

МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В ЕКОНОМІЦІ

УДК 005

ВИКОРИСТАННЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ У СИСТЕМАХ ЕЛЕКТРОННОГО ДОКУМЕНТООБІГУ

APPLICATION OF INTELLIGENT DATA ANALYSIS IN DIGITAL DOCUMENT MANAGEMENT

Беккауер А.О.

асистент кафедри комп'ютерних наук,
Таврійський державний агротехнологічний університет

У статті розглянуто можливості інтелектуального аналізу даних в цілях оптимізації робочого процесу систем електронного документообігу. Описуються методи інтелектуального аналізу даних, завданнями яких є вилучення, накопичення і обробка знань, сценаріїв, моделей поведінки інформаційних систем організацій в розглянутій предметній області. Визначаються методи інтелектуального аналізу даних, які найбільш доцільно використовувати у системах електронного документообігу.

Ключові слова: інтелектуальний аналіз даних, система електронного документообігу, інформаційна система, база знань.

В статье рассмотрены возможности интеллектуального анализа данных в целях оптимизации рабочего процесса систем электронного документооборота. Описываются методы интеллектуального анализа данных, задачами которых являются извлечение, накопление и обработка знаний, сценариев, моделей поведения информационных систем организаций в рассматриваемой предметной области. Определяются методы интеллектуального анализа данных, которые наиболее целесообразно использовать в системах электронного документооборота.

Ключевые слова: интеллектуальный анализ данных, система электронного документооборота, информационная система, база знаний.

The paper presents the possibilities of intelligent data analysis with the aim of optimizing the working process of document management systems. The methods of intelligent data analysis are described with the focus on the tasks such as retrieval, accumulation and processing of knowledge, scenarios, behavior models of information systems in the field of study. The methods of intelligent data analysis which are most practical for application in digital document management are determined.

Keywords: data mining, electronic document management system, information system, knowledge base.

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями. Для сучасних інформаційних систем характерні величезні обсяги даних і стрибкоподібні зміни або невизначеність факторів зовнішнього середовища. У цих умовах застосування традиційних статистичних підходів (таких як, наприклад, апарат математичної статистики, апарат аналітичної обробки даних OLAP), як правило, є неоптимальним з точки зору ефективності. Це актуалізує проблему забезпечення організацій засобами автоматизованого аналізу потоків даних, вилучення і накопичення знань [3].

Перед кожною організацією стоїть завдання аналізу внутрішніх і зовнішніх факторів, що

впливають на рівень її виробництва. Значно знизити тимчасові та ресурсні витрати, підвищити рівень якості діловодства дає змогу комплексне забезпечення організації засобами централізованого управління даними, зокрема засобами електронного документообігу.

Ефективність процесу автоматизованого документообігу може бути значною мірою знижена під дією низки факторів: виникнення нештатних ситуацій в процесі експлуатації, непередбачених на етапі проектування системи електронного документообігу; помилки кінцевого користувача (оператора, адміністратора системи) тощо. Наслідками таких помилок є втрата або спотворення даних, неузгодже-

ності і протиріччя в роботі різних структур організації.

Універсальні засоби ІАД досить складні та дорогі, тому вони не можуть широко застосовуватися в рамках інтегрованих систем, орієнтованих на кінцевого користувача. У зв'язку з цим зростає необхідність розроблення систем інтелектуального аналізу даних (СІАД).

Аналіз останніх досліджень і публікацій, в яких започатковано розв'язання даної проблеми і на які спирається автор. Система електронного документообігу (СЕД) – організаційно-технічна система, що забезпечує процес створення, управління доступом і поширення електронних документів в комп'ютерних мережах, а також яка забезпечує контроль над потоками документів в організації. Перевагами електронного документообігу є можливість паралельного виконання операцій над документами, безперервність руху документа; забезпечення високоєфективного пошуку документів, забезпечення звітності за різними статусами і атрибутами документів.

Складність та різноманітність методів ІАД вимагають створення спеціалізованих засобів кінцевого користувача для вирішення типових задач аналізу інформації в конкретних областях.

Системи ІАД застосовуються у наукових дослідженнях та освіті, в роботі правоохоронних органів, виробництві, охороні здоров'я та багатьох інших областях. Особливо широко технологія ІАД використовується в системах електронного документообігу.

Виділення не вирішених раніше частин загальної проблеми, котрим присвячується означена стаття. У статті подається дослідження основних методів інтелектуального аналізу даних, які можуть застосовуватись в електронному документообігу на підприємстві.

Формулювання цілей статті (**постановка завдання**). Мета статті полягає у тому, щоб розглянути системи інтелектуального аналізу даних, завданнями яких є вилучення, накопичення і обробка знань, сценаріїв, моделей поведінки інформаційних систем організації, а також провести аналіз цих систем; визначити, які методи інтелектуального аналізу даних раціонально використовувати у системах електронного документообігу.

