

УПРАВЛІННЯ ПРОЕКТАМИ ТА ПРОГРАМАМИ

УДК 004.6

І.І. Глаголева, А.Ю. Берко

Національний університет "Львівська політехніка",
кафедра інформаційних систем та мереж

ЗАСТОСУВАННЯ ЗАСОБІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ДЛЯ ПРОГНОЗУВАННЯ ВИКОРИСТАННЯ ЗЕМЕЛЬНИХ РЕСУРСІВ

© Глаголева І.І., Берко А.Ю., 2013

Описано процедури інтелектуального аналізу даних на основі прогнозування часових рядів для даних земельного кадастру. Розглянуто положення, необхідні для розроблення методу прогнозування з використанням часових рядів. Розроблено математичну модель для прогнозування рядів, а також технічно реалізовано задачу прогнозування використання земельних ресурсів у Стрийському районі Львівської області.

Ключові слова: земельний кадастр, інтелектуальний аналіз даних, прогнозування.

This article describes the procedures of data mining based on prediction of time series for land cadastre data. Principles, required for the development of the method of forecasting using time series are examined. Mathematical model for serious prediction is developed. The task of prediction of land resources used in Striy District Lviv Region is technically realized.

Key words: land cadastre, data mining, forecasting.

Вступ

Ефективність використання земельних ресурсів є важливим показником у процесах прийняття рішень з управління ними. В даних земельного кадастру міститься велика кількість відомостей, які утворюють значний потенціал для підвищення ефективності використання земельних ресурсів. Одним зі шляхів реалізації цього потенціалу є прогнозування можливих варіантів використання ресурсів – процес формування ймовірних показників визначеного об'єкта, спрямований на виявлення та вивчення можливих альтернатив майбутнього його розвитку. Прогнозування є важливою ланкою між теорією і практикою в багатьох галузях суспільної діяльності, зокрема в управлінні раціональним використанням земельних ресурсів. Прогнозування процесів використання земельних ресурсів дає змогу виконувати завдання ефективною та раціональною експлуатації земель, забезпечує баланс попиту і пропозиції на землю.

За останні десятиліття зібрано велику кількість просторових даних, які зберігаються в Головному управлінні Держкомзему. Обсяги даних настільки значні, що аналіз їх потребує істотних часових, фінансових і технічних затрат, хоча необхідність проведення такого аналізу цілком очевидна, оскільки в таких "сирих даних" закладено знання, які можна використати для ухвалення рішень, які визначають подальші процеси використання земель протягом тривалого часу.

Одним із найперспективніших сьогодні засобів розв'язання задач ефективного використання і переробки великих обсягів даних є технології, які ґрунтуються на методах інтелектуального аналізу даних (Data Mining).

Інтелектуальний аналіз (Data Mining, DM) даних являє собою процес виявлення прихованих чи не поданих явно, але придатних до використання, відомостей у великих наборах даних. В інтелектуальному аналізі даних застосовують різноманітні математичні методи для виявлення закономірностей, зв'язків і тенденцій, що існують в даних. Зазвичай такі закономірності неможливо виявити під час традиційного перегляду даних, оскільки зв'язки занадто складні або не можуть бути

визначеними через надвеликий обсяг даних [1]. Інтелектуальний аналіз даних (ІАД) – мультидисциплінарна область, що виникла та розвивається на основі таких наук, як прикладна статистика, розпізнавання образів, штучний інтелект, теорія баз даних та інших подібних інструментаріїв. Інтелектуальний аналіз даних є складовою частиною процесу видобування знань з баз даних (KDD, knowledge discovery in data bases). Він дає змогу розкрити суть прихованих залежностей у даних, виявити взаємні впливи між властивостями об'єктів, інформація про які зберігається в базах даних, виділити закономірності, властиві певному набору даних [1].

До основних методів та алгоритмів Data Mining належать, зокрема, такі: нейронні мережі, дерева рішень, символні правила, методи найближчого сусіда і k-найближчого сусіда, метод опорних векторів, байєсівські мережі, лінійна регресія, кореляційно-регресійний аналіз, ієрархічні методи кластерного аналізу, неієрархічні методи кластерного аналізу, зокрема алгоритми k-середніх і k-медіани, методи пошуку асоціативних правил, серед яких алгоритм Apriori; метод обмеженого перебору, еволюційне програмування і генетичні алгоритми, різноманітні методи візуалізації даних [1].

Більшість аналітичних методів, що використовуються в технології Data Mining, – це відомі математичні алгоритми та методи. Новим у їх застосуванні є можливість використання цих засобів для вирішення тих чи інших конкретних проблем опрацювання даних, яка зумовлена появою нових можливостей технічних і програмних засобів. Зазначимо, що більшість методів Data Mining розроблено в межах теорії штучного інтелекту [2].

