Acoustics, Speech and signal Processing. – October 1977. – Vol. ASSP-25, No. 5. – P. 423–428. 3. Vich R. and Smekal Z. Continued Fractions in Digital Filter Synthesis, Proc. of Inter. Scient. Colloquium. – Ilmenau, Germany, 18–21 September 1995. – P. 353–356. 4. Vich R. and Smekal Z. Digital Filter Realization of Nonrational Transfer Functions, Proc. of the First European Conference on Signal Analysis and Prediction, ECSAP-97. – Prague, Czech Republic, 24–27 June 1997. – P. 179–182. 5. Шмойлов В.И. Периодические цепные дроби. – Львов: Академический Экспресс, 1998. – 219 с. 6. Шмойлов В.И., Слобода М.З. Расходящиеся непрерывные дроби. – Львов: Меркатор, 1999. – 820 с. 7. Шмойлов В.И., Чирун Л.В. Комплексные числа и непрерывные дроби. – Львов: Меркатор, 2001. – 564 с. 8. Strum R.D. and Kirk D.E. First Principles of Discrete Systems and Digital Signal Processing. – Massachusetts; Addison-Wesley Publishing Company, 1988. 9. Ваврук Є.Я., Рашкевич Ю.М. Особливості реалізації пристроїв обміну інформацією в системах цифрової обробки сигналів // Моделювання та інформаційні технології: Зб. наук. праць ІПМЕ НАНУ. – 1999. – Вип. 4. – С. 119–123. 10. Ваврук Є.Я, Рашкевич Ю.М., Цмоць І.Г. Оцінка основних характеристик процесорів управління та обробки інформації на НВІС // Вісн. Держ. ун-ту "Львівська політехніка". – 1999. – № 386. – С. 5–11. 11. Рашкевич Ю. М. Перетворення часового масштабу мовних сигналів. – Львів: ТзОВ НВТ "Акад. Експрес", 1997. – 140 с. 12. Ваврук Є.Я., Рашкевич Ю.М., Цмоць І.Г. Підходи до побудови та вибору елементної бази процесорів управління та обробки сигналів // Моделювання та інформаційні технології: Зб. наук. праць ІПМЕ НАНУ. – 1999. – Вип. 3. – С. 160–168.

УДК 681.3

Н.Б. Шаховська
Національний університет "Львівська політехніка",
кафедра "Інформаційні системи та мережі"

МЕТОДИ УСУНЕННЯ НЕВИЗНАЧЕНОСТЕЙ У БАЗАХ ЗНАНЬ, ПОБУДОВАНИХ НА ОСНОВІ РЕЛЯЦІЙНОГО ПІДХОДУ

The elimination indistinct algorithmes in knowledge bases, build on relation schema, are described. The efficient gathering of sufficient statistics for classification from knowledge bases are described.

Запропоновано алгоритми усунення невизначеностей у базах знань, побудованих на основі реляційного підходу. Процес усунення невизначеностей інтерпретовано як класифікування об'єктів. Проаналізовано методи аналізу об'єктів з ієрархічною структурою.

1. ВСТУП

Як відомо, до бази даних (БД) часто формуються запити з нечітко заданими параметрами, поданими у вигляді інтервалів, лінгвістичних змінних, ступенів довіри тощо.

Крім того, усе частіше виникає проблема збереження у відношеннях реляційної БД інформації про об'єкти, які за своєю суттю є нечіткими. Прикладом таких об'єктів є різноманітні класи, причому класифікаційні ознаки можуть бути як чіткими (наприклад, загальноприйнята вікова градація *Молодий – до 28 років, Зрілий – від 28 до 40 років, Старий – 40 років і старший*), так і заданими певним розподілом. Через складність інформаційного моделювання предметних областей стало неможливим відображати такі об'єкти у традиційній реляційній моделі даних, і, як наслідок, виникли постреляційні моделі (об'єктно-орієнтовані моделі та моделі з гніздуванням).

Типовими предметними областями, у яких постає задача опрацювання (усунення) невизначених та нечітко заданих значень, є: задачі планування, біржа працевлаштування, соціологічні опитування, історичні дослідження тощо.

