Як видно з рисунків, для рівномірного закону розподілу ймовірностей звертання до сторінок значення математичного сподівання загального часу, необхідного для пошуку сторінки, є найбільшим. Математичне сподівання зменшується із зміною закону розподілу від узагальненого до закону Зіпфа. І є найменшим для "бінарного" закону розподілу ймовірностей звертання до сторінок. Порівнюючи за ефективністю обидва підходи можна зробити висновок, що ефективність другого методу є вища.

1. Кнут Д. Искусство программирования для ЭВМ. Т.3: Сортировка и поиск. – М.:Издательский дом " Вильямс ", 2000. – 840с. 2. Цегелик Г.Г. Организация и поиск информации данных. – Львов: Свит, 1990. – 186 с. 3. Цегелик Г.Г., Тичковський Р.О. Математичне моделювання оптимального доступу до інформації серверів зі сторони користувачів // Міжвідомчий збірник наукових праць "Відбір і обробка інформації" / Фізико-механічний інститут ім. Г.В. Карпенка.2004. Вип 21. – С.196–200. 4. Baeza-Yates R., Castilio C. Relating Web Structure and User Search Behavior. – Center for Web Research, Department of Computer Science, University of Chile, 2002. – 24p. 5. Hu W.-C., Chen Y., Smalz M., Ritter G. An Owerview of World Wide Web Search Technologies. Department of Computer Science. Auburn University, 2000, – 6p. 6. Kobayashi M., Takeda K. Information Retrieval on the Web. IBM Research, IBM Tokyo Research Laboratory. IBM Japan. 2000, – 47 p.

УДК 681.518:681.327.8

Л.В. Чирун, Т.В. Шестакевич
Національний університет "Львівська політехніка",
кафедра інформаційних систем та мереж

ІНТЕЛЕКТУЛЬНИЙ АНАЛІЗ ТАБЛИЦЬ ПРИЙНЯТТЯ РІШЕНЬ У СИСТЕМАХ ЕЛЕКТРОННОЇ КОМЕРЦІЇ

Проаналізовано основні проблеми електронної комерції в сфері видавництва та запропоновано методи вирішення цих проблем.

In the given article main problems of electronic commerce are analyzed. New methods for solution of discussed problems are proposed.

Вступ. Загальна постановка проблеми

Використання глобальних мереж зв'язку привело до появи нових напрямків ведення бізнесу та принципово змінило функціонування та структуру існуючих компаній, з'явилось поняття Інтернет-економіки. Внутрішня організація компанії на базі єдиної інформаційної мережі (Інтранет), яка підвищує ефективність взаємодії співробітників та оптимізує процеси планування і керування, а також зовнішня взаємодія (Екстранет) з партнерами, постачальниками і клієнтами – є складовими частинами електронного бізнесу (е-бізнесу). Одним із найважливіших складників електронного бізнесу є електронна комерція – будь-які форми ділової угоди, що проводиться за допомогою інформаційних мереж [1, 2].

До електронної комерції у широкому розумінні належать [1, 2]:

- глобальний електронний маркетинг, зокрема просування традиційних товарів і послуг;
- віддалені послуги, які можуть проводитися на відстані: послуги, пов'язані з консультуванням, юридичною і бухгалтерською підтримкою й ін.;

• дистанційна робота коли у сфері нематеріального виробництва стає можливою організація "розподілених офісів", у яких спільно працюють люди, що знаходяться в різних приміщеннях, містах і навіть країнах;

• електронну комерцію у вузькому розумінні, яка передбачає торгівлю товарами, що можуть передаватися в цифровій формі і (або) оплата яких може бути в цифровій формі. До таких товарів належить інформація в текстовій, графічній або звуковій формі; електронного бізнесу.

Сучасний етап розвитку Інтернет-економіки обумовив зростання потреб в інформації, яка тепер відіграє роль виробничого фактора та стратегічного ресурсу. Необхідність оперативного поширення та отримання інформації привела до розвитку електронних засобів масової інформації.

Зв'язок висвітленої проблеми із важливими науковими та практичними завданнями. Наукова новизна одержаних результатів

Ринок інформаційних послуг являє собою сукупність економічних, правових, організаційних і програмних відносин з продажу і купівлі інформаційних продуктів та послуг, які складаються між їхніми постачальниками і споживачами. Інформаційний продукт – це документована інформація, яка підготовлена відповідно до потреб користувачів Інтернет-послуг і призначена (або застосовується) для їх задоволення. Інформаційними послугами називають дії суб'єктів щодо забезпечення споживачів інформаційними продуктами. Одним із елементів ринку інформаційних послуг є електронні засоби масової інформації.