Аналіз останніх публікацій та досліджень

Основним чинником під час розв'язання задач прогнозування є часові показники. В середині 90-х років минулого століття розроблено принципово новий і достатньо могутній клас алгоритмів для прогнозування часових рядів. Велику частину роботи з дослідження методології та перевірки моделей виконали двоє статистиків, Г. Е. П. Бокс (G.E.P. Box) і Г. М. Дженкінс (G.M. Jenkins) [3]. З того часу побудову подібних моделей і отримання на їх основі прогнозів називають методами Бокса–Дженкінса. В цю групу входить декілька алгоритмів, найвідомішим і використовуваним з них є алгоритм ARIMA (autoregressive integrated moving average). Він вбудований практично в будь-який спеціалізований пакет для прогнозування. У класичному варіанті ARIMA не використовуються незалежні змінні. Моделі спираються тільки на інформацію, що міститься в передісторії прогнозованих рядів. У наш час в науковій літературі часто згадуються варіанти моделей ARIMA, що дають змогу враховувати незалежні змінні. На відміну від розглянутих раніше методик прогнозування тимчасових рядів, у методології ARIMA не передбачено якої-небудь чіткої моделі для прогнозування цієї тимчасової серії. Задається лише загальний клас моделей, що описують часовий ряд і дають змогу якось виражати поточне значення змінної через її попередні значення. Потім алгоритм, підлаштовуючи внутрішні параметри, сам вибирає найвідповіднішу модель прогнозування. Відповідність і зв'язки між моделями Бокса–Дженкінса [2], [3] утворюють ієрархію моделей, яку логічно можна описати послідовністю

$$AR(p) + MA(q) \rightarrow ARMA(p, q) \rightarrow ARMA(p, q)(P, Q) \rightarrow ARIMA(p, q, r)(P, Q, R) \rightarrow \dots$$

Елементи цієї послідовності є моделями певного виду, між якими підтримуються визначені співвідношення. Згідно з [3] ці моделі визначають так.

Авторегресивна модель порядку p – $AR(p)$. Модель має вигляд:

$$Y(t) = f_0 + f_1 \cdot Y(t-1) + f_2 \cdot Y(t-2) + \dots + f_p \cdot Y(t-p) + E(t), \quad (1)$$

де $Y(t)$ – залежна змінна у момент часу t ; $f_0, f_1, f_2, \dots, f_p$ – оцінювані параметри; $E(t)$ – помилка від впливу змінних, які не враховано у моделі [3].

Основне завдання під час застосування цієї моделі зводиться до встановлення значень оцінюваних параметрів $f_0, f_1, f_2, \dots, f_p$. Відомо декілька способів розв'язання такої задачі, найпоширенішим з яких є пошук значень таких параметрів через систему рівнянь Юла–Уолкера [2]. Для складання цієї системи необхідним є розрахунок значень автокореляційної функції. Простішим

способом отримання значень параметрів $f_0, f_1, f_2, \dots, f_p$ є їх розрахунок за методом найменших квадратів.

MA(q) – модель з ковзним середнім порядку q. Модель має вигляд:

$$Y(t) = m + e(t) - w_1 \cdot e(t-1) - w_2 \cdot e(t-2) - \dots - w_p \cdot e(t-p), \quad (2)$$

де $Y(t)$ – залежна змінна у момент часу t ; $w_0, w_1, w_2, \dots, w_p$ – оцінювані параметри.

Авторегресійне ковзне середнє ARMA(p,q). Під позначенням $ARMA(p,q)$ [2] розуміють модель p авторегресійних складових, що містить q ковзних середніх. У точнішому поданні до складу моделі $ARMA(p,q)$ включають моделі $AR(p)$ і $MA(q)$:

$$X_t = c + e_t + \sum_{i=1}^q q_i \cdot e_{t-i} + \sum_{i=1}^p f_i \cdot X_{t-i}, \quad (3)$$

Зазвичай значення помилки e_t вважають незалежними однаково розподіленими випадковими величинами, узятими з нормального розподілу з нульовим середнім: $e_t \sim N(0, s^2)$, де s^2 – дисперсія. Припущення можна ослабити, але це може привести до зміни властивостей моделі. Наприклад, якщо не припускати незалежності й однакового розподілу помилок, поведінка моделі суттєво змінюється.

Модель ARIMA (p,d,q). У завданні аналізу тимчасового ряду зі складною структурою часто використовуються моделі класу $ARIMA(p,d,q)$ [3] (авторегресійне інтегрування ковзаючого середнього – Autoregressive Integrated Moving Average) порядку (p,d,q) , які моделюють різні ситуації, що виникають під час аналізу стаціонарних і нестаціонарних рядів. Залежно від аналізованого ряду модель $ARIMA(p,d,q)$ може трансформуватися до авторегресійної моделі $AR(p)$, моделі ковзного середнього $MA(q)$ або змішаної моделі $ARMA(p,q)$. У разі переходу від нестаціонарного ряду до стаціонарного значення вважають, що параметр d , визначає порядок різниці, дорівнює 0 або 1, тобто цей параметр має тільки цілочислові значення. Зазвичай обмежуються вибором між $d = 0$ і $d = 1$. Проте у поле зору дослідників не входить ситуація, коли параметр d може набувати дробових значень.

