ВСТУП

Останнім часом для вирішення практичних завдань все частіше застосовуються методи інтелектуального аналізу даних (Data Mining). Інтелектуальний аналіз даних (англ. Data Mining) — виявлення прихованих закономірностей або взаємозв'язків між змінними у великих масивах необроблених даних. Підрозділяється на завдання класифікації, моделювання і прогнозування та інші.

Побудова моделі інтелектуального аналізу даних є складовою частиною масштабнішого процесу, який включає всі етапи, починаючи з визначення базової проблеми, яку модель вирішуватиме, до розгортання моделі в робочому середовищі. Даний процес може бути заданий за допомогою наступних шести базових кроків:

* постановка задачі;
* підготовка даних;
* перегляд даних;
* побудова моделей;
* дослідження, перевірка, прогнозування за допомогою моделей;
* розгортання і оновлення моделей.

До складу Microsoft SQL Server 2005 і 2008 входить цілий ряд служб, які дозволяють виконати кожен крок. Вихідна база даних , як правило, є реляційною, для побудови і наповнення даними інформаційного сховища використовується служба Integration Services, куб будується і представляється в Analysis Services, робота з моделями здійснюється в Biseness Intelligence Studio з використанням спеціальної мови DMX.

На основі цих методів були розроблені алгоритми пошуку асоціативних правил. Вперше ці алгоритми були запропоновані для знаходження типових шаблонів покупок, що здійснюються в супермаркетах. Згодом завдання було розширене, і зараз ці алгоритми вирішують проблему пошуку закономірностей між зв'язаними подіями. Прикладом асоціативного правила може служити вислів, що людина, що купила молоко, також купить хліб за один візит в магазин.

Метою даної роботи є побудова модель інтелектуального аналізу даних з використанням алгоритму асоціативних правил на базі інформаційного сховища підприємства.

Для досягнення цієї мети необхідно вирішити ряд задач:

* створити структуру інформаційного сховища на базі OLTP (Online Transaction Process) бази даних, що містить інформацію про продажі товарів;
* організувати періодичне перевантаження даних з OLTP в інформаційне сховище;
* створити модель інтелектуального аналізу структури споживчої корзини по алгоритму асоціативних правил;
* провести аналіз моделі і прогнозування.

У дипломній роботі детально розглянуто задачі асоціації. Дуже часто покупці набувають не одного товару, а декілька. В більшості випадків між цими товарами існує взаємозв'язок. Ця інформація може бути використана для розміщення товару на полицях в магазинах.

Після створення моделі можна провести її аналіз на предмет виявлення цікавих для нас (шаблонів) правил.

Метою аналізу є встановлення наступних залежностей: якщо в транзакції зустрівся деякий набір елементів X, то на підставі цього можна зробити висновок про те, що інший набір елементів Y також повинен з'явитись в цій транзакції. Встановлення таких залежностей дає нам можливість знаходити дуже прості і інтуїтивно зрозумілі правила.

Сучасні бази даних мають дуже великі розміри, досягаючи гіга- і терабайтів, і тенденцію до подальшого збільшення. І тому, для знаходження асоціативних правил потрібні ефективні масштабовані алгоритми, що дозволяють вирішити задачі за певний час. Один з алгоритмів, що ефективно вирішують подібний клас задач – це алгоритм Apriori.

На основі аналізу можемо створити прогноз даних.

Прогнозування — складання прогнозів продажів і складських запасів, виявлення взаємозалежностей між ними для усунення недоліків і підвищення прибутку.

Для створення прогнозів використовується мова Data Mining Extensions (DMX), яка є розширенням SQL і містить команди для створення, зміни моделей і здійснення передбачень на підставі різних моделей.

1 ОГЛЯД ІСНУЮЧИХ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

1.1 Визначення поняття Data Mining

Data Mining – це процес підтримки ухвалення рішень, заснований на пошуку в даних прихованих закономірностей (шаблонів інформації).

Технологію Data Mining достатньо точно визначає Григорій Піатецкий - Шапіро (Gregory Piatetsky-Shapiro) – один із засновників цього напряму: “Data Mining – це процес виявлення в сирих даних раніше невідомих, нетривіальних, практично корисних і доступних інтерпретації знань, необхідних для ухвалення рішень в різних сферах людської діяльності” [4].

Суть і мету технології Data Mining можна визначити так: це технологія, яка призначена для пошуку у великих об'ємах даних неочевидних, об'єктивних і корисних на практиці закономірностей.

Неочевидних – це значить, що знайдені закономірності не виявляються стандартними методами обробки інформації або експертним шляхом.

Об'єктивних – це значить, що знайдені закономірності повністю відповідатимуть дійсності, на відміну від експертної думки, яка завжди є суб'єктивною.

Практично корисних – це значить, що висновки мають конкретне значення, якому можна знайти практичне застосування.

Знання – сукупність відомостей, яка утворює цілісний опис, відповідний деякому рівню обізнаності про описуване питання, предмет, проблему і т.д.

Використовування знань (knowledge deployment) означає дійсне застосування знайдених знань для досягнення конкретних переваг (наприклад, в конкурентній боротьбі за ринок).

Приведемо ще декілька визначень поняття Data Mining.

Data Mining – це процес виділення з даних неявної і неструктурованої інформації і представлення її у вигляді, придатному для використовування.

Data Mining – це процес виділення, дослідження і моделювання великих об'ємів даних для виявлення невідомих до цього шаблонів (patterns) з метою досягнення переваг в бізнесі (визначення SAS Institute).

Data Mining – це процес, мета якого – знайти нові значущі кореляції, зразки і тенденції в результаті просівання великого об'єму бережених даних з використанням методик розпізнавання зразків плюс застосування статистичних і математичних методів (визначення Gartner Group).

«Mining» англійською означає «видобуток корисних копалин», а пошук закономірностей у величезній кількості даних дійсно схожий на цей процес.

Перш ніж використовувати технологію Data Mining, необхідно ретельно проаналізувати її проблеми [4]:

* Data Mining не може замінити аналітика;
* не може складати розробки і експлуатації додатку Data Mining;
* потрібна підвищена кваліфікація користувача;
* витягання корисних відомостей неможливе без доброго розуміння суті даних;
* складність підготовки даних;
* висока вартість;
* вимога наявності достатньої кількості репрезентативних даних.

Data Mining тісно пов’язана з різними дисциплінами , що засновані на інформаційних технологіях та математичних методах обробки інформаціі (рисунок 1.1).

Теорія БД

Інші дисципліни

Статистика

Машинне навчання

Data Mining

Візуалізація

Алгоритмізація

Штучний інтелект

Розпізнавання образів

Рисунок 1.1 – Data Mining як мультідісциплінарна область

Кожний з напрямів, що сформували Data Mining, має свої особливості. Проведемо порівняння з деякими з них.

* 1. Порівняння статистики, машинного навчання і Data Mining

Статистика– це наука про методи збору даних, їх обробки і аналізу для виявлення закономірностей, властивих явищу, що вивчається.

Статистикає сукупністю методів планування експерименту, збору даних, їх уявлення і узагальнення, а також аналізу і отримання висновків на підставі цих даних.

Статистика оперує даними, що отримані в результаті спостережень або експериментів.

Перевагами є:

* більш ніж Data Mining, базується на теорії;
* більш зосереджується на перевірці гіпотез.

Єдиного визначення машинного навчання на сьогоднішній день немає.

Машинне навчанняможна охарактеризувати як процес отримання програмою нових знань. Мітчелл в 1996 році дав таке визначення: «Машинне навчання – це наука, яка вивчає комп'ютерні алгоритми, автоматично що поліпшуються під час роботи».

Одним з найпопулярніших прикладів алгоритму машинного навчання є нейронні мережі.

Алгоритми машинного навчання є:

* більш евристичні;
* концентрується на поліпшенні роботи агентів навчання.

Переваги Data Mining:

* інтеграція теорії і евристик;
* сконцентрована на єдиному процесі аналізу даних, включає очищення даних, навчання, інтеграцію і візуалізацію результатів.
  1. Методи Data Mining

Методи, що використовує технологія Data Mining можна розподілити на технологічні, статистичні та кібернетичні.

Таблиця 1.1- Методи Data Mining

|  |  |
| --- | --- |
| Методи Data Mining | Характеристика |
| Технологічні методи | а) безпосереднє використання даних, або збереження даних. Методи цієї групи: кластерний аналіз, метод найближчого сусіда;  б) виявлення і використання формалізованих закономірностей, або дистиляція шаблонів - логічні методи, методи візуалізації, методи крос-табуляції, методи, що засновані на рівняннях. |
| Статистичні методи | а) дескриптивний аналіз і опис вихідних даних;  б) аналіз зв'язків (кореляційний і регресійний аналіз, факторний аналіз, дисперсійний аналіз);  в) багатовимірний статистичний аналіз (компонентний аналіз, дискримінантний аналіз, багатовимірний регресійний аналіз, канонічні кореляції і ін.);  г) аналіз тимчасових рядів (динамічні моделі і прогнозування). |
| Кібернетичні методи | а)штучні нейронні мережі (розпізнавання, кластеризація, прогноз);  б) еволюційне програмування (в т.ч. алгоритми методу групового обліку аргументів);  в) генетичні алгоритми (оптимізація);  ґ) асоціативний алгоритм;  г) нечітка логіка;  д) дерева рішень;  є) системи обробки експертних знань. |

1.4 Відмінності Data Mining від інших методів аналізу даних

Традиційні методи аналізу даних в основному орієнтовані на перевірку наперед сформульованих гіпотез (статистичні методи) і на «грубий розвідувальний аналіз», що становить основу оперативної аналітичної обробки даних (Online Analytical Processing, OLAP), тоді як одне з основних положень Data Mining – пошук неочевидних закономірностей. Інструменти Data Mining можуть знаходити такі закономірності самостійно і також самостійно будувати гіпотези про взаємозв'язки. Оскільки саме формулювання гіпотези щодо залежності є найскладнішою задачею, перевага Data Mining в порівнянні з іншими методами аналізу є очевидною.

Більшість статистичних методів для виявлення взаємозв'язків в даних використовує концепцію усереднювання по вибірці, що приводить до операцій над неіснуючими величинами, тоді як Data Mining оперує реальними значеннями.

OLAP більше підходить для розуміння ретроспективних даних, Data Mining спирається на ретроспективні дані для отримання відповідей на питання про майбутнє.

2 МАТЕМАТИЧНА ПОСТАНОВКА ЗАДАЧ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ – АЛГОРИТМ АСОЦІАТИВНИХ ПРАВИЛ

Однією з задач Data Mining є асоціація. Метою пошуку асоціативних правил (association rule) є знаходження закономірностей між зв'язаними подіями в базах даних.

Дуже часто покупці придбавають не один товар, а декілька. В більшості випадків між цими товарами існує взаємозв'язок. Так, наприклад, покупець, що придбаває макаронні вироби, швидше за все, схоче придбати також кетчуп. Ця інформація може бути використана для розміщення товару на полицях крамниці.

Наведемо простий приклад асоціативного правила: покупець, що придбаває банку фарби, придбає пензлик для фарби з вірогідністю 50%.

Вперше ця задача була запропонована пошуку асоціативних правил для знаходження типових шаблонів покупок, які придбають в супермаркетах, тому іноді її ще називають аналізом ринкової корзини (market basket analysis).

Хай є база даних, що складається з купівельних транзакцій. Кожна транзакція – це набір товарів, куплених покупцем за один візит. Таку транзакцію ще називають ринковою корзиною.