Виклад основного матеріалу дослідження з повним обґрунтуванням отриманих наукових результатів. Інтелектуальний аналіз даних (ІАД) є міждисциплінарною галуззю науки, що виникла і розвивається на базі досягнень прикладної статистики, розпізнавання образів, штучного інтелекту, теорії баз даних та інших областей сучасної науки. ІАД є спектром технологій і методик виявлення статистичних взаємозв'язків, закономірностей, правил у наборах даних. Нині реалізація і розвиток цього підходу ведуться корпорацією «Oracle», компанією «SAS», компанією «Google In,c.» тощо.

Задачі ІАД зводяться до реалізації алгоритмів класифікації, кластеризації, асоціації, прогнозування, побудови продукцій, регресійного та кореляційного аналізу. Основними завданнями СІАД є аналіз структури баз даних, запитів, журналів транзакцій, пошук закономірностей і взаємозв'язків між даними, побудова продукційних моделей, емпіричних моделей, дерев рішень та семантичних мереж [2].

Проведення інтелектуального аналізу даних можна розбити на такі етапи:

- постановка задачі;
- підготовка даних;
- вивчення даних;
- побудова моделей;
- дослідження і перевірка моделей;
- розгортання та оновлення моделей.

В основі сучасного ІАД лежить принцип побудови шаблонів гіпотез. Таким чином, здійснюється формування наборів гіпотез на основі вибірок даних, а не пошук вибірок даних, що задовольняють спочатку сформульовані гіпотези.

База даних інформаційної системи (ІС) є набором таблиць, що містять записи про властивості різних об'єктів. Ці записи містять дані, що відображають той чи інший факт у рамках предметної області, але позбавлені смислового навантаження.

Масиви даних значного обсягу, накопичені за певний часовий проміжок, можуть служити джерелами інформації про закономірності поведінки ІС, взаємозв'язками між даними, даними для побудови сценаріїв функціонування системи. Отримана таким чином інформація знаходить своє застосування в галузях прогнозування, стратегічного антикризового управління, в аналізі ризиків [1].

Виділяють такі типи закономірностей у даних, які виявляються за допомогою ІАД:

1) асоціація – виявлення залежностей між зв'язаними подіями, що вказують, що з події X випливає подія Y; такі правила називаються асоціативними;

2) секвенція – оцінка ймовірності виникнення послідовності подій;

3) класифікація – установлення функціональної залежності між вхідними й дискретними вихідними змінними. За допомогою класифікації вирішується задача віднесення об'єктів (спостережень, подій) до одного із заздалегідь відомих класів;

4) кластеризація – це угруповання об'єктів (спостережень, подій) на основі даних (властивостей), що описують сутність об'єктів; об'єкти усередині кластера повинні бути «схожими» один на одного й відрізнятися від об'єктів, що ввійшли в інші кластери;

5) регресія – установлення функціональної залежності між вхідними й неперервними вихідними змінними; прогнозування найчастіше зводиться до розв'язку задачі регресії.

Сьогодні розроблено і використовується велика кількість методів ІАД, які можуть застосовуватися в цілях оптимізації електронного документообігу. Серед них можна виділити такі:

- регресійний, дисперсійний, кореляційний аналіз;
- побудова емпіричних моделей;
- реалізація нейромережових алгоритмів ІАД;
- побудова дерев рішень;
- побудова кластерних моделей.

Метод регресійного, дисперсійного, кореляційного аналізу передбачає дослідження [4]:

- ступеня впливу декількох незалежних атрибутів на залежний атрибут (регресійний аналіз);
- ступеня впливу атрибутів один на одного (кореляційний аналіз);
- ступеня впливу факторів, що визначають атрибут, на його значення (дисперсійний аналіз).

Побудова емпіричних моделей передбачає накопичення відомостей про досвід функціонування системи. Як правило, окремо взята емпірична модель застосовується для аналізу даних строго конкретної предметної області (наприклад, для фінансового аналізу).

Метод побудови дерев рішень надає не завжди оптимальні, але завжди наочні рішення. Цей підхід також є досить популярним завдяки своїй простоті й наочності. Дерева рішень є ієрархічною структурою питань на «так» і «ні». Листям в цьому випадку є твердження про віднесення об'єкта до відповідного класу (категорії) або про настання тієї чи іншої події. Недоліком цього підходу є його обмеженість в побудові правил логічного висновку – рішення задачі зводиться до формування ланцюжка послідовного перегляду властивостей об'єктів, а не знаходження справжніх і логічно повних закономірностей поведінки об'єктів системи.

Побудова кластерних моделей є об'єднанням схожих об'єктів на групи (класи) на підставі подібних значень у наборі даних.