2. ОГЛЯД СУЧАСНИХ ДОСЛІДЖЕНЬ

Проблемою подання об'єктів із стохастичною компонентою у базах даних та знань, побудованих на основі реляційного підходу, займались Дюбуа, Кодд та ін. [7], а також ряд сучасних дослідників [1 – 4]. У представленні невизначеностей поширені такі підходи:

- Зазначення факту, що значення є невідомим (нечітким), але накопичення факту невизначеності є неможливим ані з інформативної точки зору, ані з метою подальшого аналізу.
- Панує об'єктно-орієнтований підхід, за яким допускається факт незнання про об'єкт, але нема можливості представлення нечіткості як характеристики об'єкта, тобто на рівні об'єкта нечіткість не закладається (використання апарата нечітких множин, лінгвістичних змінних тощо).

Іншою задачею, яка виникає у процесі роботи з інформаційними моделями об'єктів предметної області, є опрацювання невизначеності з метою її зменшення або й повного усунення. Зважаючи на структурованість БД, моделі, які описують залежності між даними, доцільно подавати у вигляді правил [1, 2]. Такий підхід використовується для вирішення задач пошуку асоціативних зв'язків, класифікування, прогнозування тощо [3]. Очевидно, що залежності між даними мають стохастичний характер, а тому побудовані правила описуватимуться кількісними характеристиками, такими як ступінь підтримки, точність та повнота, які є оцінками відповідних імовірностей [4].

У статті пропонується метод зменшення невизначеності у відношенні (або відношеннях) БД за допомогою модифікованого алгоритму прогонки chase [6] та з врахуванням зв'язків, які виникають між кортежами відношень (тобто на рівні описів об'єктів спорідненої структури).

3. ПОСТАНОВКА ЗАДАЧІ УСУНЕННЯ НЕВИЗНАЧЕНОСТІ

У [1] описано 14 видів невизначеностей, які можуть виникати у процесі опрацювання даних, проте вони зводяться до таких:

1. Значення невідоме (відсутнє).
2. Неповнота інформації.
3. Нечіткість (використання розподілу для встановлення істинності знань).
4. Неточність (стосується числових даних).
5. Недетермінованість процедур виведення рішень (випадковість).
6. Ненадійність даних.
7. Багатозначність інтерпретацій.

8. Лінгвістична невизначність:
   a) невизначеність значення слова;
   b) невизначеність змісту речення.

Розглянемо детальніше вказані типи невизначеностей та виявимо місця їх появи у відношенні.

Невизначеності типів 3–8 класифікують у [1] як неоднозначність даних;переважно з'являються на рівні кортежа або підмножини значень атрибутів, з яких формується кортеж.

Відсутність інформації найчастіше зустрічається на рівні значення атрибута. Неповнота є станом кортежа, у якому є відсутні значення. Нечіткість, неточність та випадковість можна віднести до фізичної невизначеності, одним із джерел якої є обмеженість у точності числових типів даних або втрата точності під час виконання математичних операцій (сюди ж відносять невизначеність, яка виникає внаслідок роботи з інтервальними величинами). Ненадійність та багатозначність інтерпретацій виникає через неповне вивчення або неоднозначне відображення характеристик сутності у відношенні. У відношенні зображається за допомогою додаткового атрибута, значення якого характеризують міру довіри до цілого кортежа або підмножини значень атрибутів у кортежі. Багатозначність інтерпретації є одним із джерел виникнення протиріч.

Лінгвістична невизначеність пов'язана із використнням природної мови для представлення знань, які мають якісний характер, і може виникати внаслідок нерозуміння (незнання) значення слова або нерозуміння змісту речення. Такий тип невизначеності зустрічається у системах обробки текстової інформації (системи автоматизованого перекладу, системи для самонавчання тощо).

Розглянуті типи невизначеностей можуть накладатись один на одного або бути джерелом появи один одного.

У зв'язку зі складністюрозрізнення деяких типів невизначеності та їх опрацювання, реляційна модель дозволяє моделювати лише невідомі, неповні, неточні та нечіткі дані.

