

функціонування економічного механізму загалом та сприяють підвищенню ефективності діяльності торговельних підприємств зокрема.

1. Войнаренко М. Концепція кластерів – шлях до відродження виробництва на регіональному рівні / М. Войнаренко // *Економіст*. – 2000. – № 1. – С. 12–15.
2. Головне управління статистики у Львівській області. Офіційний сайт: [Електронний ресурс]. – Режим доступу: <http://www.dnprstat.gov.ua>.
3. Державна установа “Агентство з розвитку інфраструктури фондового ринку України” // Офіційний сайт: [Електронний ресурс]. – Режим доступу: <http://www.smida.gov.ua/>.
4. Державний комітет статистики України // Офіційний сайт: [Електронний ресурс]. – Режим доступу: <http://www.ukrstat.gov.ua/>.
5. Ігнашкіна Т. Б. Формування галузевих кластерів промислових підприємств Дніпропетровської області з метою аналізу відтворювальних процесів / Т. Б. Ігнашкіна, Н. О. Шура // *БІЗНЕСІНФОРМ: Економіка. Економіко-математичне моделювання*. – 2011. – № 7(2). – С. 23–30.
6. Лукань Л. Застосування кластерного аналізу для оцінки розвитку малого підприємництва в регіонах України / Л. Лукань, Г. Цегелик // *Формування ринкової економіки в Україні*. – 2009. – Вип. 19. – С. 73–80.
7. Мельман И. В. Сетевые черты кластерной организации производства / И. В. Мельман // *Современные аспекты экономики*. – СПб. – 2005. – № 19(86). – С. 101–110.
8. Панченко Н. Г. Кластерний аналіз в дослідженні показників соціально-економічного розвитку міст України / Н. Г. Панченко, В. Б. Родченко // *Збірник наукових праць Української державної академії залізничного транспорту*. – 2010. – № 114. – С. 205–210.
9. Пістунов І. М. Кластерний аналіз в економіці : навч. посіб. [Електронне видання] / І. М. Пістунов, О. П. Антонюк, І. Ю. Турчанінова. – Дніпропетровськ: Національний гірничий університет. – 2008. – 84 с.
10. Тараненко О.О. Механізм забезпечення економічної стійкості торговельних підприємств / О.О. Тараненко // *Економічний простір*. – 2012. – № 63. – С. 296–302.
11. Тищенко О. М. Кластери як вектор розвитку економіки: організація, сутність і концепції / О. М. Тищенко // *Теоретичні та прикладні питання економіки*. – 2010. – Вип. 21. – С. 74–80.

УДК 004.6

І.І. Глаголева, А.Ю. Берко

Національний університет “Львівська політехніка”,
кафедра інформаційних систем та мереж

ЗАСТОСУВАННЯ КЛАСТЕРНОГО АНАЛІЗУ ДЛЯ ОПРАЦЮВАННЯ ДАНИХ ЗЕМЕЛЬНОГО КАДАСТРУ

© Глаголева І.І., Берко А.Ю., 2014

Описано процедури інтелектуального аналізу даних на основі кластерного аналізу для даних земельного кадастру. Розглянуто положення, необхідні для кластерного аналізу з використанням методу k-середніх. Розроблено математичну модель, а також технічно реалізовано задачу кластеризації земельних ділянок у Стрийському районі Львівської області.

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

Procedures of data mining based on prediction of time series for land cadastre data are described in this article. Principles required for the development of the method of forecasting using time series are examined. Mathematical model is developed. The task of predicting land resources use in Striyskyi Park in Lviv is technically realized.

Key words: land cadastre, data mining, forecasting, cluster analysis.

Вступ

Одним з основних ресурсів, які забезпечують життєдіяльність суспільства в політичній, економічній, соціальній, виробничій, комунальній, екологічній та інших сферах, є земельний ресурс. Характерною особливістю земельних ресурсів є їх значна вартість, тому адекватна оцінка

землі є однією з найважливіших умов функціонування і розвитку економіки та суспільства. Результатом проведення державної кадастрової оцінки є реальна картина вартості земель (земельних ділянок) на території визначеного регіону, що уможливить ефективне управління земельними ресурсами, та пов'язаних з ними об'єктів нерухомості, а також збалансоване планування дохідної частини бюджетів всіх рівнів, стимулювання розвитку інвестиційних процесів та розвитку економіки загалом.