Модель ARFIMA(p,d,q). Для ситуації розгляду дробових значень порядку різниці в роботах зарубіжних учених, в першу чергу, С.W.Granger, J.R.Hosking, P.M.Robinson, R. Veran, запропоновано новий клас моделей $ARFIMA(p,d,q)$ [3] (F: fractional – дріб), що допускає можливість нецілого параметра d і авторегресійний дробоінтегрований процес ковзного середнього. Такі ряди мають певну специфіку: самоподібність, дробову розмірність, кореляцію, яка повільно спадає. Прогнозування тимчасових рядів за допомогою моделі $ARFIMA(p,d,q)$ відкриває ширші перспективи для підвищення точності прогнозу.

Модель вигляду ARIMA (p,d,q)(P,D,Q)S. Таку модель описують виразом $ARIMA(p,d,q)(P,D,Q)S$ [3],

де p – авторегресійні доданки;

d – різниці;

q – доданки ковзного середнього;

P – сезонні авторегресійні доданки;

D – сезонні різниці на інтервалі S ;

Q – доданки сезонного ковзного середнього

Алгоритм часових рядів (Microsoft Time Series Algorithm)

Засоби Data Mining є стандартним інструментом розв'язання задач аналізу в СУБД Microsoft SQL Server. Їх застосування створює додаткові можливості щодо використання обсягів даних, накопичених протягом значного часу. Технології аналізу даних СУБД Microsoft ґрунтуються на застосуванні попередньо визначених і описаних моделей. Така модель являє собою набір метаданих, що відображає залежності, правила і закономірності, які підтримуються в початкових даних. При цьому структура моделі визначає набір ключових атрибутів, а вміст формують статистичні та узагальнені значення даних. Такий підхід передбачає можливості змін та побудови нових моделей. У такий спосіб реалізується процес навчання засобів Data Mining.

Однією з моделей, які застосовують засоби інтелектуального аналізу даних СУБД Microsoft SQL Server, є модель аналізу часових рядів. Алгоритм часових рядів (Microsoft) реалізує алгоритми регресії, оптимізовані для прогнозу неперервних значень в часі. На відміну від інших алгоритмів, таких як дерева рішень, модель часових рядів не вимагає додаткових стовпців нових відомостей, щоб прогнозувати тенденцію. За допомогою моделі часових рядів можна прогнозувати тенденції на основі тільки початкового набору даних, використаного для створення моделі. Під час прогнозування можна вводити в модель нові дані й автоматично задіювати їх при аналізі тенденцій.

У складі засобів аналізу даних СУБД Microsoft SQL Server 2005 в алгоритмі часових рядів (Microsoft) використано єдиний алгоритм ARTXP. Він оптимізований для короткострокових прогнозів, тому прогнозував наступне вірогідне значення в ряді. Починаючи з версії SQL Server 2008, в алгоритмі часових рядів (Microsoft) разом з алгоритмом ARTXP використовується другий алгоритм, ARIMA. Алгоритм ARIMA оптимізований для довгострокового прогнозування.

За замовчуванням в алгоритмі часових рядів Microsoft використовується комбінація алгоритмів для аналізу закономірностей і підготовки прогнозів. Алгоритм навчає дві окремі моделі тих самих даних: в одній моделі використовується алгоритм ARTXP, а в іншій – алгоритм ARIMA. Потім алгоритм об'єднує результати обох моделей, щоб сформулювати найкращий прогноз для змінної кількості тимчасових зрізів. Алгоритм ARTXP краще підходить для короткострокових прогнозів, тому, починаючи ряд прогнозів, на нього треба покладатися більшою мірою. Але у міру того, як тимчасові зрізи, що використовуються для прогнозування, йдуть все далі в майбутнє, стає кориснішим алгоритм ARIMA.

Можна управляти поєднанням алгоритмів, переносячи акцент у тимчасових рядах на короткостроковий або довгостроковий прогноз. Починаючи з випуску SQL Server 2008 Standard, для алгоритму тимчасових рядів (Майкрософт) можна вказати один з таких режимів:

- використання тільки алгоритму ARTXP для короткострокового прогнозування;
- застосування тільки алгоритму ARIMA для довгострокового прогнозування;
- використання поєднання двох алгоритмів (за замовчуванням).