Визначення 1. Хай I = {i1, i2, i3 ... in} – безліч (набір) товарів, званих елементами. Хай D - безліч транзакцій, де кожна транзакція T – це набір елементів з I, TI. Кожна транзакція є бінарним вектором, де t[k]=1, якщо ik елемент присутній в транзакції, інакше t[k]=0. Ми говоримо, що транзакція T містить X, деякий набір елементів з I, якщо XT. Асоціативним правилом називається імплікація XY, де XI, YI і XY=. Правило XY має підтримку s (support), якщо s% транзакцій з D, містять XY, supp(XY) = supp(XY). Достовірність правила показує, яка вірогідність того, що з X слідує У. Правило XY справедливо з достовірністю (confidence) з, якщо c% транзакцій з D, що містять X, також містять У, conf(XY) = supp(XY)/supp(X).

Покажемо на конкретному прикладі: "75% транзакцій, що містять хліб, також містять молоко. 3% від загального числа всіх транзакцій містять обидва товари". 75% - це достовірність (confidence) правила, 3% це підтримка (support), або "Хліб" "Молоко" з вірогідністю 75%.

Іншими словами, метою аналізу є встановлення наступної залежності: якщо в транзакції зустрівся деякий набір елементів X, то на підставі цього можна зробити висновок про те, що інший набір елементів У також повинен з'явитись в цій транзакції. Встановлення такої залежності дає нам можливість знаходити дуже прості і інтуїтивно зрозумілі правила.

Алгоритми пошуку асоціативних правил призначені для знаходження всіх правил X У, причому підтримка і достовірність цих правил повинна бути вищою за деякі наперед певні пороги, звані відповідно мінімальною підтримкою (minsupport) і мінімальною достовірністю (minconfidence).

Задача знаходження асоціативних правил розбивається на дві підзадачі:

1. знаходження всіх наборів елементів, які задовольняють порогу minsupport. Такі набори елементів називаються тими, що часто зустрічаються;
2. генерація правил з наборів елементів, знайдених згідно п. a) з достовірністю, що задовольняє порогу minconfidence.

Один з перших алгоритмів, ефективно вирішальних подібний клас задач, – це алгоритм APriori. Окрім цього алгоритму останнім часом був розроблений ряд інших алгоритмів: DHP, Partition, DIC і інші.

Значення для параметрів мінімальна підтримка і мінімальна достовірність вибираються так, щоб обмежити кількість знайдених правил. Якщо підтримка має велике значення, то алгоритми будуть знаходити правила, добре відомі аналітикам або настільки очевидні, що немає ніякого значення проводити такий аналіз. З другого боку, низьке значення підтримки веде до генерації величезної кількості правил, що, звичайно, вимагає істотних обчислювальних ресурсів. Проте, більшість цікавих правил знаходиться саме при низькому значенні порогу підтримки. Хоча дуже низьке значення підтримки веде до генерації статистично необґрунтованих правил.

Пошук асоціативних правил зовсім не тривіальна задача, як може здатися на перший погляд. Одна з проблем – алгоритмічна складність при знаходженні наборів елементів, які часто зустрічаються, оскільки із зростанням числа елементів експоненційно росте число потенційних наборів елементів.

2.1 Узагальнені асоціативні правила (Generalized Association Rules)

При пошуку асоціативних правил, ми припускали, що всі аналізовані елементи однорідні. Повертаючись до аналізу ринкової корзини, це товари, абсолютно однакові атрибути, що мають, за винятком назви. Проте не складе великих труднощів доповнити транзакцію інформацією про те, до якої товарної групи входить товар і побудувати ієрархію товарів.

Наприклад, якщо Покупець купив товар з групи "Безалкогольні напої", то він купить і товар з групи "Молочні продукти" або "Сік". Ці правила носять назву узагальнених асоціативних правил.

Визначення 2. Узагальненим асоціативним правилом називається імплікація XY, де XI, YI і XY= і де жоден з елементів, що входять в набір У, не є предком жодного елемента, що входить в X. Підтримка і достовірність підраховуються так само, як і у разі асоціативних правил (див. Визначення 1).

Введення додаткової інформації про угрупування елементів у вигляді ієрархії дасть наступні переваги:

1. це допомагає встановити асоціативні правила не тільки між окремими елементами, але і між різними рівнями ієрархії (групами);
2. окремі елементи можуть мати недостатню підтримку, але в цілому група може задовольняти порогу minsupport;
3. для знаходження таких правил можна використовувати будь-який з вищеназваних алгоритмів.

Групувати елементи можна не тільки по входженню в певну товарну групу, але і по інших характеристиках, наприклад за ціною (дешево, дорого), бренду і т.д.

Алгоритм пошуку асоціативних правил, заснований на аналізі частих наукових наборів. Спочатку в базі даних транзакцій шукаються усі частини наукових наборів, а потім генеруються асоціативні правила на основі тих з них, які задовольняють заданому рівню підтримки і достовірності.

При цьому для скорочення простору пошуку асоціативних правил використовується властивість апріорності. Воно затверджує, що якщо науковий набір Z не є частим, то додавання будь – якого нового предмету А до набору Z не робить його таким. Іншими словами, якщо набір Z не є частим, то і Z + A – теж.

2.2 Властивість анти–монотонності

Виявлення наборів елементів, що часто зустрічаються, – операція, що вимагає багато обчислювальних ресурсів і, відповідно, часу. Примітивний підхід до рішення даної задачі – простий перебір всіх можливих наборів елементів. Це потребує O(2|I|) операцій, де |I| – кількість елементів. Apriori використовує одну з властивостей підтримки, що свідчить: підтримка будь-якого набору елементів не може перевищувати мінімальної підтримки будь-якої з його підмножин. Наприклад, підтримка 3–елементного набору {Хліб, Масло, Молоко} буде завжди менше або рівно підтримці 2–елементних наборів {Хліб, Масло}, {Хліб, Молоко}{Масло, Молоко}. Річ у тому, що будь-яка транзакція, що містить {Хліб, Масло, Молоко}, також повинна містити {Хліб, Масло}, {Хліб, Молоко}, {Масло, Молоко}, причому зворотне не вірно.

Ця властивість носить назву анти–монотонності і служить для зниження розмірності простору пошуку. Не май ми в наявності такої властивості, знаходження багатоелементних наборів було б практично нездійсненною задачею у зв'язку з експоненціальним зростанням обчислень.

Властивості анти–монотонності можна дати і інше формулювання: із зростанням розміру набору елементів підтримка зменшується, або залишається такою ж. Зі всього, що було сказано раніше витікає, що будь–який k–елементний набір буде часто зустрічатися тоді і тільки тоді, коли всі його (k–1)–елементні підмножини будуть часто зустрічатись. Всі можливі набори елементів з I можна представити у вигляді грат, що починаються з порожньої множини, потім на 1 рівні 1–елементні набори, на 2–м – 2–елементні і т.д. На k рівні представлені k–елементні набори, пов'язані зі всіма своїми (k – 1) – елементними підмножинами.

Розглянемо малюнок, ілюструючи набір елементів I – {А, B, З, D}. Припустимо, що набір з елементів {А, B} має підтримку нижче заданого порогу і, відповідно, не є тим, що часто зустрічається. Тоді, згідно властивості анти–монотонності, всі його супермножини також не є тими, що часто зустрічаються і відкидаються. Вся ця гілка, починаючи з {А, B}, виділена фоном. Використовування цієї евристики дозволяє істотно скоротити простір пошуку.

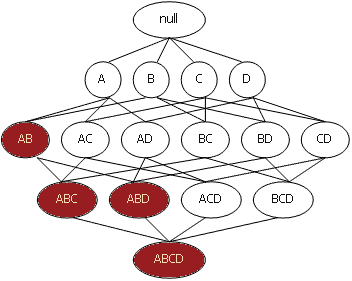


Рисунок 2.1 – Набір елементів I 2.3 Алгоритм Apriori

Для того, щоб було можливе застосувати алгоритм, необхідно провести предобработку даних: по-перше, привести всі дані до бінарного вигляду; по-друге, змінити структуру даних[9]. Вигляд транзакційної бази даних представлен в таблиці 2.1 і таблиці 2.2.

Таблиця 2.1 – Звичайний вигляд

|  |  |  |
| --- | --- | --- |
| **Номер транзакції** | **Назва елемента** | **Кількість** |
| 1001 | А | 2 |
| 1001 | D | 3 |
| 1001 | E | 1 |
| 1002 | А | 2 |
| 1002 | F | 1 |
| 1003 | B | 2 |
| 1003 | A | 2 |
| 1003 | C | 2 |
| … | … | … |

Таблиця 2.2 – Нормалізований вигляд

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **TID** | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** | **K** | **…** |
| 1001 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | … |
| 1002 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | … |
| 1003 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | … |
|  |  |  |  |  |  |  |  |  |  |  |  |

Кількість стовпців в таблиці рівно кількості елементів, присутніх в безлічі транзакцій D. Кожний запис відповідає транзакції, де у відповідному стовпці стоїть 1, якщо елемент присутній в транзакції, і 0 в осоружному випадку. Помітимо, що початковий вид таблиці може бути відмінним від приведеного в таблиці 2.1. Головне, щоб дані були перетворені до нормалізованого вигляду, інакше алгоритм не застосовний. Крім того, всі елементи повинні бути впорядкований в алфавітному порядку (якщо це числа, вони повинні бути впорядкований в числовому порядку).

На першому кроці алгоритму підраховуються 1–елементні набори, що часто зустрічаються. Для цього необхідно пройтися по всьому набору даних і підрахувати для них підтримку, тобто скільки разів зустрічається в базі.

Наступні кроки складатимуться з двох частин: генерації наборів елементів (їх називають кандидатами) і підрахунку підтримки, що потенційно часто зустрічаються, для кандидатів[4].

Описаний вище алгоритм можна записати у вигляді наступного псевдокоду:

F1 = {1-елементні набори, що часто зустрічаються}

для (k=2; Fk-1 <> ; k++) {



Ck = Apriorigen (Fk-1) // генерація кандидатів

для всіх транзакцій t T {



Ct = subset(Ck, t)// видалення надмірних правил

для всіх кандидатів с Ct



c.count ++

}

Fk = {с Ck | c.count >= minsupport} // відбір кандидатів



}

Результат Fk



Опишемо функцію генерації кандидатів. На цей раз немає ніякої необхідності знов звертатися до бази даних. Для того, щоб отримати k–елементні набори, скористаємося (k–1)–елементними наборами, які були визначені на попередньому кроці і є тими, що часто зустрічаються.

Пригадаємо, що наш початковий набір зберігається у впорядкованому вигляді.

Генерація кандидатів також складатиметься з двох кроків.

Крок 1. Об'єднання. Кожний кандидат Cк формуватиметься шляхом

розширення набору розміру (k–1), що часто зустрічається, додаванням елемента з іншого (k–1)–елементного набору.

Приведемо алгоритм цієї функції Apriorigen у вигляді невеликого SQL –подібного запиту.

insert into Ck

select p.item1, p.item2 ., p.itemk-1, q.itemk-1  
From Fk-1 p, Fk-1 q  
where p.item1= q.item1, p.item2 = q.item2 ., p.itemk-2 = q.itemk-2, p.itemk-1 < q.itemk-1

Крок 2. Видалення надмірних правил. На підставі властивості анти–монотонності, слід видалити всі набори с Ck якщо хоча б одна з його (k–1) підмножин не є тим, що часто зустрічається.



Після генерації кандидатів наступною задачею є підрахунок підтримки для кожного кандидата. Очевидно, що кількість кандидатів може бути дуже великою і потрібен ефективний спосіб підрахунку. Найтривіальніший спосіб – порівняти кожну транзакцію з кожним кандидатом. Але це далеко не краще рішення. Набагато швидше і ефективно використовувати підхід, заснований на зберіганні кандидатів в хеш–дереві. Внутрішні вузли дерева містять хеш–таблиці з покажчиками на нащадків, а листя – на кандидатів. Це дерево нам стане в нагоді для швидкого підрахунку підтримки для кандидатів.