Необхідність в отриманні достовірної оціночної вартості земельних ділянок відчувають як державні та муніципальні органи виконавчої влади під час управління земельними ресурсами, здійснення перспективного розвитку населених пунктів, проведення раціональної земельної та податкової політики, так і приватні суб'єкти земельного права під час здійснення різноманітних операцій із землею, для цілей визначення найраціональнішого, найефективнішого розвитку того чи іншого регіону або тієї чи іншої конкретної земельної ділянки.

Перелічені чинники обґрунтовують актуальність проблематики розроблення алгоритмів, засобів та інформаційних технологій для проведення ефективного, точного і достовірного оцінювання земельних ділянок. В роботі розглянуто один із підходів до розв'язання задач такого класу, який ґрунтується на застосуванні методу кластеризації даних, що описують земельні ресурси. Цей метод придатний для оцінки практично всіх категорій земель, оскільки використовує найбільш загальні та об'єктивні показники об'єктів оцінювання. Кластеризація, як один з кроків загальної процедури, дає змогу досягти істотного зменшення обсягу виконуваних робіт та підвищення точності моделі розрахунку. В результаті кластеризації об'єкти оцінювання об'єднуються в кластери на основі значень характеристик, що впливають на формування їх ринкової ціни. Ці характеристики називають оцінними факторами, або факторами.

Процес кластеризації земельних ділянок для формування оцінок їх вартості передбачає виконання таких дій:

- поділ об'єктів оцінки на кластери;
- отримання в кожному кластері розрахункових цін об'єктів оцінки;
- розрахунок в кожному кластері коефіцієнта визначеності й середньоквадратичного відхилення.

Перевагою оцінювання вартості земельних ділянок на основі їх кластеризації є можливість застосування спільних узагальнених показників, які є характерними для однотипних земельних ділянок. Це дає змогу підвищити об'єктивність та достовірність результатів оцінювання, а також значно зменшити кількість операцій визначення базових показників, необхідних для оцінювання земельних ділянок.

Аналіз останніх публікацій та досліджень

Аналіз можливостей кластеризації кадастрових даних

Термін “кластерний аналіз” вперше вжив англійський вчений Р.Тріон в 1939 році, він охоплює 100 різних алгоритмів [4]. Можливість та доцільність кластерного аналізу як одного з етапів процедури оцінювання земельних ділянок обґрунтовується, з одного боку, простотою й універсальністю методики його проведення та особливостями складу і змісту кадастрових даних. Для кадастрових даних характерні значний обсяг, велика кількість оцінок та численність однотипних показників, які характеризують земельні ділянки в процесі оцінювання. Тому доцільно розширити певні оцінювальні дії та характеристики на групи земельних ділянок. Основним завданням кластеризації земельних ділянок є формування таких однотипних груп (кластерів).

Більшість сучасних методів кластеризації запропоновано в 60-ті роки ХХ століття. Цей час характеризується величезною кількістю публікацій. Як найважливіші можна виділити роботи таких авторів: Г. Болла і Д. Холла, Дж. Мак-Кіна – щодо методу К-середніх; Р. Сокала і Дж. Снитсья, Г. Ланса і У. Вільямса, Н. Джардайн і Р. Сібсон – щодо ієрархічних процедур; Дж. Роджесу і Т. Танімото, Е.М. Бравермана, А.А. Дорофеюк, І.Б. Мучника – щодо процедур типу послідовного формування кластерів і діагоналізації; В.М. Йолкіної, Н.Г. Загоруйко – щодо еталонних алгоритмів й деякі інші дослідження [9]. Ці та багато інших авторів сформували математичну базу для застосування кластерних методів у різних галузях науки.

Кластерний аналіз (англ. **Data clustering**) – це задача розподілу заданої вибірки об’єктів (ситуацій) на підмножини, які називаються кластерами, так, щоб кожен кластер складався зі схожих об’єктів, а об’єкти різних кластерів істотно відрізнялися. Задача кластеризації потребує статистичної обробки, а також належить до широкого класу задач навчання без вчителя. Кластерний аналіз – це багатовимірний статистичний алгоритм, що виконує збір даних, які містять інформацію про вибірку об’єктів, і потім впорядковує об’єкти в порівняно однорідні групи (кластери). Кластер – група елементів, які характеризуються загальною властивістю, головна мета кластерного аналізу – знаходження груп схожих об’єктів у вибірці [12].

На відміну від завдань класифікації, кластерний аналіз у застосуванні до задач оцінювання земельних ділянок не вимагає апріорних припущень про набір даних, що не накладає обмеження на подання досліджуваних об’єктів, дозволяє аналізувати показники різних типів даних (інтервальних даних, бінарних даних). Важливим чинником є те, що змінні вимірюються в порівнянних шкалах [10].