У разі використання змішаної моделі алгоритм тимчасових рядів (Майкрософт) об'єднує два алгоритми:

- для формування двох перших прогнозів завжди використовується тільки алгоритм ARTXP;
- після перших двох прогнозів використовується поєднання алгоритмів ARIMA і ARTXP;
- з подальшими кроками прогнозування частка алгоритму ARIMA в прогнозах зростає до повної відмови від використання ARTXP;
- управляти моментом поєднання алгоритмів, швидкістю зниження частки алгоритму ARTXP і підвищенням частки алгоритму ARIMA можна, змінюючи параметр PREDICTION _ SMOOTHING.

Обидва алгоритми можуть виявляти вплив сезонних чинників на дані на декількох рівнях. Наприклад, усередині річних циклів можуть існувати місячні цикли зміни даних. Щоб визначити сезонні цикли, можна ввести відомості про періодичність або вказати режим автоматичного виявлення періодичності.

Під час підготовки даних для використання в навчанні будь-якої моделі інтелектуального аналізу даних необхідно розуміти вимоги конкретної моделі й способи використання даних.

У кожній моделі прогнозування повинен міститися набір варіантів, тобто стовпчик таблиці, що вказує зрізи часу, або інший ряд, в якому відбуваються зміни. У інших моделях набір варіантів може бути текстовим полем або яким-небудь ідентифікатором, таким як код споживача або код транзакції. Проте модель тимчасових рядів повинна завжди використати дату, час або інше унікальне числове значення, таке як набір варіантів.

Вимоги до моделі тимчасових рядів:

- Окремий ключовий стовпчик часу. Кожна модель повинна містити один стовпець числового типу або типу "date", який використовується як набір варіантів, що визначає тимчасові

зрізи, які використовує модель. Типом даних стовпчика "key time" може бути datetime або будь-який числовий тип. Проте цей стовпчик повинен містити безперервні значення, які мають бути унікальними для кожного ряду. Набір варіантів для моделі тимчасових рядів не може зберігатися в двох стовпчиках, таких як стовпчик "Рік" або стовпчик "Місяць".

- Прогнозований стовпчик. Кожна модель повинна містити принаймні один прогнозований стовпчик, на підставі якого алгоритм будуватиме модель тимчасових рядів. Прогнозований стовпчик повинен мати тип даних з безперервними значеннями. Наприклад, можна прогнозувати, як змінюються з часом числові атрибути, такі як дохід, обсяг продажів або температура. Проте не можна використати як прогнозований стовпчик, що містить дискретні значення, такі як споживчий статус або рівень освіти.

- Необов'язковий ключовий стовпчик ряду. Кожна модель має додатковий ключовий стовпчик, що містить унікальні значення, які ідентифікують ряд. Цей необов'язковий ключовий стовпчик ряду повинен містити унікальні значення. Наприклад, одна модель може містити дані про продажі багатьох моделей продуктів, якщо існує тільки один запис для кожної назви продукту в кожному тимчасовому зрізі.

Після навчання моделі результати зберігаються у вигляді набору закономірностей, які можна досліджувати або робити на їх основі прогнози [4].

Цілі та завдання статті

Основною метою статті є розроблення способів прогнозування використання земельних ресурсів із використанням інтелектуального аналізу даних. Завданнями, які визначає ця ціль, такі:

- дослідження та обґрунтування можливостей застосування методів та засобів інтелектуального аналізу даних у процесах формування прогнозних оцінок щодо використання земельних ресурсів;

- моделювання процесів аналізу кадастрових даних із застосування часових рядів⁴
- розроблення порядку формування прогнозних оцінок використання земельних ресурсів на основі моделей, які базуються на застосуванні часових рядів;
- апробація розроблених у роботі вирішень на основі даних земельного кадастру Стрийського району Львівської області.

Основні результати досліджень

Інформація про використання земельних ресурсів Стрийського району зберігається в управлінні Держкомзему, у формі "б-зем", яка є звітом про наявність земель та розподіл їх за власниками, землекористувачами, угіддями та видами економічної діяльності за рік.

Для прогнозування використання земельних ресурсів за допомогою алгоритму часових рядів використано інформацію за останні п'ять років та засобами SQL Server Management Studio побудовано модель інтелектуального аналізу.

У процесі моделювання було враховано низку обмежень щодо складу та способів подання даних, які підлягають інтелектуальному аналізу. Зокрема, модель часових рядів повинна містити один стовпчик числового типу або типу "дата". Таким стовпчиком було вибрано поле "Рік". Кожна модель повинна містити принаймні один прогнозований стовпчик, на підставі якого алгоритм будуватиме модель часових рядів. Прогнозований стовпчик повинен мати тип даних з неперервними значеннями. Прогнозованими стовпчиками вибрано такі поля таблиці:

- загальна площа сільськогосподарських земель;
- площа сільськогосподарських угідь (всього);
- землі, які перебувають на стадії меліоративного будівництва та відновлення родючості;
- забруднені сільськогосподарські землі (всього);
- ліси та інші лісовкриті площі (всього);
- забудовані землі (всього);
- вода;

- землі природоохоронного призначення;
- землі громадського призначення;
- землі оздоровчого призначення;
- землі промисловості;
- землі історико-культурного призначення.