Хеш–дерево будується кожного разу, коли формуються кандидати. Спочатку дерево складається тільки з кореня, який є листом, і не містить ніяких кандидатів – наборів. Кожного разу коли формується новий кандидат, він заноситься в корінь дерева і так до тих пір, поки кількість кандидатів в корені – листі не перевищить якогось порогу. Як тільки кількість кандидатів стає більше порогу, корінь перетвориться в хеш–таблицю, тобто стає внутрішнім вузлом, і для нього створюються нащадки – листя. І всі приклади розподіляються по вузлах – нащадкам згідно з хеш–значеннями елементів, що входять в набір, і т.п. Кожний новий кандидат хеширується на внутрішніх вузлах, поки він не досягне першого вузла–листа, де він і зберігатиметься, поки кількість наборів знову ж таки не перевищить порогу.

Хеш–дерево з кандидатами–наборами побудовано, зараз, використовуючи хеш–дерево, легко підрахувати підтримку для кожного кандидата. Для цього потрібно "пропустити" кожну транзакцію через дерево і збільшити лічильники для тих кандидатів, чиї елементи також містяться і в транзакції, тобто Ck Ti = Ck. На кореневому рівні хеш–функція застосовується до кожного елемента з транзакції. Далі, на другому рівні, хеш–функція застосовується до других елементів і т.д. На k–рівні хеширується k–елемент. І так до тих пір, поки не досягнемо листа. Якщо кандидат, що зберігається в листі, є підмножиною даної транзакції, тоді збільшуємо лічильник підтримки цього кандидата на одиницю.



Після того, як кожна транзакція з початкового набору даних "пропущена" через дерево, можна перевірити чи задовольняють значення підтримки кандидатів мінімальному порогу. Кандидати, для яких ця умова виконується, переносяться в розряд часто зустрічаємих. Крім того, слід запам'ятати і підтримку набору, вона нам стане в нагоді при витяганні правил. Ці ж дії застосовуються для знаходження (k+1)–елементних наборів і т.д.

Після того, як знайдені всі часто зустрічаємі набори елементів можна приступити безпосередньо до генерації правил.

Витягання правил – менш трудомістка задача. По–перше, для підрахунку достовірності правила достатньо знати підтримку самого набору і множини, що лежить в умові правила. Наприклад, є набір {А, B, С}, що часто зустрічається, і потрібно підрахувати достовірність для правила ABС. Підтримка самого набору нам відома, але і його множина {А, B}, що лежить в умові правила, також часто зустрічається через властивість анти–монотонності, і значить, його підтримка нам відома. Тоді ми легко зможемо підрахувати достовірність. Це позбавляє нас від небажаного перегляду бази транзакцій, який був б потрібен в тому випадку, якщо б це підтримка була невідома.



Щоб витягнути правило із часто зустрічає мого набору F, слід знайти всі його не порожні підмножини. І для кожної підмножини s ми зможемо сформулювати правило s (F – s), якщо достовірність правила conf(s (F – s)) = supp(F)/supp(s) не менше порогу minconf.



Помітимо, що чисельник залишається постійним. Тоді достовірність має мінімальне значення, якщо знаменник має максимальне значення, а це відбувається у тому випадку, коли в умові правила є набір, що складається з одного елемента. Все супермножина даної множини має меншу або рівну підтримку і, відповідно, більше значення достовірності. Ця властивість може бути використана при витяганні правил. Якщо ми почнемо витягувати правила, розглядаючи спочатку тільки один елемент в умові правила, і це правило має необхідну підтримку, тоді всі правила, де в умові стоїть супермножина цього елемента, також мають значення достовірності вище заданого порогу. Наприклад, якщо правило АBCDE задовольняє мінімальному порогу достовірності minconf, тоді ABCDE також задовольняє. Для того, щоб витягнути всі правила використовується рекурсивна процедура. Важливе зауваження: будь-яке правило, складене з набору, що часто зустрічається, повинне містити всі елементи набору. Наприклад, якщо набір складається з елементів {А, B, С}, то правило АB не повинне розглядатися.



3 DATA MINING ЯК ЧАСТИНА СИСТЕМИ АНАЛІТИЧНОЇ ОБРОБКИ ІНФОРМАЦІЇ

3.1 Сховища даних

Інформаційні системи сучасних підприємств часто організовані так, щоб мінімізувати час введення і коректування даних, тобто організовані не оптимально з погляду проектування бази даних. Такий підхід ускладнює доступ до історичних (архівним) даних. Зміни структур в базах даних інформаційних систем дуже трудомісткі, а іноді просто неможливі.

В той же час, для успішного ведення сучасного бізнесу необхідна актуальна інформація, що надається в зручному для аналізу вигляді і в реальному масштабі часу. Доступність такої інформації дозволяє, як оцінювати поточне положення справ, так і робити прогнози на майбутнє, отже, ухвалювати більш зважені і обґрунтовані рішення. До того ж, основою для ухвалення рішень повинні бути реальні дані.

Якщо дані зберігаються в базах даних різних інформаційних систем підприємства, при їх аналізі виникає ряд складнощів, зокрема, значно зростає час, необхідний для обробки запитів; можуть виникати проблеми з підтримкою різних форматів даних, а також з їх кодуванням; неможливість аналізу тривалих рядів ретроспективних даних і т.д.

Ця проблема розв'язується шляхом створення сховища даних. Задачею такого сховища є інтеграція, актуалізація і узгодження оперативних даних з різнорідних джерел для формування єдиного несуперечливого погляду на об'єкт управління в цілому. На основі сховищ даних можливо складання всілякої звітності, а також проведення оперативної аналітичної обробки і Data Mining.

Тоді загальна схема інформаційного сховища буде виглядати наступнім чином:

OLTP

БД

.doc

OLTP

БД

.xls

Загрузка

Очистка

Трансформація

Інформаційне сховище

Засоби аналізу та представлення даних

OLAP

Data Mining

Вітрина

Вітрина

Вітрина

Рисунок 3.1 – Структура інформаційної системи

Біл Інмон (Bill Inmon) визначає сховища даних як "наочно орієнтовані, інтегровані, немінливі, підтримуючі хронологію набори даних, організовані з метою підтримки управління" і покликані виступати в ролі "єдиного і єдиного джерела істини", яке забезпечує менеджерів і аналітиків достовірною інформацією, необхідною для оперативного аналізу і ухвалення рішень[3].

Наочна орієнтаціясховища даних означає, що дані з'єднані в категорії і зберігаються відповідно областям, які вони описують, а не застосуванням, що їх використовують.

Інтегрованістьозначає, що дані задовольняють вимогам всього підприємства, а не однієї функції бізнесу. Цим сховище даних гарантує, що однакові звіти, що згенерували для різних аналітиків, міститимуть однакові результати.

Прив'язка до часу означає, що сховище можна розглядати як сукупність "історичних даних": можливо відновлення даних на будь-який момент часу. Атрибут часу явно присутній в структурах сховища даних.

Незмінністьозначає, що, потрапивши один раз в сховищі, дані там зберігаються і не змінюються. Дані в сховищі можуть лише додаватися.

Річард Хакаторн, інший основоположник цієї концепції, писав, що мета Сховищ Даних – забезпечити для організації "єдиний образ існуючої реальності"[3].

Іншими словами, сховище даних є своєрідним накопичувачем інформації про діяльність підприємства.

Дані в сховищі представлені у вигляді багатовимірних структур під назвою "зірка" або "сніжинка".

3.2 Організація інформаційного сховища в реалізації бази даних. Схеми зірка та сніжинка

Схема типу зірки (Star Schema) – схема реляційної бази даних, що служить для підтримки багатовимірного представлення даних, які в ній зберігаються.

Особливості ROLAP-схеми типу "зірка":

а) одна таблиця фактів (fact table), яка сильно денормалізована. Є центральною в схемі, може складатися з мільйонів рядків і містить підсумовувані або фактичні дані, за допомогою яких можна відповісти на різні питання;

б) декілька денормалізованих таблиць вимірювань (dimensional table). Мають меншу кількість рядків, ніж таблиці фактів, і містять описову інформацію. Ці таблиці дозволяють користувачу швидко переходити від таблиці фактів до додаткової інформації;

в) таблиця фактів і таблиці розмірності зв'язані ідентифікуючими зв'язками, при цьому первинні ключі таблиці розмірності мігрують в таблицю фактів як зовнішні ключі. Первинний ключ таблиці факту цілком складається з первинних ключів всіх таблиць розмірності;

ґ) агреговані дані зберігаються спільно з початковими.

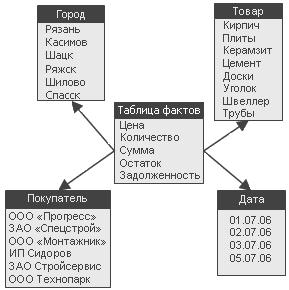


Рисунок 3.2 – Схема «зірка»

#### Схема типу сніжинки (Snowflake Schema) – схема реляційної бази даних, яка служить для підтримки багатовимірного представлення даних,що в ній знаходяться, є різновидом схеми типу "зірка" (Star Schema).

#### Особливості ROLAP-схеми типу "сніжинка":

#### а) одна таблиця фактів (fact table), яка сильно денормалізована. Є центральною в схемі, може складатися з мільйонів рядків і містити підсумовувані або фактичні дані, за допомогою яких можна відповісти на різні питання;

#### б) декілька таблиць вимірювань (dimensional table), які нормалізовані на відміну від схеми "зірка". Мають меншу кількість рядків, ніж таблиці фактів, і містять описову інформацію. Ці таблиці дозволяють користувачу швидко переходити від таблиці фактів до додаткової інформації. Первинні ключі в них складаються з єдиного атрибута (відповідають єдиному елементу вимірювання);

#### в) таблиця фактів і таблиці розмірності зв'язані ідентифікуючими зв'язками, при цьому первинні ключі таблиці розмірності мігрують в таблицю фактів як зовнішні ключі. Первинний ключ таблиці факту цілком складається з первинних ключів всіх таблиць розмірності;

#### ґ) в схемі "сніжинка" агреговані дані можуть зберігатися окремо від початкових.

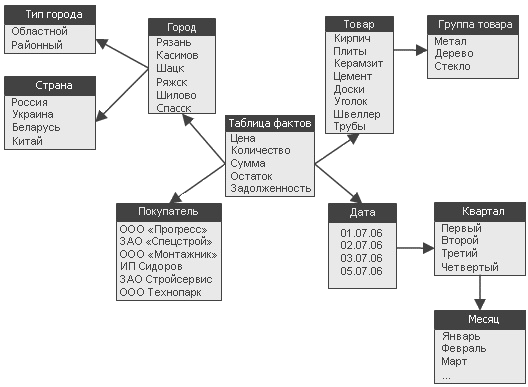


Рисунок 3.3 – Схема «сніжинка»

3.3 OLAP-системи

В основі концепції OLAP, або оперативної аналітичної обробки даних (On-Line Analytical Processing), лежить багатовимірне концептуальне представлення даних (Multidimensional conceptual view).

Термін OLAP введений Коддом (E. F. Codd) в 1993 році. Головна ідея даної системи полягає в побудові багатовимірних таблиць, які можуть бути доступний для запитів користувачів. Ці багатовимірні таблиці або так звані багатовимірні куби будуються на основі початкових і агрегованих даних. І початкові, і агреговані дані для багатовимірних таблиць можуть зберігатися як в реляційних, так і в багатовимірних базах даних. Взаємодіючи з OLAP-системою, користувач може здійснювати гнучкий перегляд інформації, одержувати різні зрізи даних, виконувати аналітичні операції деталізації, згортки, крізного розподілу, порівняння в часі. Вся робота з OLAP-системою відбувається в термінах наочної області[3].