Кластерний аналіз земельних ділянок спирається на два припущення. Перше припущення – ознаки земельного об’єкта, які розглядаються, загалом допускають бажаний розподіл сукупності об’єктів на кластер. Друге припущення – правильність вибору масштабу або одиниць вимірювання ознак.

На рис. 1 наведено узагальнену класифікацію методів кластерного аналізу, придатних для застосування у процедурах оцінювання земельних ділянок.

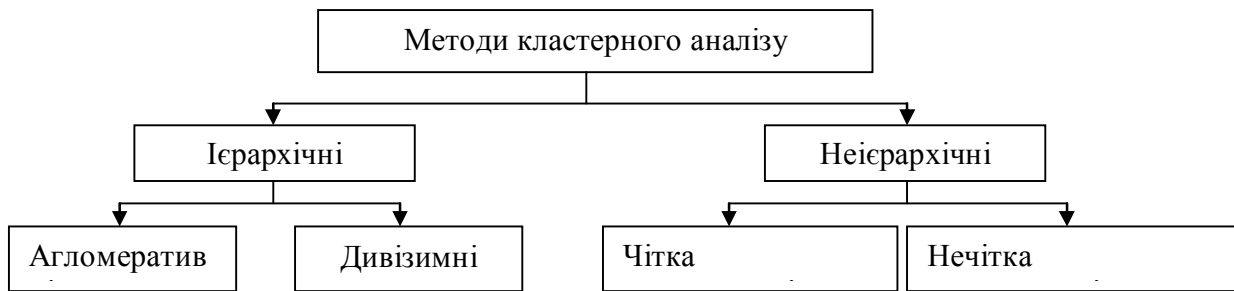


Рис 1. Класифікація методів кластеризації для оцінювання земельних ділянок

Більшість сучасних кластерних методів належать до групи ієрархічних. Головною відмінною особливістю таких методів є те, що процес об’єднання об’єктів у разі їх використання має ієрархічний характер і може бути поданий вигляді дендрограми (деревоподібної діаграми), кожен рівень якої відповідає одному кроку алгоритму. При цьому на кожному кроці кількість кластерів змінюється в бік збільшення або зменшення. Робота більшості таких методів оснований на обчисленні матриці подібності, яка містить міри відстаней – числа, що виражають “схожість” кожних двох об’єктів. Робота ієрархічних методів проілюстрована на рис. 2 [11].

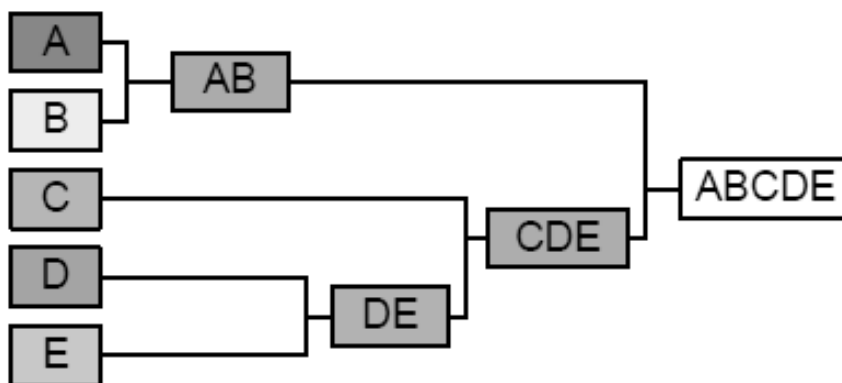


Рис. 2. Загальна схема ієрархічної кластеризації [11]

Залежно від напрямку аналізу (поділ чи об'єднання об'єктів), ієрархічні методи поділяють на агломеративні й дивізімні. Ієрархічні агломеративні методи (Agglomerative Nesting, AGNES) характеризуються послідовним об'єднанням вихідних елементів і відповідним зменшенням кількості кластерів [1].

Всі такі методи переглядають матрицю подібності розмірністю $N \times N$ (де N – кількість об'єктів) та послідовно об'єднують найсхожіші об'єкти. Послідовність об'єднань кластерів можна представити візуально у вигляді деревоподібної діаграми, яка називається дендрограмою. Кожен крок, на якому об'єднується пара об'єктів, відображається гілкою цього дерева. Дерево зображує ієрархічну організацію зв'язків між декількома точками даних. На найнижчому рівні всі точки незалежні; на наступному рівні вони об'єднуються в групи, на найвищому рівні всі об'єднуються в одну велику групу.