Традиційно усі види невизначеностей позначаються символом $\perp$. Оскільки доволі важко відчути семантичну різницю між деякими із наведених типів невизначеностей і ще важче відобразити певні типи невизначеностей у відношенні, то розглянемо такі типи невизначеностей та введемо для них відповідні позначення:

- відомо, що значення притаманне сутності, але ми його не знаємо (відсутність) – позначається символом $\perp$ (вводиться на рівні значення атрибута);
- є неповна або часткова інформація про значення, для відображення якої використовується додатковий атрибут, що характеризує рівень істинності даних та містить значення функцій розподілу, лінгвістичних змінних, ступенів істинності багатозначних логік (може вводитися на рівні значення атрибута, підмножини значень атрибутів або кортежа).

Усунення (зменшення) цих типів невизначеності можна здійснити такими методами:

1. Застосування алгоритмів видобування знань;
2. Аналіз залежностей між кортежами відношення.

У першому випадку невідоме значення атрибута можна розглядати як мітку класу [1], а сама задача усунення невизначеності трансформується у задачу віднесення до класу. Використання цього методу дозволяє усувати невизначеності типу "невідомий" та "неповнота" на рівні значення атрибута та підмножини атрибутів.

Другий метод дозволяє усувати частковість, багатозначність інтерпретацій та ненадійність даних шляхом використання об'єктно-орієнтованого підходу до моделювання сутностей та аналізу зв'язків між об'єктами та їх характеристиками.

Здійснимо формальну постановку описаних задач.

3.1. Віднесення до класу як метод усунення залежностей

Маємо деяке відношення r зі схемою R. З метою усунення невизначеності на основі відношення r необхідно побудувати множину класифікаційних правил виду $s(X \to Y)$, де $X, Y \subset R$, $X \cap Y = \varnothing$; X – підмножина атрибутів, на основі значень яких здійснюється віднесення до класу (усунення невизначеності за значеннями атрибута Y), Y – атрибут (підмножина атрибутів), значеннями якого є мітки класу.

Введемо основні поняття, що використовуються у задачі віднесення до класу.

Класифікаційним правилом назвемо залежність між підмножинами атрибутів X та Y, яка зустрічається у тестовому наборі відношення r зі ступенем відповідності (довіри) s, при якій $(X = x) \to (Y = y)$.

Класифікаційне правило можна розглядати як наближену (approximate) [4] функціональну залежність, яка підтримується у відношенні сховища даних, побудованого на основі реляційної моделі. Частковим випадком класифікаційного правила є *традиційна функціональна залежність*, коли ступінь довіри до правила набуває максимального значення (абсолютна довіра).

Будується класифікаційне правило на основі навчального набору даних у відношенні r, де значення міток класу (значення підмножини атрибутів Y) відомі.

Як математичний апарат для подання та опрацювання ступенів відповідності правила використовують:

- ймовірність;
- нечітку логіку;
- багатозначну логіку (лінгвістичні змінні).

Методами визначення ступеня довіри до правила є:

- експертні оцінювання;
- детерміновані методи (ймовірність, теорія шансів);
- статистичні залежності (припускається, що поява кортежів у відношенні є незалежною та підраховується кількість пар кортежів, між якими встановлено залежність);
- кластеризація та визначення віддалі до найбільшого скупчення даних (середньостатистичні відхилення, метод К-найближчого сусіда тощо).

Для визначення ступеня істинності правил оберемо традиційні для звичайних функціональних залежностей статистичні залежності (перевірка на істинність функціональної залежності здійснюється через порівняння пар кортежів).

Введемо поняття значущого правила.

Правило вважається *значущим* (дійсним), якщо його ступінь відповідності s не менший якогось порогового значення *UNK*, яке визначає загальну значущість навчального набору у відношенні r. Значення величин s та *UNK* лежать у межах [0, 1] і трактуються як ступені багатозначної логіки, тобто є дискретними. Це, у свою чергу, дозволяє не тільки застосовувати відповідний математичний апарат для опрацювання ступеня довіри, але й використовувати аксіоми виведення для самих правил.

Будемо шукати лише ті правила, які описують деякі закономірності (залежності) між атрибутами. Крім того, слід зазначити, що класифікаційні правила можуть будуватися експертами самостійно та заноситись у базу знань. У цьому випадку ступінь відповідності також визначається експертом, а саме правило опрацьовується так само, якби воно було згенероване на основі аналізу навчального набору у відношенні *r*.