Існує три способи зберігання даних в OLAP-системах або три архітектура OLAP - серверів:

* MOLAP (Multidimensional OLAP);
* ROLAP (Relational OLAP);
* HOLAP (Hybrid OLAP).

Таким чином, згідно цієї класифікації OLAP-продукти можуть бути представлений трьома класами систем:

* у разі MOLAP, початкові і багатовимірні дані зберігаються в багатовимірній БД або в багатовимірному локальному кубі;
* в ROLAP-продуктах початкові дані зберігаються в реляційних БД або в плоских локальних таблицях на файл-сервері. Агрегатні дані можуть поміщатися в службові таблиці в тій же БД;
* у разі використовування гібридної архітектури, тобто в HOLAP-продуктах, початкові дані залишаються в реляційній базі, а агрегати розміщуються в багатовимірній.

Логічним уявленням є багатовимірний куб — це набір зв'язаних заходів і вимірювань, які використовуються для аналізу даних[3].

OLAP-куб підтримує всі багатовимірні операції: довільне розміщення вимірювань і фактів, фільтрація, сортування, угрупування, різні способи агрегації і деталізації. Дані відображаються у вигляді крос-таблиць і крос-діаграм, всі операції відбуваються "на льоту", методи маніпулювання інтуїтивно зрозумілі.

OLAP-куб – могутній інструмент дослідження, що дозволяє проводити розвідувальний і порівняльний аналіз, виявляти тенденції, знаходити сезонність і тренд, визначати кращі і гірші товарні позиції, розраховувати їх частки у продажі.

3.4 Інтеграція OLAP і Data Mining

Обидві технології можна розглядати як складові частини процесу підтримки ухвалення рішень. Проте ці технології як би рухаються у різних напрямах: OLAP зосереджує увагу виключно на забезпеченні доступу до багатовимірних даних, а методи Data Mining в більшості випадків працюють з плоскими одновимірними таблицями і реляційними даними.

Інтеграція технологій OLAP і Data Mining "збагатила" функціональність і однієї, і іншої технології. Ці два види аналізу повинні бути тісно з'єднано, щоб інтегрована технологія могла забезпечувати одночасно багатовимірний доступ і пошук закономірностей.

Засіб багатовимірного інтелектуального аналізу даних повинен знаходити закономірності як в тих, що деталізуються, так і в агрегованих з різним ступенем узагальнення даних. Аналіз багатовимірних даних повинен будуватися над гіперкубом спеціального вигляду, вічка якого містять не довільні чисельні значення (кількість подій, об'єм продажів, сума зібраних податків), а числа, що визначають вірогідність відповідного поєднання значень атрибутів. Проекції такого гіперкуба (що виключають з розгляду окремі вимірювання) також повинні досліджуватися на предмет пошуку закономірностей. J. Han пропонує ще більш просту назву - "OLAP Mining" і висуває декілька варіантів інтеграції двох технологій[3]:

а) "Cubing then mining". Можливість виконання інтелектуального аналізу повинна забезпечуватися над будь-яким результатом запиту до багатовимірного концептуального уявлення, тобто над будь - яким фрагментом будь - якої проекції гіперкуба показників;

б) "Mining then cubing". Подібно даним, витягнутим з сховища, результати інтелектуального аналізу повинні представлятися в гіперкубічній формі для подальшого багатовимірного аналізу;

в) "Cubing while mining". Цей гнучкий спосіб інтеграції дозволяє автоматично активізувати однотипні механізми інтелектуальної обробки над результатом кожного кроку багатовимірного аналізу (переходу між рівнями узагальнення, витягання нового фрагмента гіперкуба і т.д.).

На сьогоднішній день небагато виробників реалізують Data Mining для багатовимірних даних. Крім того, деякі методи Data Mining, наприклад, метод найближчих сусідів або байєсівськая класифікація, через їх нездатність працювати з агрегованими даними незастосовні до багатовимірних даних.

4 СТРУКТУРА ІНФОРМАЦІЙНОГО СХОВИЩА ДЛЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ

4.1 Характеристика джерела даних для інформаційного сховища

У даній роботі за основу була узята БД-зразок Microsoft – Adventure Works[18]. Проект Adventure Works описує роботу виробника велосипедів - компанії "Adventure Works Cycles". Компанія займається виробництвом і реалізацією велосипедів з металевих і композиційних матеріалів на території Північної Америки, Європи і Азії. Головне виробництво, яке має в своєму розпорядженні 500 співробітників, знаходиться в місті Bothell, штат Вашингтон. Декілька регіональних офісів знаходяться безпосередньо на території ринків збуту.

Компанія реалізує продукцію оптом для спеціалізованих магазинів і на роздріб через Інтернет. Для вирішення демонстраційних завдань ми використовуватимемо в базі AdventureWorks дані об інтернет продажах, оскільки вони містять дані, які добре підходять для аналізу.

На рисунку 4.1 представлена транзакційна бази даних AdventureWorks, відділу продаж, яка містить наступні таблиці:

* таблиця SalesTaxRate – в якій містяться податкові ставки, вживані в областях або країнах і регіонах, в яких компанія Adventure Works Cycles здійснює ділову активність;
* таблиця ShoppingCartItem – містить замовлення клієнтів через інтернет до моменту виконання або відміни;
* таблиця SpecialOfferProduct – в якій приведені знижки на різні види (найменування) продукції;
* таблиця SpecialOffer – в якій містяться знижки на продаж;
* таблиця CountryRegionCurrency – зіставляє коди валют по стандартах Міжнародної організації по стандартизації (ISO) і коди країн або регіонів;
* таблиця Currency – містить описи валют по стандартах Міжнародної організації стандартизації (ISO);
* таблиця SalesTerritoryHistory – у таблиці відстежуються переміщення комерційних представників в інші комерційні території;
* таблиця SalesTerritory – в якій містяться території продажів, які обслуговуються групами продажів Adventure Works Cycles;
* таблиця SalesPersonQuotaHistory – містить зведення по історії продажів для комерційних представників;
* таблиця Store – містить список замовників, торгівельних посередників, що купують продукти в Adventure Works;
* таблиця CurrencyRate – містить курси обміну валюти;
* таблиця SalesPerson – містить поточні відомості про продажі для комерційних представників;
* таблиця SalesOrderDetail – містить окремі продукти, пов'язані з певним замовленням на продаж. Замовлення на продаж може містити замовлення на декілька продуктів;
* таблиця SalesOrderHeader – містить відомості про загальне або батьківське замовлення на продаж;
* таблиця Customer – містить поточні відомості про замовника. Клієнти розбиті на категорії по типах — приватний споживач або магазин роздрібної торгівлі;
* таблиця StoreContact – в якій зіставляються магазини і їх службовці, з якими безпосередньо співробітничають торгівельні представники компанії Adventure Works Cycles;
* таблиця SalesReason – в якій містяться можливі причини придбання клієнтом певного продукту;
* таблиця SalesOrderHeaderSalesReason – в якій замовлення на продаж зіставляються з кодами причин продажів;
* таблиця CustomerAddress – зіставляє замовників з їх адресами. Наприклад, замовник може мати різні адреси для виставляння рахунків і доставки.

4.2 Структура інформаційного сховища

Для подальшого інтелектуального аналізу було розроблено структуру інформаційного сховища на базі схеми «сніжинка». На рисунку приведена логічна схема інформаційного сховища.

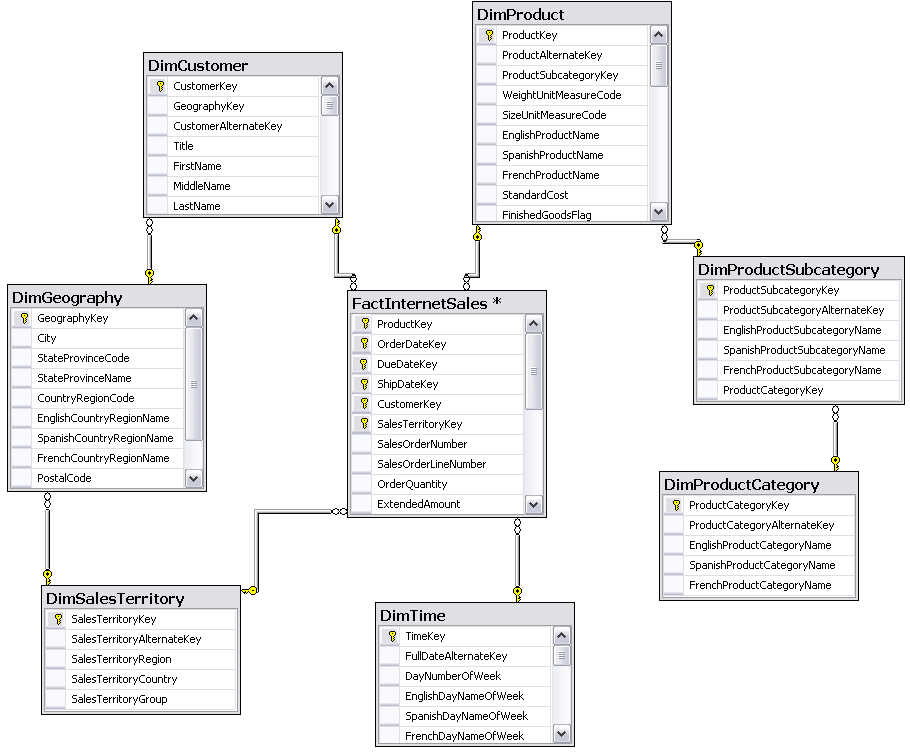


Рисунок 4.2 – Сховище даних

На цій схемі таблиці вимірювань містять інформацію про покупців (DimCustomer), про товари (DimProduct), про місце продаж (DimSalesTerritory), про час продаж (DimTime); консольні таблиці: під категорія товарів (DimProductSubcatecory), категорія товарів (DimProductCategory), узагальнене місце продажів (DimGeography) і таблиця фактів FactInternetSales містіть ключі для зв’язків с таблицямі вимірювань (ProductKey, OrderDateKey, DueDateKey, ShipDateKey, CustomerKey, SalesTerritoryKey), а також самі дані для подальшого аналізу (SalesOrderNumber, SalesOrderLineNumber, OrderQuantity, ExtendedAmount).

4.4 В`ювері для структури інтелектуального аналізу по алгоритму асоціативних правил

Для полегшення аналізу створюються 2 в`ювера vAssocSeqLineItems і vAssocSeqOrders.

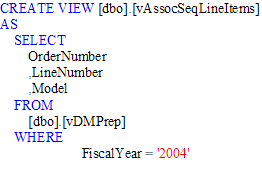


Рисунок 4.4 – SQL- на створення vAssocSeqLineItems

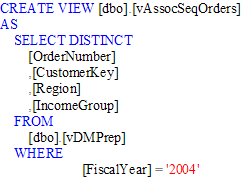


Рисунок 4.5 – на створення vAssocSeqOrders

Ці вьювери создаються на підставі вьювера vDMPrep, який у свою чергу був створений з таблиць сховища AdventureWorks.

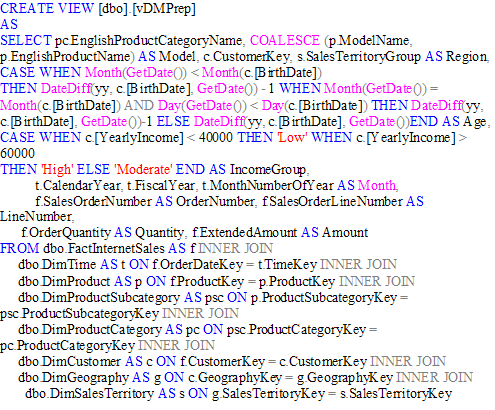


Рисунок 4.6 – на створення vDMPrep

5 РЕАЛІЗАЦІЯ МОДЕЛІ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ В СЕРЕДОВИЩІ MS SQL SERVER 2005

5.1 Принцип роботи з моделлю інтелектуального аналізу по алгоритму асоціативних правил

Для інтелектуального аналізу даних в службах Microsoft SQL Server 2005 Analysis Services використовується два основні об'єкти:

* структура інтелектуального аналізу даних;
* модель інтелектуального аналізу даних.