Інтернет-ЗМІ – це відвідувані значною аудиторією великі сайти, що поновлюються кілька разів за добу і створені для надання саме журналістської продукції, соціально значимої інформації: новин, статей тощо.

Електронне поширення інформації має такі переваги:

1. Низька вартість створення інформації – тобто допоміжних робіт при її створенні, що дозволяє заощадити ресурси власне для виробництва самої інформації, немає потреби у спеціальному устаткуванні;

2. Низька вартість тиражування інформації і відсутність надлишкового тиражування: документ фізично знаходиться в одному місці, "тиражується" і поширюється тільки адреса документа;

3. Простота і низька вартість опрацювання інформації (сортування, перетворення) – оскільки інформація знаходиться в електронному вигляді, її легко опрацьовувати як постачальникам інформації, так і кінцевим користувачам;

4. Відсутність обмежень обсягу – в Інтернеті обсяг публікацій не визначається обсягом друкованих площ, що волю мережному журналісту;

5. Екстериторіальність – оскільки носій несуттєвий, його не потрібно поширювати, будь-який сайт доступний там, де є доступ до Інтернету.

Практичне значення одержаних результатів

Інтернет-газета – це засіб масової інформації [1], спрямований на широку аудиторію, загальнодоступний, з корпоративним характером виробництва і поширенням інформації, містить багато видів комунікацій. З 2005 року функціонує Інтернет-газета "Прес-Тайм" (www.presstime.com.ua), до основних рубрик якої належать загальні новини, політика та суспільство, економіка. Своєю чергою, всі новини відсортовані за обласними центрами України: Львівська, Івано-Франківська, Закарпатська, Волинська, Рівненська, Тернопільська, Чернівецька і т.д. Газета має три основні підсистеми, які між собою взаємодіють: підсистема для редагування газети з боку журналістів; підсистема перегляду новин з боку клієнта; підсистема передплати. Клієнт (майбутній читач Інтернет-газети "Прес-тайм") укладає договір з видавництвом, в якому вказує основі рубрики, які його цікавлять, за якими областями отримувати інформацію, період отримання передплати, за бажанням вибирає, чи буде автоматично надсилатися сервером раз на добу інформація на електронну пошту. Клієнт отримує логін та виставляє пароль при першому

відвідуванні Інтернет-газети. Уся інформація про вибір користувача зберігається у клієнтській базі даних. Аналіз потреб передплатників дозволить покращити якість не лише видання, але й послуг, що надає видавництво.

Аналіз сучасних досліджень і публікацій

Інтелектуальний аналіз даних – складова частина процесу видобування знань з баз даних, що дає змогу розкрити суть прихованих залежностей у даних, виявити взаємні впливи між властивостями об'єктів, інформація про які зберігається в базах даних, виділити закономірності, властиві певному набору даних.

Загальна схема інтелектуального аналізу даних. Проблема відсутніх даних

Актуальність проблеми дослідження та опрацювання даних підтверджується широким практичним та комерційним використанням систем інтелектуального аналізу. Найчастіше їх застосовують у науковій сфері та бізнесі.

У загальному випадку процес видобування знань складається з чотирьох основних кроків.

1. Відбирання даних.
2. Попереднє опрацювання даних.
3. Інтелектуальний аналіз даних.
4. Оцінювання та інтерпретація побудованих моделей та знайдених залежностей.

Послідовність етапів [3] видобування знань зображено на рис. 1.

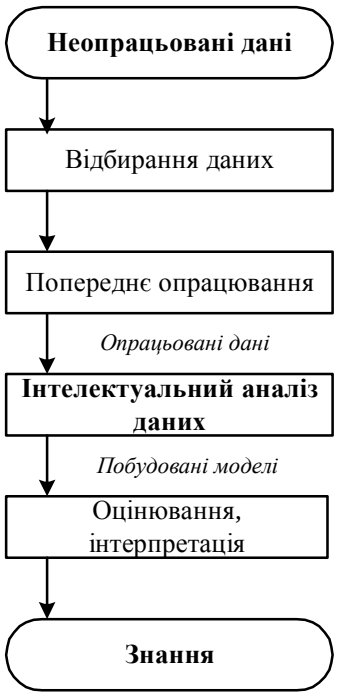

Рис. 1. Загальна схема процесу видобування знань

Неопрацьовані дані – це довільна інформація про досліджувану предметну область. Об'єкти предметної області описують множинами їхніх властивостей. Найзручніше подавати інформацію про властивості об'єктів таблицями, стовпці яких позначені іменами властивостей, а елементи рядків містять значення властивостей. Рядок таблиці в термінах машинного навчання є *прикладом*, стовпці таблиці називають *атрибутами*. Множина значень атрибута називається *доменом*. Якщо у таблиці визначено *атрибут прийняття рішення* (його значення вказує на належність прикладу до певного класу), то така таблиця є *таблицею прийняття рішень*. Тоді всі атрибути, крім атрибутів прийняття рішень, називають *умовними атрибутами*. Для цього дослідження предметною областю

є клієнтська база Інтернет-газети "Прес-тайм", де атрибутами таблиці прийняття рішень є адреса клієнта, рубрики, регіони, термін передплати тощо (усього 34 атрибути). Атрибутом прийняття рішень є висновок про зміну штату журналістів (1 атрибут). Усього в таблиці є 230 прикладів.