Міткою класу назвемо лінгвістичну змінну або типову характеристику об'єктів, яка є значеннями підмножини атрибутів *Y* і позначає об'єкти зі спільними (подібними зі ступенем *s*) значеннями підмножини атрибутів *X*. Домени атрибутів, що належать до підмножини *Y*, $y \in dom(Y) = \pi_Y(r)$, обов'язково повинні містити скінченну та наперед відому множину значень.

Побудова правил здійснюється на основі навчальних даних з максимально допустимим для предметної області ступенем наповнення, причому мітки класу обираються з наперед відомої множини значень (у межах досліджуваної області є фіксованими), а віднесення до класу об'єктів, інформація про які щойно надійшла у сховище даних, здійснюється на основі класифікаційних правил.

Усунення невизначеностей та нечіткостей, які зустрічаються серед значень атрибута *Y* відношення *r*, є класифікуванням за відомими алгоритмом *chase* [6]. Тобто, якщо зустрічається кортеж, у якому значення атрибутів $X = x$, то експерту пропонується прийняти рішення, чи здійснювати заміну (запис) у відповідні значення атрибутів $Y = y$, причому істинність такої операції становитиме *s*.

Застосування класифікаційних правил для зменшення невизначеності порівняно з використанням лише традиційних функціональних залежностей дозволить:

- Моделювати особливості конкретної предметної області;
- Значно розширити коло невизначеностей, які можуть бути усунені (нагадаємо, що за допомогою функціональних залежностей усувалися лише відсутні значення);
- Використовуючи знання (досвід) експерта, видобувати нові знання.

Наведемо приклад класифікаційних правил, побудованих на основі деякого відношення *r*.

Ідентифікаційний код	Освіта	Вік	Стаж роботи	Професія	Величина міста	Кількість попередніх місць роботи	Оклад
0123456	вища	27	4	програміст	середнє	2	1700
2340988	вища	23	0	програміст	велике	0	2000
0347689	середня	22	0	оператор	середнє	0	250
0356762	вища	27	3	програміст	велике	1	⊥

На основі аналізу інформації з наведеного відношення можна побудувати такі класифікаційні правила (слід зазначити, що цей приклад є умовний і наведені правила не перевірені експериментально, а лише демонструють наведені у статті викладки):

Освіта, Вік,Стаж роботи, Професія→Оклад
Оклад, Професія→Величина міста

Застосовуючи вищенаведені правила, для особи з індексом 0356762 буде визначено оклад у розмірі від 1700 до 2000 одиниць.

3.2. Аналіз залежностей між кортежами відношень

Коли ми говоримо про невизначеності, які є характеристиками об'єкта, доцільно *застосувати об'єктно-орієнтований підхід до проектування схеми бази даних*. Мова йде про об'єкти, які описуються

- через перелічення їхніх складових або властивостей (наприклад, географічні об'єкти, виробничі процеси тощо);
- через вказання зв'язків з іншими об'єктами.

У такому випадку фактор невизначеності та зв'язок між об'єктами має бути передбачений вже у процесі проектування схеми БД (на відміну від попереднього випадку, де ми говоримо про усунення невизначеності у відношеннях, наповнених тестовим набором [1], тобто існуючих).

Якщо вважати, що за допомогою кортежа відношення подають характеристику об'єкта, його стан у часі або властивість, то моделюванням об'єкта у сховищі даних є встановлення зв'язку між окремими кортежами. А зважаючи на те, що кожен об'єкт є системою певної складності, то за допомогою властивостей складових об'єкта можна визначати властивості самого об'єкта, або ж, навпаки, переносити властивості об'єкта на його складові. Тобто, моделювання об'єкта за допомогою перелічення його складових або властивостей та перенесення властивостей з вищого рівня ієрархії на нижчий та навпаки є одним із методів усунення невизначеності його характеристик.

Якщо на рівні описів об'єкта передбачити атрибут для подання ступеня істинності чи відповідності, то моделювання мережевої структури об'єктів та їх властивостей у сховищах даних дозволить також позбуватися нечіткостей та неоднозначностей шляхом руху по мережі та аналізу ступенів відповідності.