Для повної кластеризації цими методами на основі матриці подібності розмірністю $N \times N$ потрібно $N-1$ кроків. На першому кроці події (об'єкти) розглядаються як самостійні кластери. На останньому кроці всі події об'єднуються в одну велику групу.

Ієрархічні агломеративні методи розрізняються переважно за правилами побудови кластерів. Існує багато різних правил групування, кожне з яких породжує специфічний ієрархічний метод. Найпоширеніші чотири з них: одиничного зв'язку, повного зв'язку, середнього зв'язку та метод Варда.

Ієрархічні дивізімні (подільні) методи (DIvisive ANALysis, DIANA) є логічною протилежністю агломеративних методів [4]. На початку роботи алгоритму всі об'єкти належать одному кластеру, який на наступних кроках ділиться на менші кластери, в результаті утворюється послідовність груп. Загальна схема роботи ієрархічних дивізімних методів наведена на рис. 4.

Серед неієрархічних методів кластеризації на особливу увагу заслуговують ітеративні методи.

На відміну від ієрархічних методів, які потребують обчислення і зберігання матриці подібностей між об'єктами розмірністю $N \times N$, ітеративні методи працюють безпосередньо з первинними даними.

Тому з їх допомогою можливо обробляти доволі великі множини даних. І навіть більше, ітеративні методи виконують кілька переглядів даних і можуть компенсувати наслідки поганого вихідного розподілу даних. Ці методи породжують кластери одного рангу, які не є вкладеними і тому не можуть бути частиною ієрархії.

Більшість властивостей ітеративних методів групування можна описати за допомогою трьох основних чинників: вибору вихідного розподілу, типу ітерації та статистичного критерію. Ці фактори можуть поєднуватися по-різному, утворюючи алгоритми відбору даних для визначення оптимального розбиття. Різні комбінації ведуть до розроблення методів, що породжують різні результати під час роботи з тими самими даними [7].

Порівняльний аналіз ієрархічних і неієрархічних методів кластеризації

Перед проведенням кластеризації в аналітика може виникнути питання: якій групі методів кластерного аналізу надати перевагу?

Вибираючи між ієрархічними й неієрархічними методами, необхідно враховувати їхні особливості.

Неієрархічні методи виявляють вищу стійкість стосовно шумів і викидів, некоректного вибору метрики, введення незначущих змінних у набір, що бере участь у кластеризації. Ціною, що доводиться платити за ці переваги методу, є слово “априорі”. Аналітик повинен заздалегідь визначити кількість кластерів, кількість ітерацій або правило зупинки, а також деякі інші параметри кластеризації. Це особливо складно фахівцям-початківцям.

Якщо немає припущень щодо кількості кластерів, рекомендують використати ієрархічні алгоритми. Однак якщо обсяг вибірки не дозволяє це зробити, можливий шлях – провести низку експериментів з різною кількістю кластерів, наприклад, почати розбивку сукупності даних з двох груп і, поступово збільшуючи їх кількість, порівнювати результати. За рахунок такого “варіювання” результатів досягається доволі велика гнучкість кластеризації.

Ієрархічні методи, на відміну від неієрархічних, відмовляються від визначення кількості кластерів, а будують повне дерево вкладених кластерів.

Складності ієрархічних методів кластеризації: обмеження обсягу набору даних, вибір міри близькості, негнучкість отриманих класифікацій.

Перевага цієї групи методів порівняно з неієрархічними методами – їх наочність і можливість одержати детальне подання структури даних.

Використовуючи ієрархічні методи, можливо доволі легко ідентифікувати викиди в наборі даних й, у результаті, підвищити якість даних. Ця процедура є основою двокрокового алгоритму кластеризації. Такий набір даних надалі можна використати для проведення неієрархічної кластеризації.

Існує ще один аспект, про який необхідно згадати. Це питання кластеризації всієї сукупності даних або ж її вибірки. Названий аспект важливий для обох розглянутих груп методів, однак він критичніший для ієрархічних методів. Ієрархічні методи не можуть працювати з великими наборами даних, а використання деякої вибірки, тобто частини даних, могло б дозволити застосовувати ці методи.

Результати кластеризації можуть не мати достатнього статистичного обґрунтування. З іншого боку, під час розв'язання задач кластеризації допустима нестатистична інтерпретація отриманих результатів, а також доволі велика розмаїтість варіантів поняття кластера. Така нестатистична інтерпретація дає можливість аналітикові одержати результати кластеризації, які задовольняють його, що у разі використання інших методів часто доволі складно.