Проблема відсутніх даних у таблицях та невідомих значень атрибутів у прикладах з'являється тоді, коли хоча б одне значення атрибуту невідоме. Під невідомим розуміємо таке значення атрибута, визначити яке вже немає можливості, оскільки неможливо умови, у яких були отримані всі інші дані у таблиці, неможливо або дуже дорого повторити. Таке значення може бути довизначене на основі певних міркувань, яким присвячені подальші дослідження.

Причини появи у таблицях невідомих значень атрибутів є такими [4, 5].

1. Недбалість осіб, що збирають або вносять дані у таблиці, спричинена особистими рисами або відсутністю фінансової зацікавленості.
2. Зміна множини атрибутів у процесі збирання даних.
3. Надходження даних з різних джерел, у яких об'єкти описані різними множинами атрибутів.
4. Фізична відсутність даних. Наприклад, особа, яка не отримала водійських прав, не має запису про серію та номер посвідчення водія.
5. Логічна відсутність даних. Наприклад, керівник підприємства не може вказати в анкеті прізвище свого начальника.
6. Помилки вимірювань та обмежені можливості апаратури.
7. Значення атрибута не належить допустимій множині його значень.
8. Необхідність дотримання анонімності.

Серед зазначених причин появи відсутніх даних в описаній клієнтській базі, що представлена таблицею прийняття рішень, можна віднести, по-перше, недбалість осіб, що заповнюють форму замовлення, по-друге, зміну множини атрибутів у процесі збирання даних, а також небажання розголошувати індивідуальні дані.

Приклади в таблицях прийняття рішень можуть мати невідомі значення як умовних атрибутів, так і атрибута прийняття рішення. Надалі розглядатимемо приклади, у яких можуть бути невідомими лише значення умовних атрибутів.

Для інтелектуального аналізу таблиць даних, атрибути в яких мають невідомі значення, планується застосувати порівняно новий підхід, який ґрунтується на понятті *наближеної множини* (rough set) [6]. Наближені множини – це символьна індуктивна методологія, яка поряд з нейронними мережами, розмитими множинами, генетичними алгоритмами належить до методологій *м'яких обчислень* (soft computing), які застосовують в інтелектуальному аналізі даних та машинному навчанні.

Способи вирішення проблеми відсутніх даних

Приклади, які є описом об'єктів предметних областей та на основі яких доводиться приймати рішення, у часто містять невідомі значення атрибутів. У разі побудови систем прийняття рішень доводиться враховувати і такі дані. Це пов'язано з тим, що здійснення додаткових досліджень з метою покращення даних неможливе або вартісне.

Виділяють [7] такі основні групи методів опрацювання таблиць із невідомими значеннями атрибутів:

– іґнорування відсутніх даних;

– видалення прикладів із невідомими значеннями атрибутів;

– доповнення відсутніх даних;

– безпосереднє опрацювання таблиць з відсутніми даними.

На рис. 2 перелічено методи опрацювання таблиць із невідомими значеннями атрибутів.