Останніми об'єктами, що беруть участь в інтелектуальному аналізі даних, є стовпці структури інтелектуального аналізу і стовпці моделі інтелектуального аналізу.

Процес роботи з моделями інтелектуального аналізу полягає в наступному:

а) створення структури інтелектуального аналізу даних;

б) додання моделі;

в) завдання параметрів моделі;

ґ) перегляд моделі;

д) прогнозування моделі.

Структура інтелектуального аналізу — це структура даних, що визначає домен даних, на основі якого будуються моделі інтелектуального аналізу. Одна структура інтелектуального аналізу може містити декілька моделей інтелектуального аналізу даних, що спільно використовують один домен.

Будівельними блоками структури інтелектуального аналізу є стовпці, які описують дані, що містяться в джерелі даних. Ці стовпці містять такі відомості, як тип даних, тип вмісту і способи розподілу даних.

Структура інтелектуального аналізу також може містити певні вкладені таблиці. Вкладена таблиця представляє зв'язок «один до багатьом» між об'єктом варіанту і пов'язаними з ним атрибутами. Наприклад, якщо відомості, що описують клієнта, знаходяться в одній таблиці, а покупки цього клієнта знаходяться в іншій таблиці, то можна використовувати вкладені таблиці для комбінування відомостей в єдиний варіант. Ідентифікатором клієнта є об'єкт, а покупки — пов'язані з ним атрибути.

Модель інтелектуального аналізу даних застосовує алгоритм інтелектуального аналізу до даних, представлених структурою інтелектуального аналізу даних. Модель інтелектуального аналізу даних, як і структура інтелектуального аналізу, містить стовпці. Модель інтелектуального аналізу міститься в структурі інтелектуального аналізу і успадковує всі значення властивостей, визначених цією структурою. Модель може використовувати всі стовпці, що містяться в структурі інтелектуального аналізу даних, або підмножини цих стовпців.

На додаток до параметрів, визначених в структурі інтелектуального аналізу, модель інтелектуального аналізу містить дві властивості: Algorithm і Usage. Параметр algorithm визначений в моделі інтелектуального аналізу, а параметр usage визначений в стовпці моделі інтелектуального аналізу. Опис цих параметрів приводиться нижче:

* «Algorithm». Властивість моделі, що визначає алгоритм, використовуваний для створення моделі. У нашому випадку це алгоритм асоціативних правил.
* «Usage». Властивість стовпця моделі, що визначає те, як стовпець використовується моделлю. Можна визначити стовпці як стовпці введення, ключові стовпці або прогнозовані стовпці.

Модель інтелектуального аналізу даних до обробки є просто порожнім об'єктом. При обробці моделі дані, визначені структурою, обробляються алгоритмом. Алгоритм ідентифікує правила і закономірності в даних, а потім використовує ці правила і закономірності для заповнення моделі.

Можна створювати декілька моделей, заснованих на одній і тій же структурі. Всі моделі, побудовані на основі однієї і тієї ж структури, мають бути засновані на одному і тому ж джерелі даних. Проте моделі можуть розрізнятися по стовпцях структури, способах їх використання, типові алгоритму для створення кожної моделі і параметрах для кожного алгоритму.

Для кожного алгоритму є свій набір параметрів моделі, які не обходжений визначити. Для моделі «Споживчої корзини» це:

* максимальний/мінімальний розмір (max/min itemset size)– кількість товару в корзині;
* максимальна/мінімальна підтримка (max/min support) – кількість спостережень – покупок;
* мінімальна значущість (min importance)– поріг, нижче якого не має сенсу проводити аналіз;
* мінімальна вірогідність (min probability)– вірогідність попадання товару в корзину.

Після обробки моделі її можна проглянути за допомогою призначених для користувача засобів перегляду, що надаються в середовищах Business Intelligence Development Studio і SQL Server Management Studio або шляхом передачі запитів моделі для виконання прогнозів.

Служби Microsoft SQL Server 2005 Analysis Services дозволяють використовувати прогнозуючий запит на мові розширень інтелектуального аналізу даних Data Mining Extensions (DMX) для прогнозування невідомих значень стовпців в новому наборі даних на основі результатів моделі інтелектуального аналізу даних.

5.2 Реалізація моделі за допомогою мови DMX

Data Mining Extensions (DMX) є мовою, яку ви можете використовувати, щоб створити і, працювати з data mining models in Microsoft SQL Server 2005 Analysis Services (SSAS). Ви можете використовувати DMX, щоб створити структуру нового data mining models, щоб тренувати ці моделі, і для перегляду, управління, і прогнозу.

Структура інтелектуального аналізу даних — це структура даних, яка визначає наочну область, на основі якої будується модель інтелектуального аналізу даних. Одна структура інтелектуального аналізу може містити декілька моделей інтелектуального аналізу даних, спільно використовуючи один домен. Модель інтелектуального аналізу даних застосовує алгоритм інтелектуального аналізу до даних, представлених структурою інтелектуального аналізу даних.

Крок 1. Для створення структури інтелектуального аналізу, що містить вкладені таблиці, використовується інструкція CREATE MINING STRUCTURE (розширення інтелектуального аналізу даних). Код інструкції можна розбити на наступні частини:

* привласнення структурі імені;
* визначення ключового стовпця;
* визначення стовпців інтелектуального аналізу даних;
* визначення стовпців вкладених таблиць.

На рисунку показана інструкція DMX на створення структури «Market Basket»

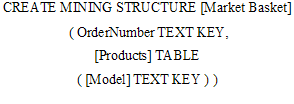


Рисунок 5.1 – Створення структури

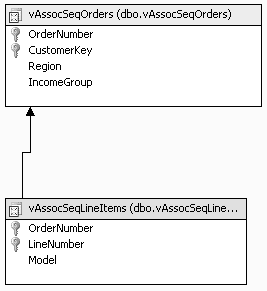


Рисунок 5.2 – Структура інтелектуального аналізу

Крок 2. На наступному етапі необхідно додати нову модель інтелектуального аналізу до структури інтелектуального аналізу «Споживацька корзина», заснованої на алгоритмі взаємозв'язків Microsoft, і змінити значення за умовчанням для параметра MINIMUM\_PROBABILTY на 0.2. Зміна цього параметра приведе до створення більшої кількості правил алгоритмом взаємозв'язків Microsoft.

Інструкція ALTER MINING STRUCTURE (розширення інтелектуального аналізу даних) використовується для додавання до структури інтелектуального аналізу моделі інтелектуального аналізу, що містить вкладену таблицю. Код інструкції можна розбити таким чином:

* визначення структури інтелектуального аналізу даних;
* вказівка імені моделі інтелектуального аналізу;
* визначення ключового стовпця;
* визначення стовпців початкових даних і прогнозованих стовпців;
* визначення стовпців вкладених таблиць;
* ідентифікація алгоритму і змін параметра.

На рисунку показана інструкція DMX на додання моделі «Modified Assocation» до структури «Market Basket»:

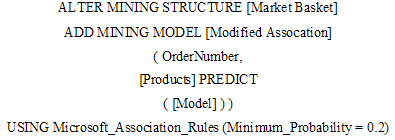


Рисунок 5.3 – Додання моделі до структури

Крок 3. На цьому занятті за допомогою інструкції INSERT INTO (розширення інтелектуального аналізу даних) і представлень vAssocSeqLineItems і vAssocSeqOrders із зразка бази даних AdventureWorksDW обробляються структури і моделі інтелектуального аналізу даних, створені Крок 1. Створення структури інтелектуального аналізу «Споживацька корзина» і Крок 2. Додавання моделей інтелектуального аналізу до структури інтелектуального аналізу «Споживацька корзина».

Код інструкції можна розбити на наступні частини:

* визначення структури інтелектуального аналізу даних;
* список стовпців структури інтелектуального аналізу даних;
* визначення повчальних даних за допомогою інструкції SHAPE.

На рисунку показана інструкція DMX на обробку структури і моделі:

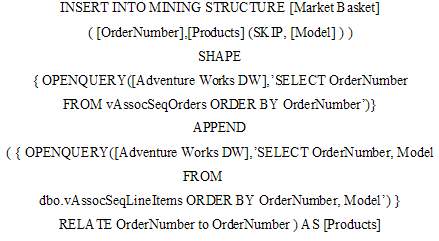


Рисунок 5.4 – Обробка структури і моделі

Після перевантаження модель підключається в Biseness Intelligence і виглядає таким чином:

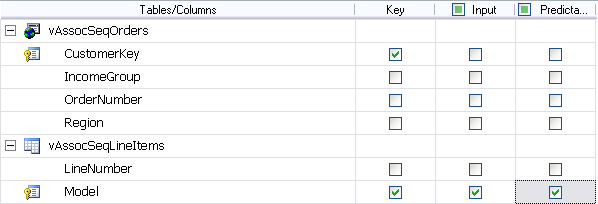


Рисунок 5.5 – Модель інтелектуального аналізу

Крок 4. На цьому етапі буде створено прогноз, заснований на моделі інтелектуального аналізу даних, якій був доданий в структуру «Споживацька корзина».

На рисунку показаний приклад інструкції прогнозування, у якій задана кількість та перелік товарів, що знаходяться в корзині (Mountain Bottle Cage, Mountain Tire Tube, Mountain-2000) і здійснюється прогнозування товарів, що може бути придбана разом з ними.

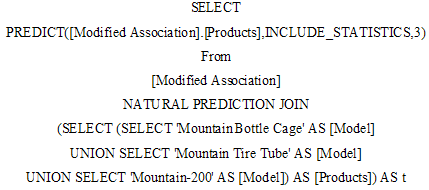


Рисунок 5.6 – Створення прогнозу

Результат візуалізації моделі має 2 вигляду: аналітичний (рисунок 5.7), «дерево» (рисунок 5.8).

Розглядаючи аналітичний вигляд, ми можемо визначити вірогідність попадання товару в корзину, якщо там вже є який-небудь товар.

Наприклад: при покупці товару Touring Tire Tube з вірогідністю 54,3% буде куплений товар Touring Tire. А при покупці Touring-1000 і Touring Tire Tube, товар Touring Tire буде преобретен з вірогідністю 100%.

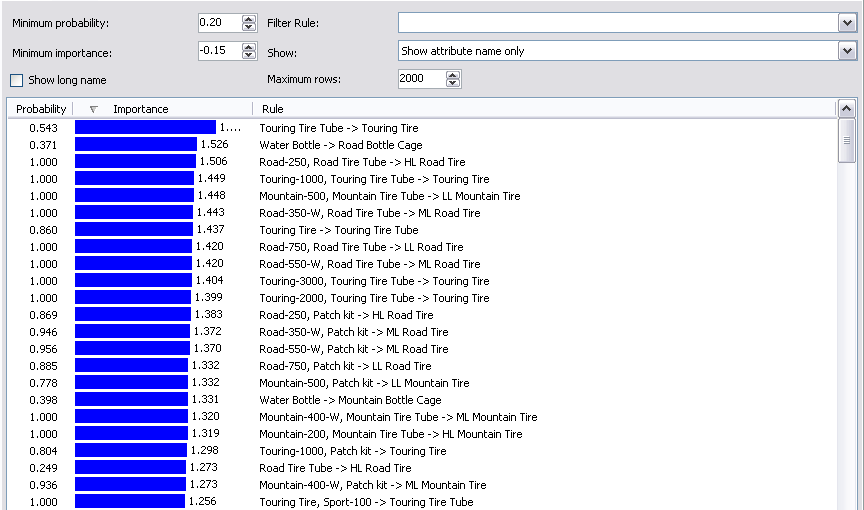


Рисунок 5.7 – Аналітичний вигляд

При візуалізації моделі в вигляді «дерево» ми можемо проглянути взаємозв'язок товарів при покупці.