Цілі та завдання статті

Основною метою статті є розроблення процедур та алгоритму розбиття земельних ділянок на кластери із використанням інтелектуального аналізу даних. Для досягнення цілі в роботі виконано такі завдання: дослідження та обґрунтування можливостей застосування та вибір методів і засобів аналізу даних у процесах формування ціни земельних ділянок; розроблення порядку формування кластерів земельних ділянок на основі моделей, які ґрунтуються на застосуванні методу k-середніх; апробація розроблених вирішень на основі даних земельного кадастру Стрийського району Львівської області.

Основні результати досліджень

Вибір та обґрунтування методики кластеризації земельних ділянок

Для розв'язання задачі кластеризації земельних ділянок вибрано деревоподібну кластеризацію. Об'єднання, або метод деревоподібної кластеризації, використовується для формування кластерів відмінності або відстані між об'єктами. Саме відстань між земельною ділянкою та певним центром відліку є одним з чинників, який поряд з іншими її характеристиками суттєво впливає на ціну ділянки. Тому для проведення кластеризації земель в цій роботі використано показники відстані від центра населеного пункту та відстані від обласного центра. Ці відстані можуть визначатися в одновимірному або багатовимірному просторі. Для розв'язання задачі кластеризації земельних ділянок найприроднішим є прямиий шлях обчислення евклідових відстаней між об'єктами в багатовимірному просторі [5].

Евклідова відстань є геометричною відстанню в багатовимірному просторі й обчислюється за формулою:

$$L = \sqrt{\sum_i (x_i - y_i)^2} . \quad (1)$$

Особливістю кластеризації земельних ділянок є те, що евклідова відстань (і її квадрат) обчислюється за початковими, а не за стандартизованими даними. Це звичайний спосіб її обчислення, який має певні переваги (наприклад, відстань між двома об'єктами не змінюється після введення в аналіз нового об'єкта, який може виявитися викидом). Проте на відстані можуть сильно впливати відмінності між осями, за координатами яких обчислюють ці відстані [7, 8].

Для оцінювання вартості земельних ділянок із застосуванням кластеризації для підвищення відносної ваги віддалених об'єктів застосовують такий прийом, як квадрат стандартної евклідової відстані. В такому випадку відстань подають як

$$L = \sum_i (x_i - y_i)^2. \quad (2)$$

В окремих випадках, для виконання кластеризації земельних ділянок, розміщених на незначних віддальх одна від одної, доцільно застосовувати як параметр кластеризації відстані міських кварталів (мангеттенську відстань). Ця відстань є сумою модулів різниць по координатах. Здебільшого ця міра відстані призводить до таких самих результатів, як і використання відстані Евкліда. Проте у разі використання такого способу формування відстані вплив окремих великих різниць (викидів) зменшується. Мангеттенську відстань обчислюють за формулою:

$$L = \sum_i |x_i - y_i|. \quad (3)$$

Якщо необхідно визначити дві земельні ділянки, вони мають бути кваліфіковані як елементи різних кластерів за умови їх відмінності за значеннями одного з вимірів, доцільним є застосування відстані Чебишева. Відстань Чебишева за одним з показників земельних ділянок обчислюється за формулою:

$$L = \text{Max}|x_i - y_i|. \quad (4)$$

Для випадків, у яких параметри кластеризації земельних ділянок визначено у номінативній шкалі, виправдане застосування міри віддалі у формі відсотка невідповідності (Percent disagreement), який обчислюється за формулою (5):

$$L = \frac{\text{Кількість}(x_i \neq y_i)}{i}. \quad (5)$$

Якщо абсолютні значення та різниці між оцінюваними параметрами земельних ділянок є несуттєвими, а важливішою є наявність зв'язку між ними – $L = 1 - r$, доцільно використовувати 1-коефіцієнт кореляції Пірсона (1-Pearson r). Ця міра чутлива лише до схожості профілів об'єктів, а її застосування у кластеризації земельних ділянок призводить до результатів, близьких до факторного аналізу, – кластери можуть наближатися до факторів.

Оцінка якості кластеризації земельних ділянок

Оцінка якості кластеризації земельних ділянок у задачах оцінювання є важливим кроком, який обґрунтовує та підтверджує достовірність результатів і їх якість. Розв'язання задачі оцінювання якості здійснюють на основі таких процедур:

- ручна перевірка;
- установа контрольних точок і перевірка на отриманих кластерах;
- визначення стабільності кластеризації додаванням у модель нових змінних;
- створення й порівняння кластерів з використанням різних методів.

Різні методи кластеризації можуть створювати різні кластери, і це є нормальним явищем. Однак створення схожих кластерів різними методами вказує на правильність кластеризації [8].