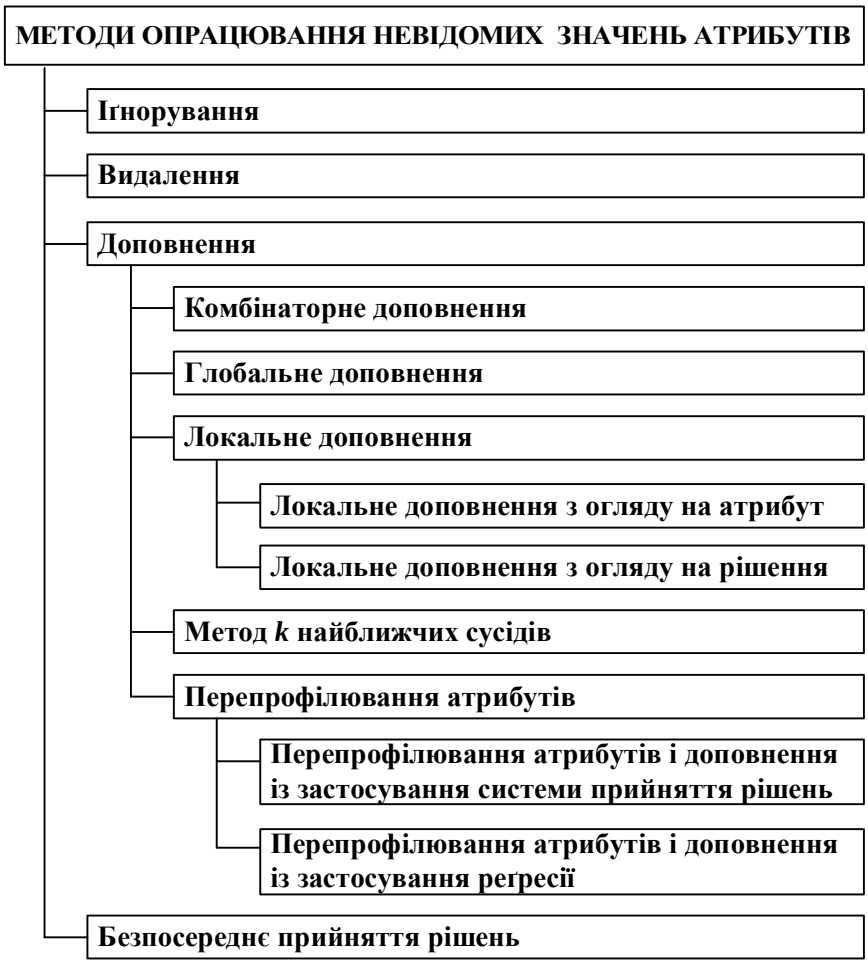

Рис. 1. Класифікація методів опрацювання відсутніх даних

Ігнорування відсутніх даних полягає у доповненні переліку значень атрибута величиною, яка символізує невідоме значення цього атрибута.

Видалення прикладів або атрибутів. Використовують два підходи до видалення даних. Перший з них здійснює спеціаліст, який, з огляду на свої знання та досвід, приймає рішення про видалення прикладу чи атрибута (часткове видалення). Також обсяги та пропорції видалення прикладів із невідомими значеннями атрибутів залежать від конкретних даних та задач. Такий підхід не є алгоритмічним, оскільки прийняття рішення про видалення залежить від досвіду людини-експерта.

Другий підхід до видалення даних можна назвати автоматичним: допускається видалення кожного атрибута і кожного прикладу, якщо вони містять хоча б одне невідоме значення (повне видалення). У результаті такого видалення у таблиці будуть залишені лише заповнені рядки.

І повне, і часткове вилучення прикладів чи атрибутів із невідомими значеннями не виключають видалення і таких, що мають суттєві властивості для досліджуваної таблиці прийняття рішень. Вилучення прикладів чи атрибутів, що містять навіть одне невідоме значення, може істотно зменшити розмірність таблиці.

Доповнення таблиць даних [7]. Невідомі значення атрибута заповнюють за певним критерієм, який формують на основі відомих значень атрибута. У разі доповнення таблиці необхідно розрізняти дані, що об’єктивно існують, та такі, що не існують. До перших належать дані, які можна отримати, але вони з певних причин не були внесені до таблиці (наприклад, інформація про вік працівника – її можна доповнити на основі інших відомих даних). Доповнення таблиць із відсутніми даними не змінює розмірності таблиці, але вносить інформаційний шум у дані.

Універсальні методи доповнення таблиць з відсутніми даними дають змогу застосовувати відомі методи опрацювання заповнених таблиць прийняття рішень.

Існують такі основні методи доповнення таблиць з відсутніми даними:

- комбінаторне доповнення;
- глобальне доповнення;
- локальне доповнення з огляду на атрибут та на рішення;
- доповнення методом k найближчих сусідів;
- перепрофілювання змінних і використання системи прийняття рішень.

Комбінаторне доповнення. Метод комбінаторного доповнення дозволяє доповнити таблицю заміною прикладу із невідомим значенням атрибуту кількома прикладами із усіма відомими значеннями атрибутів.

Кількість додаткових прикладів, що утвориться застосуванням методу комбінаторного доповнення, обчислюють за формулою

$$F = \sum_{i=1}^{n} \left(\prod_{j=1}^{m} z_{ij} - 1 \right),$$

де n – кількість прикладів у таблиці, m – кількість атрибутів таблиці, z_{ij} дорівнює 1, якщо значення i-го прикладу на j-му атрибуті відоме, і потужності домену j-го атрибута, якщо невідоме.

Обмеженням на застосування методу комбінаторного доповнення є велика кількість невідомих значень атрибутів і (або) велика потужність доменів атрибутів, значення яких невідомі.