Процес створення моделі Data Mining з використанням алгоритму часових рядів реалізовано за допомогою такого DMX запиту:

```
CREATE MINING MODEL [Forecasting_MIXED]
(
  [Рік] LONG KEY TIME,
  [Води всього]DOUBLE CONTINUOUS PREDICT,
  [Забуднені сг землі які не використовуються]DOUBLE CONTINUOUS PREDICT,
  [Забудовані Землі всього]DOUBLE CONTINUOUS PREDICT,
  [Землі Громадського Призначення]DOUBLE CONTINUOUS PREDICT,
  [Землі Історико-культурного Призначення]DOUBLE CONTINUOUS PREDICT,
  [Землі Оздоровчого Призначення]DOUBLE CONTINUOUS PREDICT,
  [Землі Природоохоронного призначення]DOUBLE CONTINUOUS PREDICT,
  [Землі Промисловості]DOUBLE CONTINUOUS PREDICT,
  [Землі які перебувають в стадії меліоративного будівництва та відновлення родючості]
DOUBLE CONTINUOUS PREDICT,
  [Ліси та інші лісовкриті площі(всього)] DOUBLE CONTINUOUS PREDICT,
  [Сільськогосподарські Землі всього] DOUBLE CONTINUOUS PREDICT,
  [Сільськогосподарські Угіддя всього] DOUBLE CONTINUOUS PREDICT
)
USING Microsoft_Time_Series (AUTO_DETECT_PERIODICITY = 0.8,
FORECAST_METHOD = 'MIXED')
WITH DRILLTHROUGH
```

У першому рядку запиту вказано назву моделі Forecasting_MIXED. Служби Analysis Services автоматично формують ім'я базової структури, додаючи рядок "structure" до імені моделі, що гарантує незбіг імені структури і моделі. У наступному рядку коду задається ключовий стовпець моделі інтелектуального аналізу даних, який у випадку з моделлю тимчасових рядів унікально визначає часовий етап у початкових даних. Часовий етап визначається за допомогою ключових слів KEY TIME після імені стовпця і типів даних. Якщо модель тимчасових рядів має окремий ключ ряду, вона визначається за допомогою ключового слова KEY. Отже, стовпець "Рік" містить значення тимчасових етапів, які використовують для визначення порядку слідування значень. Наступні рядки коду використовуються для визначення стовпців у прогнозованій моделі. Параметр алгоритму AUTO _ DETECT _ PERIODICITY = 0,8 вказує на те, що алгоритм повинен виявляти цикли в даних. Встановлення значення, близького до 1, дає змогу виявити безліч шаблонів, але може уповільнити процес обробки. Параметр алгоритму FORECAST _ METHOD задає алгоритм аналізу даних ARTXP або ARIMA або поєднання обох алгоритмів.

Ключове слово WITH DRILLTHROUGH забезпечує можливість перегляду детальної статистики в початкових даних після завершення створення моделі.

Процес побудови прогнозних показників поділяється на дві стадії – перша навчання засобів Data Mining на основі вибраної моделі часових рядів, друга – формування результатів прогнозування із застосуванням вивченої моделі. Графічно результати прогнозування подано на рис. 1–4.

На діаграмі (рис. 1) показано як минулий час, так і майбутній. Частина графіка, яка відповідає за прогнозовані значення, виділена темнішим фоном. Кожне прогнозоване значення виділене іншим кольором.