Особливої уваги заслуговує реалізація нейромережових алгоритмів ІАД, заснованих на застосуванні апарату штучних нейронних мереж (ШНМ). Це великий клас систем, архітектура яких має аналогію (як тепер відомо, досить слабку) з побудовою нервової тканини з нейронів. В одній з найбільш поширених архітектур, багат шаровому перцептроні зі зворотним поширенням помилки, імітується робота нейронів у складі ієрархічної мережі, де кожен нейрон більш високого рівня з'єднаний своїми входами з виходами нейронів нижчого шару. На нейрони найнижчого шару подаються значення вхідних параметрів, на основі яких потрібно приймати якісь рішення, прогнозувати розвиток ситуації.

Перевагою цього напрямку є одна з основних властивостей ІНС – здатність до навчання. Навчання ШНМ полягає в тренуванні на раніше

отриманих наборах, що включають вхідні і відповідні їм вихідні дані. За допомогою різних алгоритмів навчання ваги міжнейронних зв'язків встановлюються таким чином, щоб для будь-яких наборів вхідних даних одержували відповіді були максимально близькими до правильних відповідей. Недоліками є необхідність у великій навчальній вибірці даних; складність аналізу структури ШНМ (людський мозок не здатний інтерпретувати набір з тисяч значень ваг зв'язків всередині мережі). Цей підхід широко використовується в задачах ІАД.

На рівні реалізації метою процесу ІАД є побудова моделі, що відбиває поведінку системи та дає змогу прогнозувати поведінку системи залежно від зміни зовнішніх і внутрішніх факторів.

У системах електронного документообігу доцільно використовувати регресійний, дисперсійний, кореляційний аналіз даних, нейромережові та генетичні моделі. Реалізація СІАД передбачає розробку програмного забезпечення – набору бібліотек утиліт, що розширюють функціональні можливості СЕД.

Впровадження на підприємствах цієї методики дає змогу вирішити такі проблеми, як мінімізація інформаційних ризиків, узгодження порядку експлуатації СІАД з законодавством України, зміна програмного і апаратного забезпечення згідно з виникаючими у процесі функціонування системами інтелектуального аналізу даних характером завдань і рівнем навантаження. Поряд з перерахованими слід відзначити найважливіші питання забезпечення безпеки, конфіденційності, а також запобігання некоректного використання і спотворення даних.

Використання комплексу методів ІАД дає змогу системно досліджувати питання впливу зовнішніх і внутрішніх факторів на ефективність виробничих процесів. Серед основних переваг ІАД стосовно електронного документообігу слід особливо відзначити можливість здійснення ситуаційного аналізу з метою оптимізації виробничих процесів.

Висновки з цього дослідження і перспективи подальших розвідок у даному напрямі. Безумовною перевагою ІАД є його універсальність, тобто можливість застосування в найрізноманітніших областях сучасних інформаційних технологій. Широке впровадження систем інтелектуального аналізу даних створює передумови для переходу до якісно нового рівня управління виробничими процесами, зокрема роботою систем електронного документообігу. Це дає підставу розглядати розробку систем інтелектуального аналізу даних як перспективний напрям розвитку обчислювальної техніки у модернізації виробництва.

У системах електронного документообігу доцільно використовувати регресійний, дисперсійний, кореляційний аналіз даних, нейромережові та генетичні моделі.

БІБЛІОГРАФІЧНИЙ СПИСОК:

1. Rabunal J.R. Artificial Neural Networks in Real-Life Applications / J.R. Rabunal, J. Dorrado (eds.). – Hershey – London – Melbourne – Singapore: Idea Group Publishing, 2006.
2. Wang J. Data Mining: Opportunities and Challenges / J. Wang (eds.). Hershey – London – Melbourne – Singapore – Beijing: Idea Group Publishing, 2003.
3. Митчелл К. Программирование и отладка в Delphi: Учебный курс / К. Митчелл. – М.; СПб.: Киев, 2003.
4. Радченко С.Г. Методология регрессионного анализа: [монография] / С.Г. Радченко. – К.: Корнийчук, 2011. – 376 с.
5. Математические методы прогнозирования экономических показателей: [учебное пособие] / [А.Р. Саяпова, Е.А. Гусельникова, И.А. Лакман, Н.М. Шамуратов]. – Уфа, 2000 – 128 с.
6. Фёрстер Э. Методы корреляционного и регрессионного анализа / Э. Фёрстер, Б. Рёнц. – М.: Финансы и статистика, 1981. – 302 с.
7. Хаммер М. Реинженеринг корпорации: Манифест в бизнесе / М. Хаммер, Дж. Чамли; пер. с англ. Ю.Е. Корнилович. – М.: Манн, Иванов и Фербер, 2006.