Для прикладу розглянемо фрагмент відношень:

Admin_unit

Адміністративно-територіальна одиниця	Клімат	Ступінь відповідності	Посилання
Галичина	⊥		
Львів	помірний	0.9	Галичина
Тернопіль	помірно-континентальний	0.8	Галичина
Івано-Франківськ	помірний	0.8	Галичина
Поділля	⊥		
Хмельницький	помірно-континентальний	0.5	Поділля
Київ	помірно-континентальний	0.9	Поділля
Чернівці	помірний	0.6	Галичина

Як бачимо, за атрибутом *Клімат* у нас є невідомі значення. Ставиться задача усунення невизначеності цього типу

Після виконання кроків:

1) аналіз значень атрибутів *Ступінь відповідності*, *Посилання*;

2) визначення кількості появ значень атрибута *Клімат* для об'єктів, у яких значення атрибута *Посилання* є відсутнє

вдалося визначити значення атрибута Клімат для об'єктів Галичина та Поділля (помірний та помірно-континентальний відповідно).

Отже, усунення невизначеності кортежів-предків може відбуватися шляхом аналізу кортежів-нащадків.

У наступному розділі розглядаються методи розв'язання поставлених задач.

4. МЕТОДИ РОЗВ'ЯЗАННЯ ЗАДАЧІ УСУНЕННЯ НЕВИЗНАЧЕНОСТІ

4.1.Застосування алгоритмів класифікації

Видобування класифікаційних правил на основі аналізу навчального набору.

Розглянемо *алгоритм побудови класифікаційних правил*.

1) нехай Y – атрибут, значення якого є мітками класу, за яким проводиться усунення невизначеності.; X – атрибут (множина атрибутів відношення), за значеннями яких відбуватиметься віднесення до класу, $X \cap Y = \varnothing$.

Пересортуємо відношення r за X-атрибутами так, щоб зібрати кортежі з однаковими X-значеннями разом.

Послідовно перебираючи атрибути з підмножини X, для кожного правила визначаємо величини:

s_c – кількість кортежів у відношенні;

s_{xy} – кількість кортежів, для яких виконується $X = x \to Y = y$;

$s_{\bar{x}y}$ – кількість кортежів, для яких $X = \bar{x} \to Y = y$;

$s_{x\bar{y}}$ – кількість кортежів, для яких $X = x \to Y = \bar{y}$;

$s_{\bar{x}\bar{y}}$ – кількість кортежів, для яких $X = \bar{x} \to Y = \bar{y}$.

Кількість кортежів за X та за Y, для яких значення присутні і достовірні (значення спеціального атрибута є більше дорівнює *UNK*) позначаються, відповідно, s_x та s_y;

2) ступінь істинності правила s на основі знайдених у пункті 1 величин може бути визначений як міра невизначеності з нечіткими квантифікаторами (застосовується для опрацювання лінгвістичних змінних) [5, с. 139]:

$$s = 1 - s_c \left| \frac{s_{xy}^{\ 2}}{s_x s_y} - 1 \right|;$$

3) шляхом експертного опитування визначаємо порогове значення *UNK*, за яким правило може вважатися існуючим (значущим) на цьому тестовому наборі. Тобто, якщо

$$s \geq UNK,$$

то говоримо, що знайдене правило підтримується у базі знань із ступенем істинності s.

У [3] розглянуто інші методи визначення порогового значення *UNK*.

Розглянутий метод реалізується на навчальному наборі і тому корелює з вхідними даними.

Для функціональної залежності $s_{xy} = s_x = s_y$, оскільки для кожної пари, у якій $X = x$, підтримується $Y = y$. Звідси $s = 1$.

Побудова класифікаційних правил на основі існуючих правил та функціональних залежностей.

Відомо, що за допомогою логічного виведення можна отримувати нові правила, задані у вигляді "якщо, то" [5]. Класифікаційні правила є саме такого вигляду. У [4] введено поняття наближеної функціональної залежності, яка має встановлене значення ступеня довіри. Якщо класифікаційні правила розглядати як нечіткі (наближені) функціональні залежності зі ступенем довіри s, то до них можна застосувати основні аксіоми виведення [6]. Оскільки ми маємо справу з дискретними величинами, то для опрацювання ступенів довіри до правил використаємо апарат багатозначної логіки Лукасевича (проаналізовано у [8]). Тоді використання правил виведення та застосування логічних операції "і" для правих частин та "або" для лівих дасть можливість генерувати нові правила на основі існуючих та автоматично визначати до них ступені довіри (які можуть бути перевірені експериментально).