Глобальне доповнення. Таке доповнення використовують для заповнення відсутніх даних на основі відомих значень атрибутів. Для цього на основі усіх відомих значень атрибута обчислюють певний параметр s . Значенням параметра s для чисельних атрибутів може бути середнє або медіана, для символьних атрибутів – значення, що зустрічається найчастіше. Обчисленим параметром заміняють відсутні значення.

Локальне доповнення з огляду на рішення. Метод передбачає поділ множини прикладів таблиці на підмножини з однаковим значенням атрибута прийняття рішення. Для кожної підмножини обчислюється власний параметр s та ним заповнюються невідомі значення.

Локальне доповнення з огляду на атрибут. Метод розглядає умовні атрибути із відсутніми значеннями як атрибути прийняття рішення. Пошук пов’язаних між собою атрибутів ускладнений необхідністю оцінювати зв’язки між атрибутами не лише однакового, але й різного типу. Для оцінювання міри взаємозв’язку пари числових атрибутів можна використати коефіцієнт кореляції, а двох символьних атрибутів – ентропію. Проте, немає ефективного методу порівняння між собою числових та символьних атрибутів.

Доповнення за допомогою методу k найближчих сусідів. Метод передбачає, що приклади з близькими значеннями одних атрибутів найімовірніше мають близькі значення й на інших атрибутах. Метод k *найближчих сусідів враховує* подібність між прикладами, тоді як попередні методи доповнення невідомих значень спирались на існування залежностей між атрибутами.

У разі перепрофілювання атрибутів і доповнення із застосуванням системи прийняття рішень чи регресії умовні атрибути з невідомими значеннями розглядають як атрибут прийняття рішень. Існування зв’язків між невідомими та відомими значеннями можна використати для доповнення невідомих значень атрибута з допомогою регресійного аналізу [11].

Доповнення невідомих значень є універсальним способом розв'язування задачі про неповний опис об'єктів. Водночас доповнення невідомих даних має небезпеку внесення істотних змін у дані, що ускладнює пошук зв'язків між умовними атрибутами та розв’язком.

Безпосереднє прийняття рішень на основі даних з відсутніми значеннями. Одним із способів безпосереднього опрацювання даних з невідомими значеннями є методи поділу, за допомогою яких таблицю прийняття рішень з відсутніми даними поділяють так, щоб утворити заповнені таблиці. Такі таблиці опрацьовують методами для заповнених таблиць.

Існують також методи опрацювання прикладів з невідомими значеннями атрибутів, які полягають у прийнятті рішень безпосередньо на основі неповних даних. Такий підхід реалізовано в алгоритмах MLEM2 [10], URG-2 [11]. Негативний аспект такого підходу полягає у необхідності індивідуального налаштування алгоритму до заданого набору даних.

Математичний апарат теорії наближених множин також забезпечує механізми безпосередньої роботи з відсутніми даними [9]. Фундаментальний принцип алгоритму навчання на прикладах з використанням наближених множин полягає у виявленні надлишковості серед наявних ознак, що описують приклад. Так виявляють сильні ознаки, які впливають на класифікацію прикладу. Далі вилучають стовпці, атрибути яких не впливають на класифікацію. Для аналізу залишають лише такі атрибути, від яких залежить класифікація прикладів таблиці. Множину атрибутів, що залишилися, називають *редуктом*. Іншими словами, редукт – це підмножина усіх атрибутів таблиці, які забезпечують той самий результат класифікації всіх прикладів таблиці, як і вся множина прикладів. Однак існують достатньо ефективні методи, які дають змогу розв'язувати цю задачу за прийнятний час. До таких методів належать, зокрема, *логічне виведення* (boolean reasoning) [6], алгоритм Джонсона [10].

Виділення проблем

Контент-комерція – придбання або продаж інформації (контенту) за допомогою мережі та електронних носіїв. *Система електронної контент-комерції* – організація та технологія купівлі/продажу контенту з використанням мереж та електронних фінансово-економічних інструментів [1]. Система електронної контент-комерції реалізована на прикладі Інтернет-газети "Прес-Тайм", яка має три підсистеми: редагування газети журналістами, перегляд новин клієнтами та передплата. Інтернет-газета розташована за адресою www.presstime.com.ua. До основних рубрик газети належать: загальні новини, політика та суспільство, економіка. Статті поділені за дванадцятьма регіонами України (Львівський, Закарпатський, Чернівецький, Рівненський тощо). Клієнт-передплатник укладає угоду із зазначенням регіонів та рубрик, статті з яких хоче передплачувати, періоду передплати і т.д. З огляду на потреби кожного передплатника редактор приймає рішення про штат журналістів, тобто виникає потреба аналізу існуючої клієнтської бази для визначення залежностей, що істотно впливають на зміну штату журналістів.