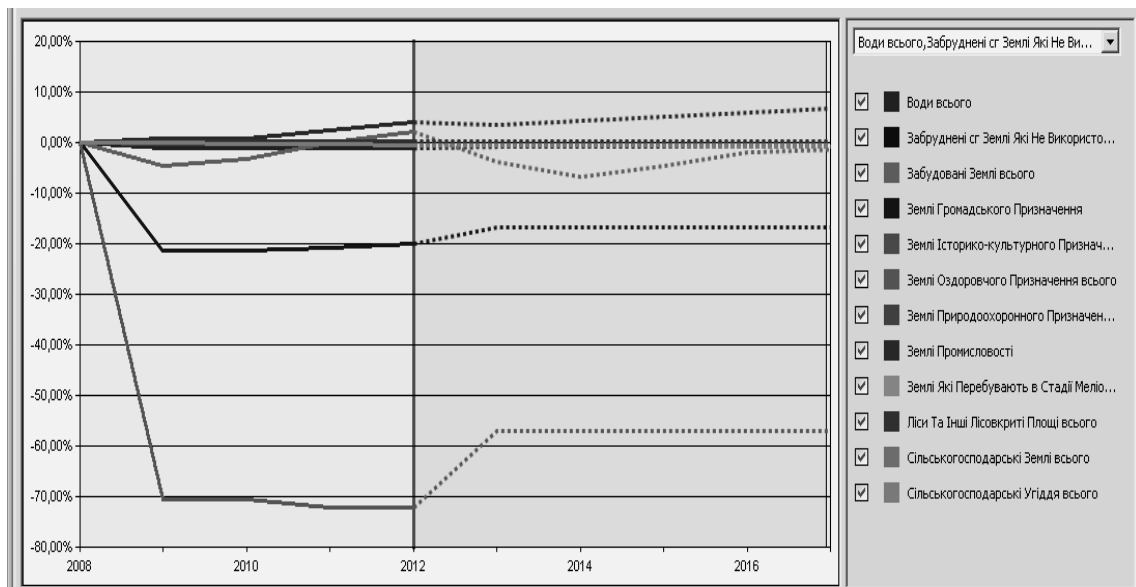


Рис. 1. Сторінка результатів застосування моделі до даних земельного кадастру

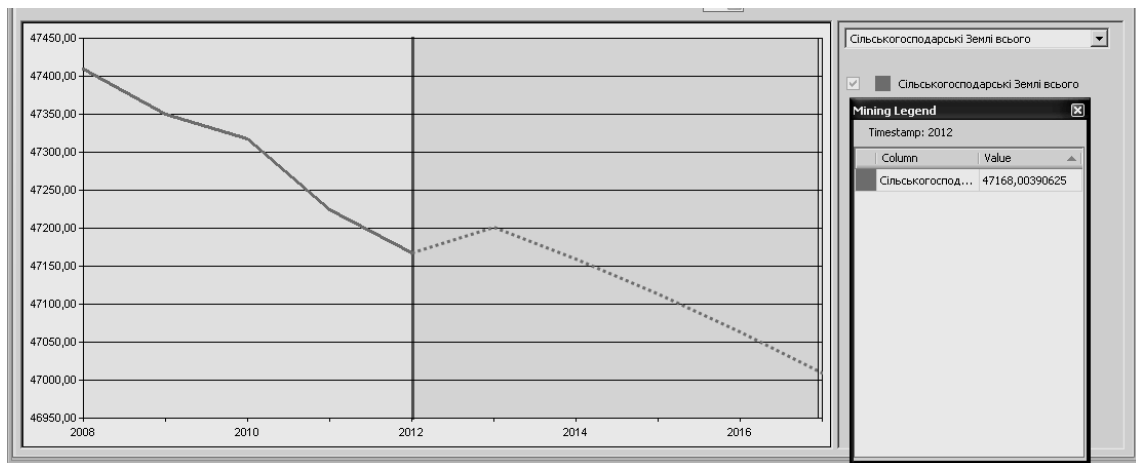


Рис. 2. Діаграма зміни земель сільськогосподарського призначення у разі позиціонування курсора на 2012 рік

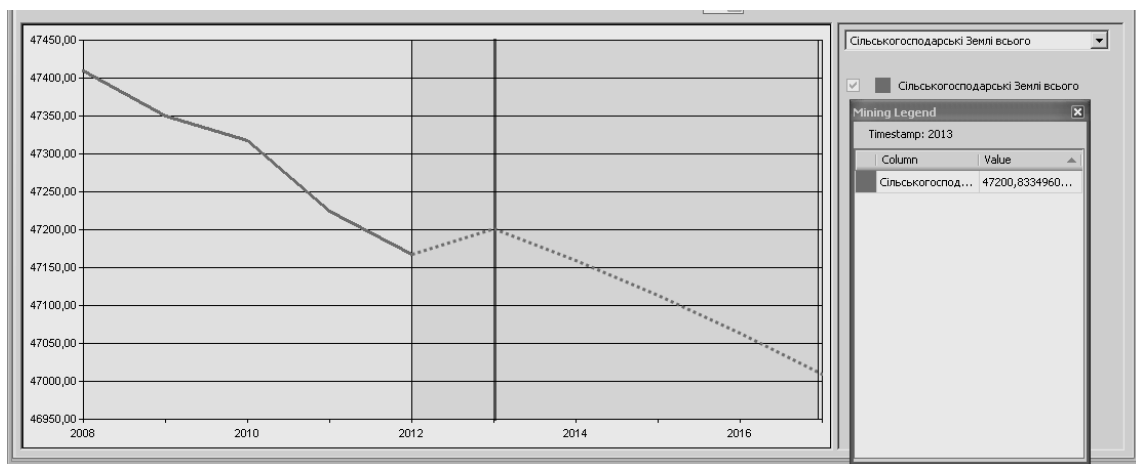


Рис. 3. Діаграма зміни земель сільськогосподарського призначення у разі позиціонування курсора на 2013 рік