Застосування алгоритму k -середніх для кластеризації земельних ділянок

Для розв'язання задачі кластеризації земельних ділянок у процесі їх оцінювання вибрано метод, що ґрунтується на застосуванні алгоритму k -середніх. Цей метод є найпоширенішим серед неієрархічних методів (також називається швидким кластерним аналізом). Повний опис алгоритму можна знайти в роботі Хартігана і Вонга [2]. Основним чинником, який обґрунтовує доцільність такого вибору, є те, що для можливості використання цього методу необхідно мати гіпотезу про найімовірнішу кількість кластерів [6] на відміну від ієрархічних методів, що не потребують попередніх припущень щодо кількості кластерів. Під час оцінювання земельних ділянок попередньо визначена кількість кластерів є важливим показником, оскільки дає змогу встановлювати та реалізувати обмеження на кількість цінкових градацій, яка не може бути невизначеною або дуже великою.

Алгоритм k -середніх будує k кластерів, які містять дані земельних ділянок, розташованих на можливо великих відстанях один від одного. Основний тип завдань, які вирішує алгоритм k -середніх – наявність припущень (гіпотез) щодо кількості кластерів, при цьому вони повинні бути різні настільки, наскільки це можливо. Вибір числа k може ґрунтуватися на результатах маркетингових досліджень вартості земель, нормативних документах щодо оцінювання вартості земель, даних земельного кадастру тощо.

Кластеризація земельних ділянок методом k -середніх розподіляє вхідний набір векторів, які містять їх ключові характеристики, за k кластерами $S_i = i(1,2,...k)$, з кожним з яких пов'язано центроїд c_i . Позначимо безліч вхідних векторів $S = \{x\}$, $|S| = n$. Нехай $D(x, c)$ – відстань між вектором x і центроїдом c . У цьому випадку для формування міри відстані між векторами використано незважену евклідову відстань між вектором та центроїдом.

Позначимо множину центроїдів, отриманих на ітерації t , $SC_t = \{c_i\}$. Алгоритм кластеризації k -середніх в застосуванні до земельних ділянок передбачає виконання таких кроків:

1. Встановлюємо $t=0$ і задаємо початкове розташування центроїдів SC_0

2. Для заданої множини центроїдів SC_t виконуємо дії, описані в п. 2.1 і 2.2, й отримуємо покращену множину центроїдів SC_{t+1} :

2.1. Знаходимо таке розбиття множини векторів характеристик земельних ділянок S , яке розподіляє їх за k кластерами $S_i = i(1,2,...k)$ і задовольняє умову $S_i = \{x | D(x, c_i) \leq D(x, c_j) \forall j \neq i\}$

2.2. Обчислюємо центроїд c_i для кожного кластера $S_i = i(1,2,...k)$, щоб отримати нову множину центроїдів SC_{t+1} :

$$c_{ij} = \frac{1}{m_i} \cdot \left(\sum_{l=1}^{m_i} x_{ij} \right), j = 1, 2, \dots, d \quad (6)$$

де m_i – кількість векторів, що належать кластеру S_i .

3. Обчислюємо сумарне спотворення $E^2 = \sum_{x \in S} D^2(x, c)$. Якщо воно відрізняється від отриманого на попередній ітерації на досить малу величину, завершуємо процес. В іншому випадку присвоюємо $t \leftarrow t + 1$ і повертаємося до кроку 2.

Алгоритм гарантовано збігається за скінченну кількість ітерацій. Похибка кластеризації та кількість ітерацій залежать від початкового вибору центроїдів, тому звичайною практикою є запуск k -середніх кілька разів з різними початковими кандидатами в центроїди [2].

Кластеризація земель населених пунктів Стрийського району. Ціна земельної ділянки є обов'язковою умовою будь-якої угоди, вона виконує функції механізму, який регулює ринковий оборот об'єктів земельної власності. Ціна землі завжди є конкретною величиною і може істотно відрізнятись як у часі, так і в просторі, перебуваючи під впливом багатьох факторів: загальної економічної ситуації, рівня доходів та платоспроможного попиту, правового середовища, коливань попиту та пропозиції, особливих мотивів і особистих інтересів конкретного покупця і продавця, які тісно пов'язані й взаємообумовлюють один одного.

За ступенем впливу на ціноутворення перераховані фактори можна поділити на загальні й локальні. Загальні фактори, що визначають стан і динаміку ринку, впливають на ціни і однаковою мірою діють протягом певного періоду часу на різні типи ділянок, розташованих у конкретній місцевості. Локальні фактори впливають на ціну конкретної земельної ділянки чи групи земельних ділянок і пов'язані передусім з характеристикою власне ділянки і місця її розташування.