У статті наведено результати опрацювання даних клієнтської бази Інтернет-газети "Прес-Тайм". Клієнтську базу вивчає редактор, який оцінює потребу в збільшенні або зменшенні штату журналістів у відповідних регіонах, які охоплює Інтернет-газета. Основними недоліками даних клієнтської бази є відсутність частини даних, а також суб'єктивність оцінок, зроблених редактором на основі власного досвіду. Накопичені дані перед здійсненням аналізу потребують додаткового опрацювання. Метою досліджень клієнтської бази є виявлення характерних особливостей видання, які впливають на зміну штату журналістів. За результатами досліджень можна вдосконалити та прискорити процес прийняття рішень редактором, виявити такі елементи структури Інтернет-газети, яким належить приділяти увагу при удосконаленні чи модернізації структури та змісту газети.

Формування цілей (постановка задачі)

У результаті проведеного дослідження вирішено проаналізувати існуючу клієнтську базу Інтернет-газети "Прес-Тайм" для визначення залежностей, що істотно впливають на зміну штату журналістів. Для цього було відібрано відповідні дані та здійснено їх опис. Опис наявних даних здійснено також з метою уточнення подальших завдань аналізу. З усього набору даних клієнтської бази сформовано таблицю прийняття рішень.

Необхідно проаналізувати клієнтську базу Інтернет-газети „Прес-тайм". Клієнти укладають із газетою угоду, в якій зазначають тематику статей, інтерв'ю, поточних новин та реклами (загальні, політика та суспільство, економіка), вказують, з якого регіону постачати інформацію, обирають форму отримання газети, термін передплати, потребу доступу до архіву видання, також зазначають форму оплати за видання. Кожен клієнт в угоді вказує також свою адресу, назву отримувача та логін, за яким його розрізняють в клієнтській базі.

Аналіз отриманих наукових результатів

Разом опрацьовано та зібрано в таблицю дані про 230 осіб, що уклали угоди. Таблиця даних містить 230 прикладів та 35 атрибутів, тобто загальна кількість значень в таблиці із врахуванням невідомих значень становить 8050. Серед атрибутів є 34 умовні атрибути та один атрибут прийняття рішень. За своєю структурою множину умовних атрибутів можна розбити на три підмножини за принципом заповнення значень таблиці. До першої підмножини належить 21 атрибут, значення яких заповнюється автоматично, навіть якщо клієнт не зробив вибір. Ці атрибути описують тематику статей (три атрибути), регіон (дванадцять атрибутів), форму доставки інформації (три атрибути), форму оплати, термін передплати та логін клієнта – по одному атрибуту. До другої підмножини умовних атрибутів належать одинадцять атрибутів, що можуть містити відсутні дані, якщо клієнт не зробив свого вибору. Це атрибути, що описують термін доступу до архіву статей (один атрибут), новини (три атрибути), інтерв'ю (три атрибути) та рекламу (чотири атрибути). Третя підмножина умовних атрибутів містить два атрибути, значення яких можуть бути або порожніми, або довільними. Це адреса клієнта та назва отримувача передплати. Атрибут прийняття рішення містить висновок про те, чи доцільно збільшувати або зменшувати штат журналістів у кожному регіоні. Структура набору даних зображена на рис. 3.

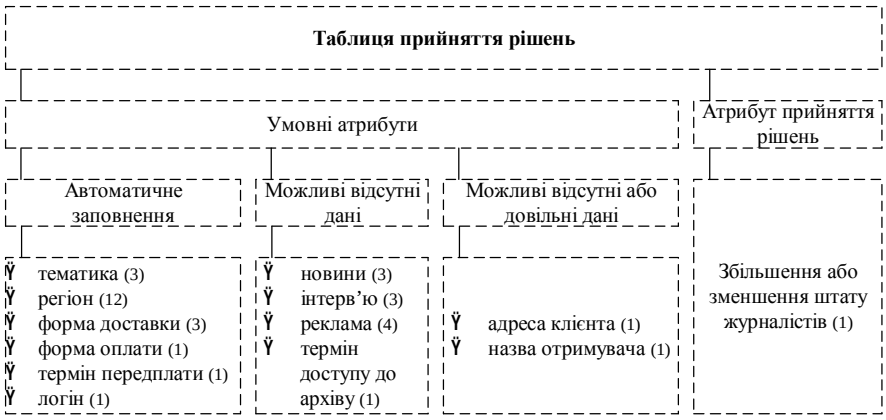

Рис. 3. Структура таблиці даних

Недоліки даних. Описані дані мають кілька недоліків.

По-перше, існує певна надлишковість даних. Деякі атрибути містять неістотну для аналізу інформацію, наприклад, адресу клієнта чи логін. На етапі відбирання даних необхідно вилучити атрибути, що містять надлишкову інформацію.