Наприклад при покупці Mountain-500 буде куплений Sport-100. В зворотну сторону цей вислів не вірний.

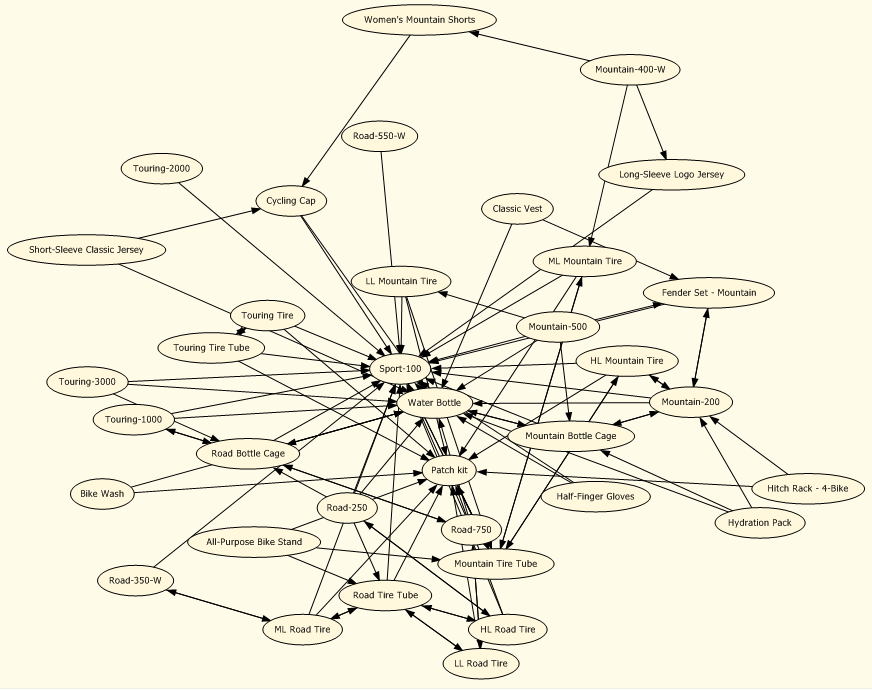


Рисунок 5.8 – Вигляд «дерево»

6 РЕАЛІЗАЦИЯ ІНТЕРФЕЙСУ КОРИСТУВАЧА ДЛЯ РОБОТИ З МОДЕЛЛЮ І ЙОГО ВЕРИФІКАЦІЯ

Програмний продукт є простим у використанні, тому з ним легко може працювати навіть малокваліфікований користувач. Для повноцінної роботи програми необхідно встановити на робочому комп'ютері платформу Visual Studio 2005 та MS SQL Server 2005.

Для безпосереднього початку роботи із програмою необхідно запустити додаток «AdventureWorksIn». При запуску з'явиться головна форма програми:



Рисунок 6.1 – Головна екрана форма

Крок 1. Після натискання кнопки «Работа с данными» відкривається вікно, в якому можливо проглянути інформацію про наявні замовлення:



Рисунок 6.2 – Вікно замовлень

або створити нове замовлення:

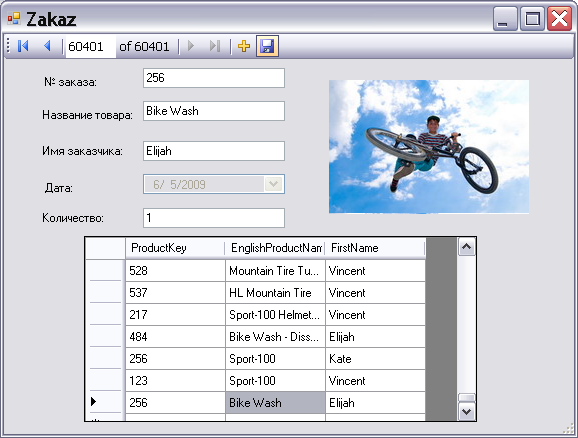


Рисунок 6.3 – Вікно нового замовлення

Крок 2. Після натискання кнопки «Создание модели» відкривається вікно, в якому необхідно ввести назву нової структури, нової моделі і 2 параметри («Минимальная вероятность» та «Минимальная поддержка»), останні параметри стоять за умовчанням:

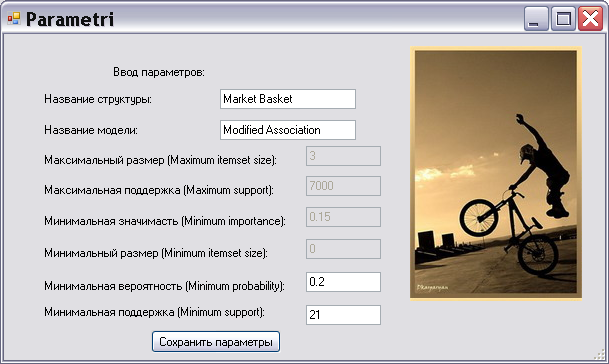


Рисунок 6.4 – Створення моделі

Потім натиснути кнопку «Сохранить параметры», після чого у базі будить створена модель з необхідними параметрами.

Крок 3. Для того щоб вибрати необхідний нам товар, треба натиснути кнопку «Прогнозирование». У вікні, що з'явилося, вибираємо 3 товари, які нам потрібні і вводимо назву моделі. Потім натискаємо кнопку «Сохранить параметры».

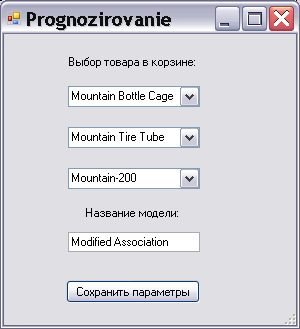


Рисунок 6.5 – Вибір товарів

Після чого з'являється результат прогнозування, в якому показані товари, які найбільш часто продаються з товарами, вибраними раніше.

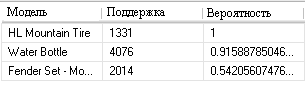


Рисунок 6.6 – Результат прогнозування7 ОХОРОНА ПРАЦІ

7.1 Стисла характеристика об'єкта дослідження з питань охорони праці

# Об’єктом дослідження є офіс бухгалтерського відділу ТОВ «ТБ - Молдавкабель», яке виконує основну діяльність:

згідно з кодом галузі за ЗКГНГ – 71130

згідно з кодом за КВЕД :

51.70.0 – «Інші види оптової торгівлі» (23 клас проф. ризику)

Основна робота зв'язана з розумовою працею, тому відповідно до гігієнічної класифікації праці - праця на даному робочому місці відноситься до категорії 1*б*.

Це підприємство зареєстровано в Приморському відділу ФССНВВ і ПЗУ м. Одеса та віднесено до 23 -го класу професійного ризику.

Назначен страховий тариф СТ 1,16% % від фонду оплати праці. Фонд оплати праці за 2008 рік склав 5\*12\*3500=42000 тис. грн ;

страховий внесок – 1,16\*42000 /100 = 487,2 тис. грн.

Середньосписочна чисельність персоналу в еквіваленті повної зайнятості на кінець 2008 р. – 5 чоловік.

Робітники основного виробництва – 3 чол.

Робітники допоміжного виробництва – 1 чол.

Фахівці і службовці – 1 чол.

Служба охорони праці на об’єкті відсутня. Відповідно до пп. НПАОП 0.00-4.35-04 “Типове положення про службу охорони праці” – функції служби охорони праці виконує в порядку сумісництва (суміщення) особа с відповідної підготовкою.

Наявність робіт підвищеної небезпеки згідно з переліком ДНАОП 0.00-8.02-93 – відсутня.

Наявність таких робіт, де необхідне проведення професійний відбір згідно з ДНАОП 0.03-8.06-94 – відсутня.

Організація та проведення попередніх та періодичних медичних оглядів згідно з ДНАОП 0.03-4.02-94

Попередній медичний огляд при пристрої на роботу на об’єкті проходили всі робітники організації. Періодичний медичний огляд відповідно до ДНАОП 0.03-4.02-94 проводиться 1 раз на 6 місяців, останні 1 раз в 2 роки.

Відповідно до пп. 5.1.“Типове положення про порядок проведення навчання і перевірки знань з питань охорони праці” виконується проведення попередніх та періодичних (1 раз у три році) навчань посадових осіб і спеціалістів з питань охорони праці. Постійно діючих комісій з перевірки знань немає. Мається протокол перевірки знань. Час останньої перевірки знань – 02.2008 року. Наступна перевірка – 02.2010(2009).

За весь період існування підприємства не було зареєстровано жодного нещасного випадку та професійного захворювання. Відповідно НПАОП 00.0-6.02-04 взято на облік:

|  |  |
| --- | --- |
| Акти | Знаходиться на обліку |
| Акт н/в Н-1 | 0 |
| Акт розслідування н/в Н-5 | 0 |
| Акт П-4 | 0 |
| Акт НПВ | 0 |

Планові перевірки службами Держгірпромнагляд, Державний санітарно-епідеміологічний нагляд МОЗ, Держпожнагляд МНС України – проводяться раз у три роки. За 2008 р. була здійснена дві оперативна перевірка Держпожнаглядом ГУ МНС України в Одеській області та Держгірпромнаглядом.

7.2 Виявлення і аналіз небезпечних і шкідливих виробничих чинників в даній робочій зоні

* + 1. Аналіз стану повітряного середовища

Приміщення з ЕОМ обладнані системами опалювання, кондиціонування повітря або приточування-витяжною вентиляцією відповідно до СНіП 2.04.05-91 "Опалювання, вентиляція і кондиціонування". Параметри мікроклімату, іонного складу повітря, зміст шкідливих речовин на робочих місцях, оснащених відеотерміналами, відповідають вимогам ДСН 3.3.6-042-99 «Державні санітарні норми мікроклімату виробничих приміщень».

ГОСТ 12.1.005-88 "ССБТ. Загальні санітарно-гігієнічні вимоги до повітря робочої зони", СН 2152-80 "Санітарно-гігієнічні норми допустимих рівнів іонізації повітря виробничих і суспільних приміщень".

Для підтримки допустимих значень передбачені установки кондиціонування повітря.

## Таблиця 7.1 – Параметри мікроклімату для приміщень з ВДТ і ПЕВМ

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Пору року | Категорія робіт у відповідності з ГОСТ 12.1-005 | Темп-ра повітря °С, оптим. | Відносить. вогкість повітря %  оптимальна | Швидкість руху повітря, м/с, оптимальна | Факт.  знач. темп. повітря. °С |  |
| Холодне | легка - 1а | 22–24 | 40–60 | 0,1 | 17–21 |
| легка – 1б | 21–23 | 40–60 | 0,1 | 17–21 |
| Тепле | легка - 1а | 23–25 | 40–60 | 0,1 | 20–26 |
| легка – 1б | 22–24 | 40–60 | 0,2 | 20–26 |

Іонізація повітря в робочому приміщенні не відповідає СН 2152–80 «Санітарно-гігієнічні норми допустимих рівнів іонізації повітря виробничих і суспільних приміщень».