Місцю розташування належить провідна роль у формуванні цін на земельні ділянки, оскільки за умови сприятливого значення цього фактора власник землі має можливість отримати найвищий дохід з одиниці площі ділянки.

З економічного погляду поняття місця розташування охоплює:

- транспортну доступність;
- наявність необхідної інфраструктури;
- характер використання прилеглої території,

Цілком природно, що ціна на земельні ділянки залежатиме від місця населеного пункту в системі виробництва і розселення, що прямо пов'язано з чисельністю його населення. Порушити це правило може хіба що особливий статус поселення (адміністративний центр чи курорт) або входження до складу агломерації великого міста.

Експертна грошова оцінка земельної ділянки здійснюється на основі таких методичних підходів:

- капіталізація чистого операційного або рентного доходу (пряма і непряма);
- зіставлення цін продажу подібних земельних ділянок;
- урахування витрат на земельні поліпшення.

За основу визначення вартості земельних ділянок зіставленням цін продажу подібних земельних ділянок беруть ціни продажу тих ділянок, які за факторами, що впливають на їх вартість, достатньою мірою збігаються з оцінюваною ділянкою.

Для реалізації експертної грошової оцінки на основі зіставлення цін продажу подібних ділянок зібрано інформацію щодо вартості ділянок у Стрийському районі та застосовано метод кластерного аналізу для розподілу на кластери типових ділянок із зазначеною вартістю.

Для проведення кластеризації розроблено програму в середовищі C++Builder та використано дані щодо земельних ділянок Стрийського району. Вхідними даними вибрано відстань до обласного центру та розміщення відносно центру населеного пункту.

Головне вікно програми зображено на рис. 3. Для виконання програми використано файл з вхідними даними кадастру, кількість кластерів та максимальну кількість ітерацій, яку необхідно виконати (якщо стабілізація кластерів не відбудеться за меншу кількість ітерацій).

Після проведення кластеризації земельних ділянок населених пунктів Стрийського району отримано такі результати розбиття на кластери (рис. 4).

	Відстань_до_Львова	Відстань_до_центру	Вартість_1м.кв.
Стрий	70	25	60
Нежухів	76	5	133
Угерсько	65	7	200
Добряни	66	5	107
Станків	77	7	125
Довге	80	10	200
Моршин	75	10	480
Дашава	80	14	60
Подорожжє	92	22	80
Нежухів	75	5	160
Моршин	80	14	200

Рис. 3. Вікно програми із завантаженими даними

	Кластер 1	Кластер 2	Кластер 3	Кластер 4
	Відстань_до	Відстань_до	Вартість_1м	
Моршин	75	10	480	
Моршин	80	10	440	

Рис. 4. Результат кластеризації земельних ділянок Стрийського району

До першого кластера ввійшли землі міста Моршин, вартість яких коливається в межах 440–480 грн. за м. кв. Незважаючи на те, що ці земельні ділянки розташовані на достатньо великій відстані до обласного центру, вартість їх найвища, оскільки Моршин має особливий статус поселення (курорт).

До другого кластера ввійшли землі таких населених пунктів: Угерсько, Нежухів, Довге. Вартість ділянок коливається у межах 130–200 грн. за м. кв.

До третього кластера: Ходовичі, Гірне, Любинці, Грабовець, Дуліби, Подорожнє, Добряни, Станків.

Вартість 80–125 грн. за м. кв.

До четвертого кластера ввійшли землі таких населених пунктів: Фалиш, Завадів, Верчани, Голобутів, Бережниця, Братківці, Жулин, Стрільків, Дашава, Вівня, Стрий, Лисятичі, Розгірче.

Вартість 30–60 грн. за м. кв. Вхідження земельних ділянок м. Стрий до цього кластера зумовлена їх місцем розташування відносно центра міста.

Висновки

У роботі розв'язано задачу кластеризації земельних ділянок у процесах визначення їх ринкової вартості. Для досягнення мети проведено аналіз та обґрунтування можливості й доцільності застосування кластеризації як елемента процедури оцінювання вартості земельного ресурсу, визначено основні підходи та способи кластеризації, описано застосування алгоритму k-середніх для розв'язання задач кластеризації земельних ділянок.

Метод кластерного аналізу, що використовує алгоритм k-середніх, дає змогу здійснити кластеризацію земель за незначною кількістю ключових характеристик з достатньою точністю та достовірністю. На основі досліджень розроблено комплекс програмних засобів та проведено апробацію запропонованих вирішень на даних про земельні ділянки населених пунктів Стрийського району. Результатом застосування розробленої методики та засобів, які її реалізують, є отримані в кожному кластері розрахункові ціни земельних ділянок для здійснення експертної грошової оцінки зіставленням цін подібних ділянок.