По-друге, у таблиці даних невідомі деякі значення атрибутів, через що незастосовні методи інтелектуального аналізу даних, які працюють лише із заповненими таблицями.

Нарешті, загальним недоліком прийнятих рішень є їх суб'єктивність. Власне, суб'єктивність прийнятих рішень вимагає розв'язування задачі знаходження залежностей у даних та атрибутів, які реально впливають на прийняття рішень. Знаходження таких атрибутів дозволить вдосконалити форму замовлення та саму структуру Інтернет-газети „Прес-тайм" у майбутньому.

Розглянемо детально етап відбирання даних, що відповідає загальному процесу видобування знань, зображеному на рис. 1.

Етап відбирання даних. Відбирання даних полягає у видаленні атрибутів таблиці, не важливих для аналізу. Під час відбирання даних було вилучено атрибути „адреса клієнта", „назва отримувача" та „логін". З вилученням вказаних атрибутів змінилася і структура таблиці (див. рис. 4).

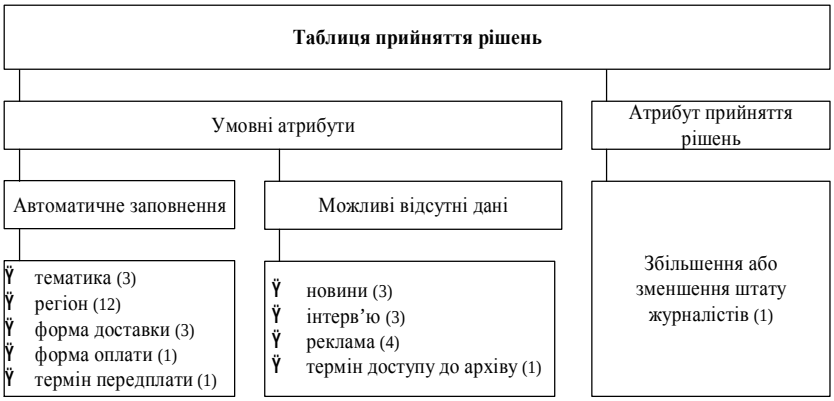

Рис. 4. Структура таблиці після етапу відбирання даних

Після відбирання даних загальна кількість атрибутів таблиці зменшена, і разом з атрибутом прийняття рішення становить 32 замість 35. Отже, отримана таблиця даних має 230 прикладів та 32 атрибути.

Наступним є етап попереднього опрацювання даних, у процесі якого вилучають зайві дані, замінюють типи даних, кодують окремі значення, опрацьовують відсутні дані тощо. Для подальшого дослідження необхідно здійснити інтелектуальний аналіз даних з метою пошуку прихованих залежностей та оцінити ці залежності. Для аналізу даних та виявлення атрибутів, які найбільше впливають на прийняте рішення, пропонується використати теорію наближених множин та алгоритм Джонсона, реалізовані в системі ROSETTA [8]. Серед даних, які залишились після етапу відбирання даних з початкової таблиці, будуть приклади із невідомими значеннями атрибутів. Необхідність опрацювання таких неповних таблиць даних зумовлює здійснення кількох серій експериментів над таблицями з різними способами опрацювання невідомих значень атрибутів для порівняння якості підходів опрацювання відсутніх даних.

Висновки і перспективи подальших наукових розвідок

Автори розглядають особливості електронної комерції та місце Інтернет-ЗМІ на ринку інформаційних послуг, переваги функціонування видань в електронній формі. Описано предметну область, що відповідає клієнтській базі Інтернет-газети "Прес-Тайм"; згідно з потребами аналізу сформовано таблицю прийняття рішень. Дані таблиці опрацьовано згідно з вимогами алгоритму, з допомогою якого здійснюватиметься інтелектуальний аналіз. Наступним є етап попереднього опрацювання даних, у процесі якого вилучають зайві дані, замінюють типи даних, кодують окремі значення, опрацьовують відсутні дані тощо. Для подальшого дослідження необхідно здійснити інтелектуальний аналіз даних з метою пошуку прихованих залежностей та оцінити ці залежності. Для аналізу даних та виявлення атрибутів, які найбільше впливають на прийняте рішення, пропонується використати теорію наближених множин та алгоритм Джонсона, реалізовані в системі ROSETTA. Серед даних, які залишились після етапу відбирання даних з початкової таблиці, будуть приклади із невідомими значеннями атрибутів. Необхідність опрацювання таких неповних таблиць даних зумовлює здійснення кількох серій експериментів над таблицями з різними способами опрацювання невідомих значень атрибутів для порівняння якості підходів опрацювання відсутніх даних.