Для прикладу розглянемо наведене вище правило

Освіта, Вік,Стаж роботи, Професія→Оклад

Шляхом застосування аксіом визначення, з цього правила можна отримати такі правила:

Освіта, Вік,Стаж роботи, Професія, Кількість попередніх місць роботи →Оклад
Освіта, Вік,Стаж роботи, Професія, Величина міста→Оклад

Звідси випливає, що у базі знань потрібно зберігати лише мінімальне покриття наближених функціональних залежностей (тобто класифікаційних правил), а усі решту можуть бути виведені на основі їх комбінацій з використанням операцій багатозначної логіки (ґрунтується на мінімаксному підході) та аксіом виведення.

Усунення невизначенсті через застосування алгоритму прогонки

Усунення (зменшення) невизначеності відбувається шляхом застосування класифікаційних правил та алгоритму прогонки (*chase*).

Опишемо алгоритм застосування модифікованого методу прогонки.

Нехай у відношенні r підтримується класифікаційне правило $s(X_1, ..., X_n \rightarrow Y)$. Символом $\downarrow$ позначимо визначене значення, $\perp$ – відсутнє значення; t_i – кортеж відношення r (послідовність кортежів значення не має)

1. **Якщо** $\{t_1(X_1)\downarrow, ..., t_1(X_n)\downarrow\}$ **і** $\{t_2(X_1)\downarrow, ..., t_2(X_n)\downarrow,\}$
   і $\{t_1(X_1)\downarrow, ..., t_1(X_n)\downarrow = t_2(X_1)\downarrow, ..., t_2(X_n)\downarrow\}$
   і $\{t_1(Y)\downarrow\}$ і $\{t_2(Y) = \perp\}$,
   то заміняємо $\perp$ на $t_1(Y)$.
2. **Якщо** $\{t_1(X_1)\downarrow, ..., t_1(X_n)\downarrow\}$
   і {в t_2 m з n значень атрибутів – $\downarrow$, $n - m$ значень атрибутів – $\perp$, $m \le n$ }
   і $\{s \ge 1 - m/n\}$ – ступінь довіри до правила більший, ніж міра необхідності кортежа t_2 (5)
   і {за визначеними значеннями $t_1(X^m)\downarrow = t_2(X^m)\downarrow$}

і { $t_1(Y)\downarrow$} **і** { $t_2(Y) = \perp$},

то заміняємо $\perp$ у r на $t_1(Y)$.

3. **Якщо** {в t_i m_i з n значень атрибутів – $\downarrow$, $m_i \le n$ }

і {в t_j m_j з n значень атрибутів – $\downarrow$, $m_j \le n$ }

і {за визначеними значеннями $t_i(X^m)\downarrow = t_2(X^m)\downarrow$}

і {за визначеними значеннями $t_j(X^m)\downarrow = t_2(X^m)\downarrow$}

і $\{ m_i/n \le m_j/n \}$ **і** $\{ s \ge 1 - m_i/n \}$

і { $t_i(Y)\downarrow$} **і** { $t_j(Y)\downarrow$}

і { $t_2(Y) = \perp$},

то заміняємо $\perp$ на $t_j(Y)$.

4.2. Аналіз зв'язків між кортежами відношення

Зв'язок як між кортежами одного відношення, так і між окремими кортежами різних відношень можна встановити за допомогою відношення dc_link:

Dc_link

Id	код
Evdate	Дата занесення зв'язку
TableType	Назва або префікс відношення
Table_id	Код кортежа відношення
Prior_id	Посилання на код

Атрибут Prior_id є посиланням на ключ відношення dc_link та застосовується для встановлення зв'язку між кортежами відношень, назва яких вказана у TableType.

Описана структура (dc_link та використання атрибуту prior_id) дозволяє моделювати складні об'єкти, наприклад, шляхом перерахування їхніх властивостей, та враховувати невизначеність як їхню характеристику, вказавши для кожної властивості ступінь її впиву (відповідності) на об'єкт-предок.

Розглянемо детальніше *алгоритм усунення невизначеності об'єкта-предка на основі аналізу характеристик нащадків*.