Перевагами використання цього методу є простота та швидкість використання, зрозумілість і прозорість алгоритму. Недоліками є чутливість до вибору початкових значень центрів кластерів та труднощі, пов'язані з визначенням прогнозованої кількості кластерів.

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. Hartigan J. A. *Algorithm 136. A k-means clustering algorithm* / J. A. Hartigan, & M. A. Wong // *Applied Statistics* No 28, 100.– 1978. 3. *Refining initial points for KMeans clustering [Електронний ресурс]*. – Режим доступу: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.44.5872>. 4. Tryon R.C. *Cluster analysis.[Text]* / R.C. Tryon L, D.E.Bailey. – McGraw-Hill Inc.,US – 1970. 5. *Алгоритмы кластеризации на службе Data Mining [Електронний ресурс]*. – Режим доступу: <http://www.olap.ru/home.asp?artId=154>.

6. Барсегян А.А. Методы и модели анализа данных OLAP и DataMining [Текст] / А. А Барсегян, М.С. Куприянов, В. В. Степаненко, И. И.Холод. – СПб: БХВ-Петербург. – 2008. 7. Енюков И.С. Факторный, дискриминантный и кластерный анализ [Текст] / Енюков И.С. – М.: Финансы и статистика, 1989. 8. Кулаичев А. П. Методы и средства комплексного анализа данных [Текст] / Кулаичев А.П. – М: ИНФРА-М. 2006. 9. Мандель И.Д. Кластерный анализ [Текст] /Мандель И.Д. – М.: Финансы и статистика. – 1988. 10. Пакет статистической обработки. Кластерный анализ при большом количестве наблюдений [Электронный ресурс]. – Режим доступа: http://www.sati.archaeology.nsc.ru/stat/methods_info.php. 11. Уиллиамс У. Т., Ланс Д. Н. Методы иерархической классификации. Статистические методы для ЭВМ [Текст] // под ред. М. Б. Малютов. – М.: Наука, 1986. 12. Чубукова И. А. Data Mining: учебное пособие [Текст] / И. А. Чубукова. — М.: Интернет-университет информационных технологий : БИНОМ: Лаборатория знаний, 2006.

УДК 004.896

М.В. Євланов

Харківський національний університет радіоелектроніки

ПАТЕРНИ ПРОЕКТУВАННЯ ВИМОГ ДО ІНФОРМАЦІЙНОЇ СИСТЕМИ

© Євланов М.В., 2014

Запропоновано визначення терміна “патерн проектування вимог до інформаційної системи”. Запропоновано теоретико-категорну модель такого патерна, визначено місце патернів у інформаційній технології формування та аналізу вимог до інформаційної системи.

Ключові слова: інформаційна система, вимога, патерн, теоретико-категорна модель.

This paper is devoted to the definition of "pattern design requirements for an information system". A theoretic-categorical model of the pattern is proposed, the place of patterns in the information technology development and analysis of requirements for information systems is defined.

Key words: information system, requirement, pattern, theoretic-categorical model.

Вступ

Створення сучасних інформаційних систем (ІС) являє собою сукупність процесів, описаних у міжнародних стандартах ISO/IEC 15288:2002 та ISO/IEC 12207:2008. Ці стандарти визначають суть та основні особливості процесів проектування архітектури ІС та систем, що її забезпечують [1, 2]. Тепер ці стандарти визнані багатьма країнами світу, серед яких й Україна.

Відповідно до положень стандартів ISO/IEC 15288:2002 та ISO/IEC 12207:2008, основним ресурсом процесів проектування архітектури ІС є вимоги, що ставляться до цих систем. Тому розв’язання задачі синтезу архітектури ІС як сукупності функцій системи значною мірою залежить від того, наскільки ефективно та якісно сформульовано вимоги до цієї системи, а витрати часу та ресурсів на створення цієї ІС великою мірою визначаються результатами виконання процесів визначення вимог правласників системи та аналізу вимог.

Одним з напрямів зменшення витрат ресурсів на виконання процесів визначення вимог правласників системи та аналізу вимог є повторне використання вимог, поставлених до розроблених раніше ІС, під час створення нової ІС або модернізації чи розвитку ІС, що експлуатуються. Цей підхід не тільки може привести до скорочення часу виконання ІТ-проекту створення ІС, але й дозволить зменшити витрати ресурсів за рахунок використання основних компонентів видів