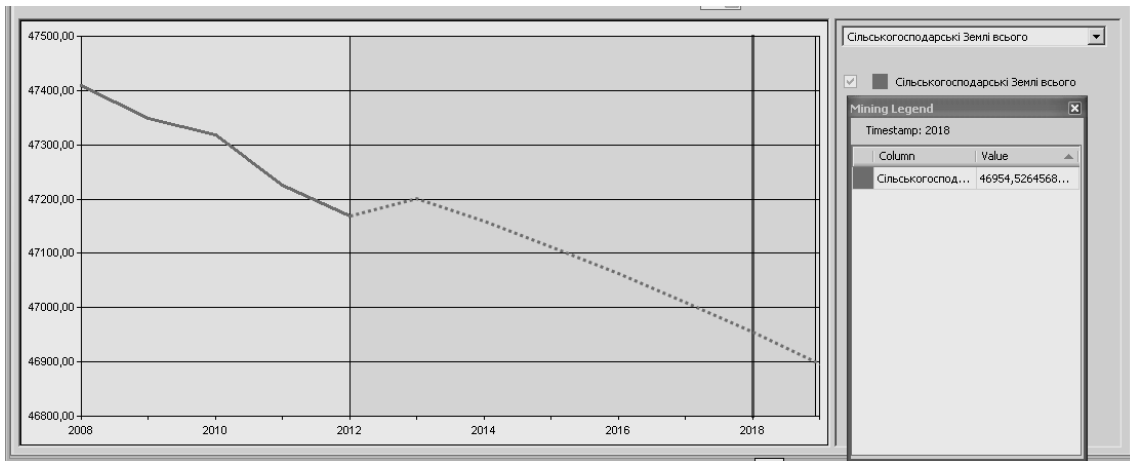


Рис. 4. Діаграма зміни земель сільськогосподарського призначення у разі позиціонування курсора на 2018 рік

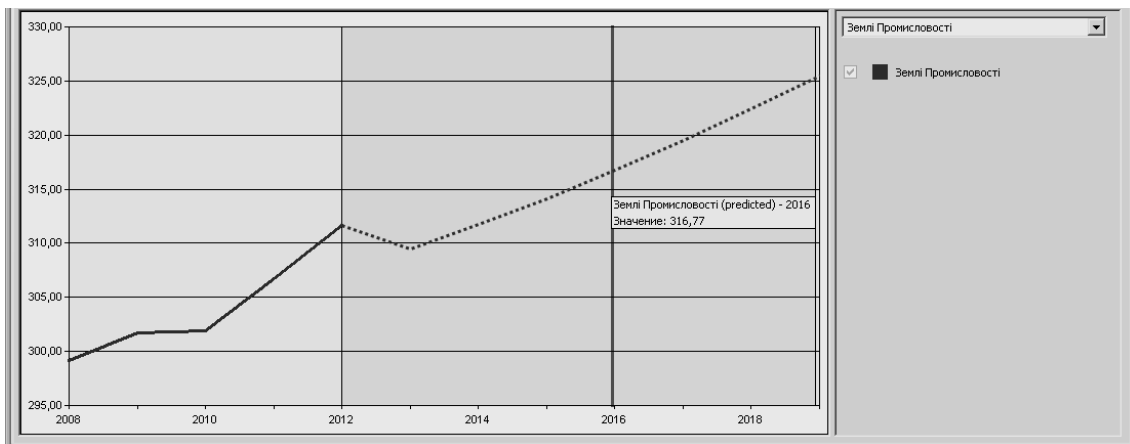


Рис. 5. Діаграма зміни земель промисловості у разі позиціонування курсора на 2016 рік

Для земель сільськогосподарського призначення побудовано діаграми із позиціонуванням курсора на 2012, 2013, 2018 роки відповідно (рис. 2, 3, 4), коли вибрано певний рік, у легенді відображається відповідне значення площі земель сільськогосподарського призначення.

Як бачимо значення площі земель сільськогосподарського призначення у 2012 р. (рис.2) становить 4768,003 га, а прогнозоване значення площі у 2014 р. – 47200,83 га (рис. 3), у 2018 р. – 46954, 52 га (рис. 4).

З діаграми (рис. 5) видно, що частково площі земель промисловості зменшаться у 2013 р., а в наступні періоди відбудеться їхнє зростання.