7.2.2 Аналіз виробничого освітлення

В даному приміщенні використано суміщене освітлення. Природне освітлення бічне, одностороннє; розташовано 2 бічні світлові отвори з подвійним склінням. На робочому місці в світлий час доби коефіцієнт природної освітленості (КЕО) > 1,5 %. Штучне освітлення виконано у вигляді загального освітлення. В темний час доби рівень освітлення не достатній, оскільки він не відповідає ДБН В.2.5-28-2006 Державні будівельні норми України «Пріродне і штучне освітлення» і ДСанПіН 3.3.2.-007-98. Рівень освітленості робочої поверхні при роботі з документами повинен складати 300–500 лк. Фактична освітленість в робочому приміщенні рівна 350 лк.

7.2.3 Виявлення і аналіз наявності шуму, вібрації, інфра- і ультразвуку

В приміщеннях з ЕОМ рівні звукового тиску, рівні звуку і еквівалентні рівні звуку на робочих місцях відповідають вимогам ГОСТ 12.1.003 ССБТ "Шум. Загальні вимоги безпеки", ДСН 3.3.6-037-99 «Державні санітарні норми виробничого шуму, ультразвуку та інфразвуку» СН 3223-85 "Санітарні норми допустимих рівнів шуму на робочих місцях", ГР № 2411-81 "Гігієнічні рекомендації по встановленню рівнів шуму на робочих місцях з урахуванням напруженості і тяжкості праці". Рівні шуму на робочих місцях осіб, що працюють з відеотерміналами і ЕОМ, визначені ДСанПіН 3.3.2-007-98. Для забезпечення нормованих рівнів шуму у виробничих приміщеннях і на робочих місцях застосовані шумопоглинаючі засоби, вибір яких був визначений спеціальним інженерно-акустичним розрахунком.

Рівні вібрації при виконанні робіт з ЕОМ у виробничих приміщеннях не перевищують допустимих значень, вказаних в ГОСТ 12.1.012-90 «Вібраційна безпека. Загальні вимоги», ДСН 3.3.6-039-99 «Державні санітарні норми виробничої загальної та локальної вi6paції», СН 3044-84 "Санітарні норми вібрації робочих місць" і ДСанПіН 3.3.2-007-98. Інфра- і ультразвук відповідають відповідно ДСН 3.3.6-037-99, СН 22-74-80 і ГОСТ

7.2.4 Аналіз виробничого шуму

Таблиця 7.2 – Рівні звукового тиску

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № | Робочі місця | Рівні звукового тиску, дБ, в октавних  смугах, з среднегеометрическими частотами, Гц | | | | | | | | Рівні звуку  дБ |
| 63 | 125 | 250 | 500 | 1000 | 2000 | 4000 | 8000 |
| 1 | Приміщення конструкторських бюро розраховувачів  Програмно – обчислювальних  машин, лабораторій для теоретичних робіт і  обробки експериментальних  даних | 71 | 61 | 54 | 49 | 45 | 42 | 40 | 38 | 50 |
| 2 | Фактичні значення | 70 | 60 | 53 | 48 | 44 | 41 | 40 | 37 | - |

Вібрація в робочій зоні відповідає ГОСТ 12.1.012-90.

7.2.5 Виявлення і аналіз рівнів неіонізуючих електромагнітних випромінювань, електростатичних і магнітних полів

Наявні 5 відео дисплейних терміналів (22 LCD монітори” Samsung ) знаходяться в експлуатації з 12.2008 року не повинні проходити оцінку (експертизу) його безпеки і нешкідливості для здоров'я людини. 22 LCD монітори” Samsung WideScren - мають державну сертифікацію системи УкрСЕПРО і екологічні мітки MPR II, TCO 03, що свідчать про виконання рекомендацій MPR II Swedac, мітку TUV – німецькою експертною організацією TUV і був випробуваний на відповідність стандарту по безпеці EN60950:1992 (IEC950), основному стандарту по ергономіці ZH 1/618 (німецький стандарт), стандарту по випромінюваннях MPR II і стандарту ISO9241/3.

Рівні електромагнітного випромінювання і магнітних полів відповідають вимогам ГОСТ 12.1.006-84 «ССБТ. Електромагнітні поля радіочастот. Допустимі рівні на робочих місцях і вимоги до проведення контролю», СН №3206-85 «Гранично допустимі рівні магнітних полів частотою 50 Гц», ДСанПін 3.3.2-007-98 і ДНАОП 0.00-1.31-99 п. 2.2.1, 3.7.

Рівні інфрачервоного випромінювання не перевищують граничних у відповідності з ГОСТ 12.1.005-88 і СН № 4088-86 значень з урахуванням опромінюваної площі тіла і ДСанПіН 3.3.2-007-98.

Рівні ультрафіолетового випромінювання не перевищують допустимих відповідно до СН № 4557-88 «Санітарні норми ультрафіолетового випромінювання у виробничих приміщеннях» і ДСанПіН 3.3.2-007-98.

Гранично допустима напруженість електростатичного поля на робочому місці не перевищує рівнів, приведених в ГОСТ 12.1.0045 «ССБТ. Електростатичні поля. Допустимі рівні на робочих місцях і вимоги до проведення контролю», СН № 1757-77 «Санітарно-гігієнічні норми допустимої напруженості електростатичного поля» і ДСанПіН 3.3.2-007-98.

7.2.6 Аналіз електробезпеки

При проектуванні систем електропостачання, монтажу силової електроустаткуванні і електричного освітлення будівлі і приміщення для ЕОМ соблюдені вимоги ПУЕ, ПТЕ, ПБЕ. СН 357-77 "Інструкція по проектуванню силового освітлювального устаткування промислових підприємств", ГОСТ 12.1.006. ГОСТ 12.1.030 "ССБТ. Електробезпека. Захисне заземлення, занулення", ГОСТ 12.1.019 "ССБТ. Електробезпека. Загальні вимоги і номенклатура видів захисту", ГОСТ 12.1.045, ВСН 59-88 Держкомархітектури СРСР "Електроустаткування житлових і суспільних будівель. Норми проектування", Правил пожежної безпеки в Україні, ДНАОП 0.00-1.31-99, розділів СНіП, що стосуються штучного освітлення і електротехнічних пристроїв, а також вимог нормативно-технічної і експлуатаційної документації заводу - виготівника ЕОМ.

Лінія електромережі для живлення ЕОМ, периферійних пристроїв ЕОМ виконана як окрема групова трьохдротяна сіть. Нульовий захисний провідник прокладений від стійки групового розподільного щита, розподільного пункту до розеток живлення.

Заземлення відповідає вимогам ДНАОП 0.00-1.21-98 "Правила безпечної експлуатації електроустановок споживачів".

7.2.7 Аналіз вибухопожежобезпеки

Категорія приміщення в яких експлуатуються відеотермінали та ЕОМ визначена категорія «Д» з вибухопожежної і пожежної безпеки відповідно до ОНТП 24-86 «Визначення категорій приміщень і будівель по вибухопожежної і пожежна небезпека», та клас зони П-IIа згідно з ПВЕ. Відповідні позначення нанесені на вхідні двері приміщення.

Будівлі і ті їх частини, в яких розташовуються ЕОМ, мають II ступень вогнестійкості. Приміщення відокремлені від приміщень іншого призначення протипожежними стінами, то межа їх вогнестійкості визначена відповідно до СНиП 2.01.02-85.

Для гасіння пожеж призначені у кожному будинку ділянки - внутрішній протипожежний водопровід згідно з СНиП 2.04.01-85 «Внутрешній водопровід і каналізація будівель», та СНиП 2.04.02-84 «Водопостачання. Поверхневі мережі і споруди».

Приміщення з ЕОМ оснащені системою автоматичної пожежної сигналізації відповідно до вимог переліку однотипних за призначенням об’єктів, які підлягають обладнанню автоматичними установками пожежогасіння та пожежної сигналізації, затвердженого наказом МВС України від 20.11.97 № 779, та СНиП 2.04.09-84 «Пожежна автоматика будівель и споруди» з димовими пожежними сповіщувачами та переносними вуглекислотними вогнегасниками ОУ-2 (ТУ У 29.2-13485476-012-2003) з розрахунку 2 шт, на кожні 20 м2 площі приміщення і відповідають ДСТУ 3675-98, ДСТУ 3734-98. Наступна дата перезаряджування –2009 р. Годин до гасіння пожеж класу «В» відповідно ГОСТ 27331-87.

В інших приміщеннях встановлені теплові пожежні сповіщувачі.

7.2.8 Ергономічні характеристики робочого місця

# Робочі місця (5 РМ) користувачів ВДТ і ПЕВМ і їх взаємне розташування відповідають п. 4 ДНАОП 0.00-1.31-99 «Правила охорони праці при експлуатації ЕОМ», ГОСТ 21958–76 «Система «людина-машина. Зал і кабіни операторів. Взаємне розташування робочих місць. Загальні ергономічні вимоги» і ГОСТ 21839–76 «Система «Людина-машина. Крісло людини-оператора. Загальні ергономічні вимоги». Крісло користувача - модель «Bridge Chrome», стіл комп`ютерний відповідають ГОСТ 12.2.032-78. ССБТ «Робоче місце при виповнені сидячих робіт. Загальні ергономічні вимоги».

7.3 Розрахунок часу евакуації першого поверху, дев’яти поверхової офісноїбудови і запобігання від нещасних випадків при пожежі

Час евакуації:

tp = t0 + t1 + t2 + t3 +...+ tі , (7.1)

де t0 –час руху людського потоку на першій ділянці, хвилин;

tі – на послідуючих етапах руху.

Час руху першого поверху, дев’яти поверхової офісної будови ділиться на дві ділянки руху: по коридору, тамбур виходу на вулицю,

tp = t0 + t1 + t2 (7.2)

Необхідний час евакуації людей можна визначити по формулі:

tнб = 0,115 \*w1/3 хв, (7.3)

де w1/3- об 'єм приміщення, м3 .

tнб = 0,115\* (7200)1/3 = 2,22 хв.

Щільність людських потоків:

* кількість людей на одиницю площі пола (на 1 м2) :

D = N \* f / δ\*L чол / м2 , (7.4)

де N - кількість людей в потоці;

δ і L – ширина і довжина потоку в м;

f - площа горизонтальної проекції одного чоловіка в м2 ;

N = 25 чоловік;

δ = 3,0 м;

L = 31 м;

f = 0,113;

D = (25\* 0,113) / (3,0 \*31) = 0.0304 чол / м2 ;

* площа пола яка відводиться для одного чоловіка

D = δ\*L/N чол / м2; (7.5)

D = (3,0 \*31) / 25 = 3,72 чол / м2;

* відношення суми горизонтальних проекцій людей до площі пола, займає - мого потоком :

D = Σ f / δ \*L м2 / м2; (7.6)

Приміряється при любому составі потоку ця формула.

Σ f = N \* f м2; (7.7)

Максимальна щільність, установлена на моделі і перевірена в натуральних умовах, виявилась рівною:

D = Σ f / δ \*L = 0.0304.

При визначенні значення максимальної щільності прийнято предположення, що еліпс, вражаючий горизонтальну проекцію чоловіка, не піддержується деформації в час стиснення потоку. Ця умова в подальшому покладена в основу методики розрахунку. В дійсності, оскільки людське тіло упруге , то при значному стисненні змінюється форма і зменшується площа його горизонтальної проекції, внаслідок чого фізичний ліміт щільності Dфп більше , чим 0,0304.

В практичних умовах щільність потоку має величини від близькою до нуля і закінчує максимальною. І частота повторення щільності далеко не однакова.