1. Берко А.Ю., Висоцька В.А., Чирун Л.В. Алгоритми опрацювання інформаційних ресурсів в системах електронної комерції // Вісник Нац. ун-ту "Львівська політехніка" Інформаційні системи та мережі. – 2004. – № 519. – С.10–20. 2. Берко А.Ю., Висоцька В.А. Проектування навігаційного графу web-сторінок бази даних систем електронної комерції // Вісник Нац. ун-ту "Львівська політехніка" Комп'ютерні науки та інформаційні технології. – 2004. – № 521. – С.48–57. 3. Mitra

Sushmita, Pal Sankar K., Mitra Pabitra. Data mining in soft computing framework: a survey, IEEE Transactions on Neural Networks, Vol. 13, Issue 1, 2002. 4. Acuna E., Rodriguez C. The Treatment of Missing Values and its Effect in the Classifier Accuracy. // In D. Banks, L. House, F.R. McMorris, P. Arabie, W. Gaul (Eds), Classification, Clustering and Data Mining Applications, Springer-Verlag Berlin-Heidelberg, 2004, c.639-648. 5. Grzymala-Busse J.W. Rough Set Strategies to Data with Missing Attribute Values. // Proceedings of the Workshop on Foundation and New Directions in Data Mining, associated with the third IEEE International Conference on Data Mining, November 19-22, 2003, Melbourne, FL, USA, c.56-63. 6. Jan Komorowski, Lech Polkowski, Andrzej Skowron, Rough Sets: A Tutorial. // Eds. S.K.Pal and A. Skowron, Rough Fuzzy Hybridization: A New Trend in Decision-Making, Spriner-Verlag, Singapore, 1998. 7. Latkowski R. Metody wnioskowania w oparciu o niekompletny opis obiektow, Warszawa, 2001. http://logic.mimuw.edu.pl/Grant2003/prace/BMscLatkowski.pdf. 8. Øhrn A. ROSETTA Technical Reference Manual, 2001. http://www.idi.ntnu.no/~aleks/. 9. Pawlak Z. Rough Sets – Theoretical Aspects of Reasoning about Data, volume 9 of Series D: System Theory, Knowledge Engineering and Problem Solving. Kluwer Academic Publishers, Dordrecht, 1991. 10. Grzymala-Busse J.W. MLEM2 – Discretization During Rule Induction. // Proceedings of the IIPWM'2003, International Conference on Intelligent Information Processing and WEB Mining Systems, Zakopane, Poland, June 2-5, 2003, Springer Verlag, c.499-508. 11. Konias S., Chouvarda I., Vlahavas I., Maglaveras N. A Novel Approach for Incremental Uncertainty Rule Generation from Databases with Missing Value Handling: Application to Dynamic Medical Databases. Medical Informatics & The Internet in Medicine, *Taylor & Francis, Vol. 2005, Issue 5, 2005.*

УДК 621.396.6. 519.2

А.О. Левченко, О.І. Кравчук
Львівський інститут сухопутних військ ім. П. Сагайдачного

ПРОЦЕДУРА СИНТЕЗУ МОДЕЛЕЙ ПАРАМЕТРА ПОТОКУ ВІДМОВ ІНФОРМАЦІЙНО-ДОВІДКОВОЇ АВТОМАТИЗОВАНОЇ СИСТЕМИ ВИЗНАЧЕННЯ СТАНУ РАДІОЕЛЕКТРОННИХ ЗАСОБІВ ПІД ЧАС ОДНОРЕЖИМНОГО УТРИМАННЯ

Актуальність теми

Одним з першочергових завдань у процесі розбудови Збройних сил України стало створення сучасної високоефективної системи управління логістикою в межах Єдиної автоматизованої системи управління Збройних сил України. За умов зростання вартості новітніх видів озброєння та військової техніки (ОВТ) досягти й підтримувати потрібний рівень бойової могутності вигідніше не нарощуванням кількісного складу військ та ОВТ, а забезпеченням структурної цілісності, високого ступеня автоматизації військами та бойовими засобами з використанням сучасних новітніх інформаційних технологій, а також розроблення програмного й математичного забезпечення. Крім того, необхідно передбачити заходи, які забезпечували б подальшу експлуатацію існуючих засобів і систем автоматизованого управління, і зафіксувати фінансові, матеріальні та інші ресурси, необхідні для підтримання їх у боєздатному стані, подовження термінів експлуатації та доопрацювання (модернізації) з метою введення їх до Єдиної системи управління. Для управління технологічними процесами зберігання, обслуговування та відновлення стану складних технічних засобів впроваджуються інформаційно-довідкові автоматизовані системи (ІДАС) підтримки прийняття рішень у межах Єдиної системи управління логістики,