Для отримання числових результатів прогнозованих значень було розроблено запит такого виду:

```

SELECT
PredictTimeSeries([Forecasting_MIXED].[Води всього], 6),
PredictTimeSeries([Forecasting_MIXED].[Забруднені сг землі], 6),
PredictTimeSeries([Forecasting_MIXED].[Забудовані Землі всього], 6),
PredictTimeSeries([Forecasting_MIXED].[Землі Громадського Призначення], 6),
PredictTimeSeries([Forecasting_MIXED].[Землі Історико-культурного Призначення], 6),
PredictTimeSeries([Forecasting_MIXED].[Землі Оздоровчого Призначення], 6),
PredictTimeSeries([Forecasting_MIXED].[Землі Природоохоронного призначення], 6),
PredictTimeSeries([Forecasting_MIXED].[Землі Промисловості],6),
PredictTimeSeries([Forecasting_MIXED].[Ліси та інші лісовкриті площі(всього)], 6),

```


PredictTimeSeries([Forecasting_MIXED].[Сільськогосподарські Землі всього], 6)
 PredictTimeSeries([Forecasting_MIXED].[Сільськогосподарські Угіддя всього], 6)
 FROM
 [Forecasting_MIXED]

Функція PredictTimeSeries в кожному рядку прогнозування виконується для раніше вказаних полів. Числа, після імен прогнозованих атрибутів, вказують кількість часових кроків, необхідних для прогнозування. Результати виконання запиту подано в таблиці.

Прогнозовані значення використання земельних ресурсів

Земельні ресурси	Прогнозовані значення	
	Рік	Значення
1	2	3
Води всього	2013	2030,67204589844
	2014	2030,67204589844
	2015	2030,67204589844
	2016	2030,67204589844
	2017	2030,67204589844
	2018	2030,67204589844
Забруднені сільськогосподарські землі	2013	0
	2014	0
	2015	0
	2016	0
	2017	0
	2018	0
Забудовані землі всього	2013	2960,33473463929
	2014	2873,26946803778
	2015	2934,96439428538
	2016	3023,45628791804
	2017	3041,52843721532
	2018	3045,18295326832
Землі громадського призначення	2013	350,295520019531
	2014	350,295520019531
	2015	350,295520019531
	2016	350,295520019531
	2017	350,295520019531
	2018	350,295520019531
Землі історико-культурного призначення	2013	1,16999995708466
	2014	1,16999995708466
	2015	1,16999995708466
	2016	1,16999995708466
	2017	1,16999995708466
	2018	1,16999995708466
Землі оздоровчого призначення	2013	24,3473003387451
	2014	24,3473003387451
	2015	24,3473003387451
	2016	24,3473003387451
	2017	24,3473003387451
	2018	24,3473003387451
Землі природоохоронного призначення	2013	3392,28803710938
	2014	3392,28803710938
	2015	3392,28803710938
	2016	3392,28803710938
	2017	3392,28803710938
	2018	3392,28803710938
Землі промисловості	2013	309,506332397461
	2014	311,71667770516
	2015	314,147994091302

1	2	3
	2016	316,766467991486
	2017	319,542720543041
	2018	322,451269483727
Ліси та інші лісовкриті площі (всього)	2013	24807,279296875
	2014	24807,279296875
	2015	24807,279296875
	2016	24807,279296875
	2017	24807,279296875
	2018	24807,279296875
Сільськогосподарські землі всього	2013	47200,8334960938
	2014	47159,0185515904
	2015	47112,8627400318
	2016	47063,0257777211
	2017	47010,0811308795
	2018	46954,5264568717
Сільськогосподарські угіддя всього	2013	46183,0568359375
	2014	46145,1242591658
	2015	46103,2522098726
	2016	46058,0393566741
	2017	46010,00610202
	2018	45959,6040567214

Аналіз отриманих результатів показує, що незмінними залишаються площі забруднених сільськогосподарських земель, громадського, історико-культурного, оздоровчого, природоохоронного призначення, землі під лісами та водами, збільшуються площі земель промисловості й зменшуються площі сільськогосподарських земель та сільськогосподарських угідь.

Отже, для майбутнього землекористування увага має бути зосереджена на підвищенні ефективності використання земель із застосуванням таких заходів, як трансформація земель.

Висновки

Описаний у цій статті метод інтелектуального аналізу даних на основі прогнозування часових рядів для земельного кадастру дає змогу аналізувати дані та прогнозувати майбутні значення показників використання земельних ресурсів у Стрийському районі Львівської області.

Недоліком використання цього методу прогнозування є те, що він не використовує незалежні змінні й спирається лише на передісторію прогнозованого ряду.

1. Fayyad U.M. *Advances in Knowledge Discovery and Data Mining [Text]* / U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth. – Menlo Park, Calif.: AAAI/MIT Press, 1996. 2. Барсегян А.А. *Методы и модели анализа данных OLAP и DataMining [Текст]* / А.А. Барсегян, М.С. Курпьянов, В.В. Степаненко, И.И. Холод. – СПб: БХВ-Петербург, 2008. 3. Бокс Дж., Дженкинс Г. *Анализ временных рядов, прогноз и управление: пер. с англ. [Текст]* // под ред. В.Ф. Писаренко. – М.: Мир, 1974. 4. *Time Series Prediction DMX Tutorial [Електронний ресурс]*. – Режим доступу: <http://technet.microsoft.com/en-us/library/cc879270.aspx>.