7.3.1 Пропускна можливість шляху і інтенсивність руху людських потоків

Величиною, пов’язуючих параметри руху: щільність D, швидкість ν і параметри шляху δ, являються пропускною можливістю Q. Q - кількість людей приходящих в одиницю часу через шлях шириною δ:

Q = D \*ν \*δ м2/хв.; (7.8)

Q = 0,0304 \* 78,35 \* 3,0 = 7,15 м2/хв.;

Добуток щільності і швидкості:

q = D \*ν м/хв. (7.9)

Називається q інтенсивністю руху, та як його значення не залежить від ширини шляху, характеризують кінетику процесу руху людського потоку. Значення інтенсивності руху відповідає значенням пропускної можливості шляху шириною 1м.

При визначенні щільності, різної для кожного виду шляху, q досягається максимуму qМакс, а потім падає. Звідси слідує важний висновок, що горизонтальні і похилі шляхи руху, а також отвори мають границю пропускної здібності, визначає мий щільністю при qМакс. Ця закономірність має важливе значення, оскільки щільність , перевищуюча щільність при qМакс,  визивають затримку рухуі скопичення людей на тих ділянках шляху, де ця границя перевищена формулами:

* для горизонтальних шляхів:

q = D\*ν м/хв; (7.10)

q = 0,0304\*78,35 = 2,38 м/хв.

* для отворів:

q о=Dо\*νо м/хв; (7.11)

q о= 0,0304 \* 91,66 = 2,79 м/хв.

* по драбині вниз:

q ↓ = D↓\*ν↓ м/хв; (7.12)

q ↓ = 0,019\*58,75 = 1,11 м/хв.

Розрахунок руху потоку першого поверху.

Час руху з кімнат до коридору:

D0 = (5 \* 0,113) / (1,2 \*8,75) = 0,04 чол / м2 ; (7.13)

Відповідно:

q0 = 2,89 м/хв ;

ν0 = 72,31 м/хв ;

δ1= 1,2 м;

L0 = 8,75 м.

Q0 = q0\*δ0 м2 / хв; (7.14)

Q0 = 1,2 \*2,89 = 1,6 м2 / хв;

Поток досягає коридора:

t0 = L0/ν0 хв; (7.15)

t0 = 8,75/72,31 = 0,12 хв;

Тоді, на другому поверсі розміщено 7 кабінетів і час евакуації людей до коридору складе:

t0 = 0,12\* 7 = 0,84 хв.

Час руху замикаючої частини потоку по коридору:

t1= L/ν1 хв; (7.16)

t1= 60/78,35 = 0,76 хв.

D1= 0.019;

q1= 1,57 м/хв ;

ν1= 78,35 м/хв ;

δ1= 2,5 м.

Кількість людей приходящих в одиницю часу через шлях шириною δ1:

Q1 = q1\*δ1чол.; (7.17)

Q1 = 2,5 \* 1,57 = 3,92 чол.

Отвір 01:

q01= q1\* δ1 / δ01 м/хв; (7.18)

q01= 1,57\* 2,5/1,8 = 2,18 м/хв;

q01> qМакс (Затримка руху перед отвором).

Відповідно:

D01= 0.92;

q01= 10,50 м/хв ;

ν01= 11,42 м/хв ;

δ01= 1,8 м.

Q01 = q01\*δ01 м2 / хв; (7.19)

Q01 = 10,50\*1,8 = 18,9 м2 / хв.

Швидкість накопичення людей визначається:

ν'с1=(q01 \* δ01 / δ1 - q1) / (D01-D1) м/хв; (7.20)

ν'с1=(10.50\*01.8/2.5 – 1,57) / (0.92-0.019) = 6,6 м/хв;

Швидкість розсмокчування накопичення людей:

νс1= ν01 \* δ01 / δ1 м/хв; (7.21)

νс1= 11,42 \* 1,8/2,5 = 8,22 м/хв.

Відрізок ділянки , на якому розповсюджується накопичення:

l c1= N/Q1 \* (ν'с1\*ν1)/ (ν'с1+ν1) м; (7.22)

l c1= 2,82/3,92 \* ( 6,6 \* 78,35) / ( 78,35 +6,6) = 4,32 м.

Максимальна кількість накопичуючих перед отвором:

Nc1= Dmax\* δ1 \*l c1  м2; (7.23)

Nc1= 0,92\*2,5\*4,32 = 9,93 м2.

Час затримки:

τ1= N (1/Q01 – 1/Q1) хв; (7.24)

τ1= 25\*0,113( 1/18,9 – 1/3,92) = 0,005 хв.

Час затримки через отвір:

t01= N/Q01 = 2,82/ 18,9= 0,15 (7.25)

t1+ τ1= 0,76+ 0,005= 0,77 хв.

По драбині вниз:

q01= q1\* δ01 / δ1 м/хв; (7.26)

δ1 =2,4 м;

q01= 10,50\* 1,8/2,4=7,87 м/хв ;

q01 > qМакс (Затримка руху).

Відповідно:

D2 = 0.92;

q2 = 5,38 м/хв;

ν2 = 5,86 м/хв;

δ2 = 2,4 м.

Q2 = q2\*δ2 м2 / хв; (7.27)

Q2 = 5,38\*2,4 = 21,91 м2 / хв.

Швидкість накопичення людей визначається:

ν'с2 = (q2 \* δ1 / δ02 - q1) / (D02-D2) м/хв; (7.28)

ν'с2 = (5,38\*1.8/2.4 – 1,57) / (0.92-0.019) = 2,7м/хв.

Швидкість розсмокчування накопичення людей:

νс2= ν2 \* δ1 / δ2 м/хв; (7.29)

νс2 = 5,86 \* 1,8/2,4 = 4,39 м/хв.

Відрізок ділянки , на якому проходить накопичення:

l c2 = N/Q2 \* (ν'с2\*ν2)/ (ν'с2+ν2) м; (7.30)

l c2 = 2,82/21,91 \* (2,7 \* 5,86) / (2,7 + 5,86) = 0,23 м.

Максимальна кількість людей накопичених на драбині:

Nc2= Dmax\* δ2 \*l c2  м2; (7.31)

Nc2 = 0,92 \* 2,4\* 0,23 = 0,5 м2.

Час руху по драбині вниз:

t2= L2/ν2 хв; (7.32)

t2= 8/5,86 = 1,36 хв.

Розрахунок руху потоку від місця злиття з потоком першого поверху до виходу на вулицю:

q3.1= q2\* δ02 / δ03 м/хв; (7.33)

δ03 = 3 м;

q3.1= 7,87\* 4/3= 10,4 м/хв.

Відповідно

D3= 0.43;

ν3= 23,43 м/хв ;

Q3.1 = q3.1\*δ03 м2 / хв; (7.34)

Q3.1 = 10,4\*3 = 31,48 м2 / хв.

Час руху потоку другого поверху по вестибюлю до виходу на вулицю:

t3 = L3/ν3 хв; (7.36)

t3= 2,6/23,43 = 0,11 хв.

В тамбур виходить з першого поверху 30 людей, таким чином потік другого поверху зливається:

q3.2= q1.1\* δ02 / δ03 м/хв; (7.37)

q3.2= 6,99\*4 /3= 9,79 м/хв.

Відповідно:

D3.2 = 0,4

ν3.2= 24,48 м/хв.

Час руху потоку другого поверху по вестибюлю до виходу на вулицю:

t4= L4/ν3.2 хв; (7.38)

L4= 8,75 м;

t4= 8,75 /24,48 = 0,35 хв.

Час евакуації :

tp =0,87+ 0,77+1,36+0,11+0,35= 3,46 хв.

Він перевищує необхідний час евакуації на 1,42 хв, тому необхідно щоб люди експлуатувались в строгому визначеному порядку, який вказаний на плані евакуації приміщення. Потрібно, щоб шлях евакуації не загромаджувався меблями, і якщо люди будуть евакуюватися по строго визначеній схемі евакуації й використовувати допоміжні сходи і драбини які призначенні для пожежі, то можливо запобігти від нещасних випадків при пожежі.

ВИСНОВКИ

Розглянутий алгоритм пошуку асоціативних правил є типовою ілюстрацією завдання аналізу купівельної корзини. В результаті її рішення визначаються набори товарів, що часто зустрічаються, а також набори товарів, що спільно набувають покупцями. Знайдені правила можуть бути використані для вирішення різних завдань, зокрема для розміщення товарів на прилавках магазинів, надання знижок на пари товарів для підвищення об'єму продажів і, отже, прибули і інших завдань.

У роботі була побудована модель інтелектуального аналізу даних по алгоритму асоціативних правил. В ході побудови моделі були визначені можливі подальші покупки клієнтів, ґрунтуючись на інформаціях про позиції товару, що вже знаходяться в його корзині.

Для роботи з моделлю було створено інформаційне сховище, розроблені процедури перевантаження даних з транзакційної бази даних в інформаційне сховище. Інформаційне сховище можна використовувати не лише для інтелектуального аналізу, але і для подальшого OLAP-аналізу. Розроблений призначений для користувача інтерфейс, що дозволяє легко створювати, змінювати модель і міняти параметри.

Візуалізація моделі здійснюється в Biseness Intelligence Studio, представляючи аналітичний і графічний вигляд моделі. Цінність полягає в спільному використанні різних служб MS SQL Server 2005:

* Analysis Services;
* Integration Services;
* язык запросов DMX;
* Biseness Intelligence Studio.

В ході проведеної роботи мета була досягнута, а всі поставлені завдання вирішено.

СПИСОК ЛІТЕРАТУРИ

1. Брайан Ларсон. Разработка бизнес – аналитики в Microsoft SQL Server 2005. – СПб.: Питер**,** 2008. – 688 с.
2. Каленик А.И. Использование новых возможностей Microsoft SQL Server 2005. – СПб.: Питер,2006 – 334 с.
3. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных: OLAP и Data Mining. – СПб: БХВ – Петербург, 2004. – 336 с.
4. Введение в анализ ассоциативных правил – Доступно з: <http://www.basegroup.ru/library/analysis/association\_rules/intro/>
5. Дейт К. Дж Введение в системы баз данных, 8-е издание. – М.: Издательский дом "Вильямс", 2005. – 1328 с.
6. Основы баз данных: курс лекций: учеб. пособие / С.Д. Кузнецов. – М.: Интернет–Ун–т Информ. Технологий, 2005. – 488с.
7. С. Я. Архипенков, Д. В. Голубев, О. Б. Максименко. Хранилища данных. От концепции до внедрения. – М.: Диалог-МИФИ, 2002. – 528 с.
8. Брюс Эккель. Философия JAVA, Библиотека программиста. 3 – е изд. – СПб.: Питер, 2003 г. – 638 с.
9. Microsoft SQL Server 2005 Analysis Services.OLAP и многомерный анализ данных. Под общей редакцией А.Бергера, И.Горбач. – СПб.: БХВ – Петербург,2007 – 908 с.
10. Maclennan J., Tang Z. Data Mining With SQL Server 2005. – Indianapolis.: Wiley, 2005 – 296 с.
11. Ульман Д., Уиндом Д. Введение в системы баз данных. – М.: Лори, 2000 – 1328 с.
12. Хансен Г., Хансен Г. Базы данных: разработка и управление. – М.: БИНОМ, 1999 – 296 с.
13. Наталия Елманова, Алексей Федоров. Введение в OLAP – технологии Microsoft. – М.: Диалог – Мифи, 2002 – 272 с.
14. Маклаков С.В. Проектирование реляционных хранилищ данных – М.: Диалог – Мифи, 2007 – 333 с.
15. Гаврилова Т.А., Хорошевский В.Ф. Базы данных интеллектуальных систем. – СПб.: Питер,2001 – 384 с.
16. Башмаков А.И., Башмаков И.А. Интеллектуальные информационные технологии: Учеб. пособие. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2005. – 304 с.
17. Чубукова И.А. Data Mining: БИНОМ. Лаборатория знаний, Интернет-университет информационных технологий, 2008 – 382 с.
18. Офіційний сайт Microsoft – msdn.microsoft.com