знань, отриманих на основі попередніх.

Для пошуку асоціативних правил, на основі маси радіоактивної речовини, її здатності до взаємодії з навколишнім середовищем та здатності до ланцюгової реакції визначатимуться залежності між об'єктами надзвичайної ситуації та подіями, що можуть бути ними викликані (виникнення пожежі, вибух, викид радіоактивних речовин у атмосферу). Такі залежності не зможуть бути отримані на етапі класифікації, так як вони не є наперед відомими правилами.

Опишемо аварію множиною значень

$$\langle s_1, s_2, \dots, s_n, f, l_1, l_2, \dots, l_n \rangle, \quad (2)$$

де s_1, s_2, \dots, s_n – множина значень параметрів, що описують початковий стан небезпечної ситуації,

f – ознака фінального стану небезпечної ситуації,

l_1, l_2, \dots, l_n – множина параметрів, що описують необхідні для виконання дії при ліквідації небезпечної ситуації.

Оберемо аварійну ситуацію з деяким визначеним набором значень f_0 та l_1, l_2, \dots, l_n . Підтримкою множини є відношення кількості аварій, опис яких містить у собі підмножину $f_0, l_1, l_2, \dots, l_n$, до загальної кількості аварій, що розглядаються.

На основі обраних частих множин будуються асоціативні правила виду "якщо (умова), то (результат)", де умова – набір значень параметрів аварії, що входять до частотної множини на основі якої будується правило, результат – значення параметрів набору $f_0, l_1, l_2, \dots, l_n$, що визначає розвиток аварії та необхідні дії для її ліквідації.

З метою врахування можливих умов надзвичайної ситуації, створюється набір з усіх можливих перестановок значень параметрів, що описують фінальний стан аварії та дій при її ліквідації. Для кожної множини значень параметрів знаходиться величина підтримки. Потім обираються ті з них, значення підтримки яких більше деякого обраного значення, що буде збільшуватись по мірі росту кількості аварій, що розглядалася поточного моменту.

Задача розбиття заданої вибірки ситуацій на такі, що є підмножинами, які не перетинаються, тобто кластерами (при цьому кожний кластер повинен складатися з об'єктів, а об'єкти різних кластерів суттєво відрізняються), зводиться до пошуку незалежних груп параметрів предметної області за невизначеними характеристиками, що дасть змогу виділяти їх в окремі групи. Це звільнить від необхідності пошуку залежностей в даних, що не відносяться до заданих вхідних параметрів про надзвичайну ситуацію. Відповідно, збільшиться швидкість роботи модулю та виникає можливість використовувати знайдені групи параметрів при подальшому аналізі надзвичайної ситуації.

Таким чином, узагальнений алгоритм, який відображає процес виконання означених інтелектуальних задач Data Mining при аналізі надзвичайних ситуацій, що виникають при перевезенні радіоактивних речовин залізничним транспортом, передбачає такі дії:

1. Визначення початкових параметрів.
2. Застосування правил зв'язаних з відомими параметрами.
3. Розбиття надзвичайних ситуацій на класи в залежності від їх фінального стану.
4. Для поточної ситуації визначити значення ймовірностей входження до визначених класів за формулою 1.
5. Визначення класу з максимальною ймовірністю входження для поточної ситуації.

6. Якщо дана ймовірність більша за визначену, то віднести поточну надзвичайну ситуацію до цього класу, інакше помітити як невизначений клас.
7. Визначення асоціативних правил.
8. Застосування знайдених асоціативних правил до відомих параметрів надзвичайної ситуації.
9. Якщо знайдено нові правила, то перехід до п. 2, інакше перехід до п. 10.
10. Якщо змінювались параметри надзвичайної ситуації, то перехід до п. 2, інакше перехід до п. 11.
11. Визначення фінального стану.
12. Визначення фінального стану надзвичайної ситуації, що відповідає знайденим параметрам.

Схема узагальненого алгоритму, який відображує процес виконання означених інтелектуальних задач Data Mining при аналізі надзвичайних ситуацій, що виникають при перевезенні радіоактивних речовин залізничним транспортом, наведена на рисунку 1.

Висновки

В ході проведеного дослідження було запропоновано підхід до аналізу радіоактивно-небезпечних надзвичайних ситуацій на залізниці, що базується на використанні алгоритмів Data Mining. Це надає можливість збільшити швидкість аналізу надзвичайних ситуацій такого типу та виконувати пошук нових правил, що з урахуванням причинно-наслідкових зв'язків між об'єктами надзвичайної ситуації, яка аналізується, формувати висновок щодо підтримки прийнятих рішень керівниками ліквідації таких ситуацій при значному зменшенні впливу суб'єктивного фактору на цей процес.

Список літератури

1. Черпанов А. Т. Радиационные аварии. Минск:Технопринт, 1993, 215 с.
2. Плахотник В.М., Сорока Н.Ю., Тригуб Л.М. Правила безпеки та порядок ліквідації наслідків аварійних ситуацій при перевезенні їх залізничним транспортом. – К. : Мінтранс України, 2001. – 885 с.
3. Юхимчук С. В., Кацман М. Д., Моделі автоматизації вироблення рекомендацій керівнику гасіння пожежі на залізничному транспорті. – УНІВЕРСУМ-Вінниця, 2008. – 144 с.
4. Зиновьев А. Ю. Многомерные данные. Красноярск, 2000, 180 с.
5. Чубукова И. А. Data mining. С.-Петербург:Бином, 384 с.
6. Дюк В., Самойленко В. Д. Data mining: учебный курс. С.-Петербург:Питер, 368с.

Відомості про авторів

Юхимчук Сергій Васильович – д.т.н., проф., завідувач кафедри інтелектуальних систем, Вінницький національний технічний університет, вул. Хмельницьке шосе 95, м. Вінниця, Україна, 21021, тел.: (0432) 598243, uhim@vstu.vinnica.ua

Савчук Тамара Олександрівна – к.т.н., доцент кафедри інтелектуальних систем, Вінницький національний технічний університет, вул. Хмельницьке шосе 95, м. Вінниця, Україна, 21021, тел.: (0432)598488, savchtam@vstu.vinnica.ua.

Вознюк Андрій Вікторович – магістрант кафедри інтелектуальних систем, Вінницький національний технічний університет, вул. Хмельницьке шосе 95, м. Вінниця, Україна, 21021, kenny@ukr.net.