1. Як і в попередній задачі, виділимо підмножини атрибутів X та Y, а також атрибут A, який міститиме значення нечіткої змінної як характеристики об'єкта.

2. Приведемо ієрархічну структуру до лінійної шляхом застосування операції об'єднання за кодом кортежів відповідного відношення та значеннями атрибута prior_id.

3. Групування кортежів отриманого відношення за значеннями атрибутів X.

4. Групування кортежів всередині групи $X = x$ за значеннями атрибутів Y.

5. Виконання всередині групи за значенням атрибута $Y = y$ логічної операції "і" для значень атрибута A.

6. Виконання всередині групи за значенням атрибута X логічної операції "або" для значень атрибута A.

7. Заміна значення атрибута A для всіх кортежів всередині групи за X, для яких значення атрибута prior_id є порожнім, на розраховане значення.

8. Заміна (занесення) значення атрибута *Y* для всіх кортежів всередині групи за X, для яких значення атрибута *A* є найвищим.

9. Усунення кортежів, що повторюються.

Слід зазначити, що зв'язок між записами означає не лише фізичний зв'язок, зафіксований або у відношеннях бази даних, або обмеженнями цілісності. Такий зв'язок фактично означає операцію об'єднання за множиною *X*. Звідси випливає, що кортежі нижчих рівнів не обов'язково матимуть ті ж значення по множині атрибутів *Y*, що і записи вищих рівнів, навіть у тому випадку, коли вони задовольнятимуть параметри запиту. Чим більша кількість атрибутів з множини *Y* використовується в операції агрегування, тим більшим є збіг між кортежем-нащадком та кортежем-предком.

У результаті руху по мережі від кортежа-предка до кортежа-нащадка за допомогою операцій багатозначної логіки ступінь істинності (атрибут *A*) предка модифікується з врахуванням ступенів впливу на нього нащадків. За кортежами – безпосередніми нащадками (односторонній зв'язок) виконується операція "і", а за рівноцінними записами (двосторонній зв'язок) – операція "або". У результаті проходження мережі отримуємо нові ступені приналежності кортежів стосовно нечітко заданих параметрів відбору.

5. Напрямок подальших досліджень

Проблема усунення невизначеності розглядалася у багатьох наукових працях. У результаті застосування алгоритмів, поданих у цій статті, вдалося отримати такі результати:

1. Застосування чисел багатозначної логіки для подання ступеня довіри до правил дозволило зберігати у базі даних лише мінімальне покриття класифікаційних правил.

2. Усунення невизначеностей у базі даних можна вважати класифікуванням за певними ознаками.

Залишилися нерозв'язаними проблеми:

1. Вибір методу визначення порогового значення значущості класифікаційних правил.

2. Формальне доведення можливості застосування до класифікаційних правил правил реляційного виведення.

Ці проблеми будуть розглянуті у подальших дослідженнях.

1. Advances in knowledge discovery and data mining. Fayyad U.M., Piatetsky-Shapiro G., Smyth P., Uthurusamy R. (editors), AAAI/MIT Press, 1996. 2. Netz A., Chaudhuri S. Integration of Data Mining and Relational Databases. Proceedings of the 26th International Conference on Very Large Databases, Cairo, Egypt, 2000 3. Grsrfe G., Fayyad U. On the Efficient Gathering of Sufficient Statistics for Classification from Large SQL Databases. – 1998. – www.aaai.org. 4. Huhtala Y., Karkainen J. Tane: An Efficient Algoritm for discovering Functional and Approximate Dependencies // The Computer Journal. – 1999. – Vol. 42. – № 2. 5. Дюбуа А., Прад А. Теория вероятностей. Приложение к представлению знаний в информатике. – М.: Радио и связь, 1990. 6. Мейер Д. Теория реляционных баз данных. – М.: Мир, 1987. 7. Шаховська Н.Б. Застосування апарату багатозначної логіки у системах баз даних // Вісн. Нац. ун-ту "Львівська політехніка". – 2001. – № 438. 8. Шаховська Н.Б., Вовчина А.Я. Обмеження на застосування тризначної логіки у базах даних // Вісн. Нац. ун-ту "Львівська політехніка". – 2001. – № 438.