

Urszula Litwin, Alicja Malczewska

Zastosowanie drzew decyzyjnych do oceny wpływu cech niezabudowanych nieruchomości gruntowych na ich wartość

Acta Scientiarum Polonorum. Administratio Locorum 3/1, 113-124

2004

Artykuł został zdigitalizowany i opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

ZASTOSOWANIE DRZEW DECYZYJNYCH DO OCENY WPLYWU CECH NIEZABUDOWANYCH NIERUCHOMOŚCI GRUNTOWYCH NA ICH WARTOŚĆ

Urszula Litwin¹, Alicja Malczewska²

¹Katedra Geodezyjnego Urządzania Terenów Wiejskich, Akademia Rolnicza
w Krakowie

Biurow Szacowania i Obsługi Nieruchomości, Bolechowice

Abstrakt: Artykuł przedstawia możliwości zastosowania drzew decyzyjnych do redukcji liczby atrybutów wartościotwórczych uwzględnianych w masowej wycenie. Poruszono problem oceny wpływu cech nieruchomości na ich wartość oraz przedstawiono sposób redukcji liczby atrybutów wartościotwórczych w procesie powszechnej taksacji, a także możliwości zastosowania technik KDD (Knowledge Discovery in Databases) – konkretnie drzew decyzyjnych – do analizy danych transakcyjnych na rynku nieruchomości.

Słowa kluczowe: atrybuty wartościotwórcze, wycena nieruchomości, masowa wycena, drzewa decyzyjne, sztuczna inteligencja

WSTĘP

Wprowadzenie podatku *ad valorem* będzie wymagać oszacowania dziesiątków milionów nieruchomości. Wycena tak znaczącej ich liczby wiąże się z koniecznością zastosowania metod statystyki i ekonometrii. Idąc jeszcze dalej, trudno oprzeć się pokusie i nie podjąć próby wykorzystania w tym celu jeszcze bardziej zaawansowanych metod analizy danych, np. sztucznej inteligencji.

Problem oceny wpływu cech poszczególnych nieruchomości uwzględnianych podczas szacowania nieruchomości nie jest zagadnieniem nowym – występuje przy wycenie nieruchomości w podejściu porównawczym i od kilku

Adres do korespondencji – Corresponding author: dr hab. inż. Urszula Litwin, Katedra Geodezyjnego Urządzania Terenów Wiejskich, Akademia Rolnicza w Krakowie, ul. Balicka 253 a, 30-198 Kraków

lat jest przedmiotem badań naukowych. Nabiera jednak szczególnego znaczenia przy masowej wycenie, gdzie ze względów czysto technicznych (trudność i pracochłonność pozyskania danych) niezbędne jest ograniczenie do minimum liczby atrybutów.

Planowana w Polsce powszechna taksacja, w której świetle analizowano problem, jest procesem zdefiniowanym przez przepisy w sposób ogólny. Ustawa o gospodarce nieruchomościami oraz rozporządzenie w sprawie powszechnej taksacji nieruchomości [Ustawa... 1997, Rozporządzenie... 2001] nie określają w precyzyjny sposób ani metod masowej wyceny, ani wymaganego stopnia, w jakim wartość katastralna ma być zbliżona do wartości rynkowej.

Obowiązujące przepisy wskazują na to, że w odniesieniu do gruntów wartość katastralna będzie obliczana wg następującego algorytmu:

- wyznaczenie stref taksacyjnych charakteryzujących się podobnymi czynnikami cenotwórczymi,
- ustalenie jednostkowych wartości gruntów położonych w tych strefach,
- obliczenie wartości katastralnych poszczególnych działek jako iloczynu skorygowanej wartości jednostkowej w strefie i współczynników korygujących odpowiadających poszczególnym cechom wartościotwórczym.

PRZEDMIOT BADAŃ I ZASTOSOWANE METODY REALIZACJI ZADAŃ

Przedmiotem badań były niezabudowane nieruchomości gruntowe w terenach wiejskich i podmiejskich. Analizą objęto rynek nieruchomości podkrakowskiej gminy Zabierzów. Jego sześcioletni monitoring pozwolił stwierdzić, że liczba zawieranych na tym rynku transakcji jest wystarczająca, aby na tej podstawie móc wnioskować o pewnych ogólnych zjawiskach oraz by do jego analizy z powodzeniem stosować metody statystyczne.

Koncepcja badań. Merytoryczny zakres badań w analizowanym obszarze badawczym obejmuje określenie, które z cech niezabudowanych nieruchomości gruntowych mają największy wpływ na wartość oraz możliwości wykorzystania drzew decyzyjnych do analiz rynku nieruchomości. Poszczególne atrybuty wartościotwórcze oceniono poprzez porównanie wyników klasyfikacji dokonanej za pomocą drzew decyzyjnych (algorytm M5 Prime) z opinią ekspertów (rzeczoznawców majątkowych). Badania przebiegały według schematu:

1. Ustalenie, na podstawie literatury przedmiotu, które czynniki wartościotwórcze wymienia się jako istotne.
2. Ustalenie wpływu atrybutów na ceny nieruchomości na podstawie analizy danych transakcyjnych za pomocą drzewa decyzyjnego.
3. Ocena wpływu cech nieruchomości na ich wartość dokonana na podstawie testu ekspertów (wywiadu przeprowadzonego w środowisku rzeczoznawców majątkowych).
4. Porównanie wyników.

METODA BADAWCZA

Sztuczna inteligencja. Drzewa decyzyjne. Tam, gdzie zachodzi potrzeba odnalezienia prawidłowości w dużych i niekoniecznie kompletnych zbiorach danych, zasadne jest zastosowanie sztucznej inteligencji. Rozpoznanie prawidłowości zachodzących w zbiorze danych (tzw. zbiorze uczącym) pozwala na przewidywanie zjawisk reprezentowanych przez ten zbiór. Algorytmy odkrywania wiedzy w bazach danych (KDD) umożliwiają między innymi wyodrębnianie w zbiorach danych grup charakteryzujących się naturalnym podobieństwem (klasteryzacja, klasyfikacja), określanie zależności między atrybutami (reguły asocjacyjne), określanie trendów i odnajdywanie anomalii [Cichosz 2000]. W badaniach do analiz danych transakcyjnych posłużono się drzewami decyzyjnymi, a wynika to z następujących przesłanek:

- wśród technik KDD drzewa decyzyjne charakteryzują się bardzo przejrzystą strukturą, umożliwiającą śledzenie przebiegu analizy;
- nie podjęto jeszcze prób ich stosowania do analiz rynku nieruchomości.

Analiza danych za pomocą drzew decyzyjnych. Do klasyfikacji zastosowano algorytm M5.Prime zbudowany na bazie C4.5 Quinlana. Algorytm działa na zasadzie drzewa decyzyjnego i służy do predykcji ciągłych wartości numerycznych [Witten, Frank 1999]. Za jego pomocą można obliczyć nieznanne wartości atrybutów numerycznych na podstawie przykładów pochodzących ze zbioru uczącego. W wyniku działania algorytmu zbiór danych jest dzielony na grupy, tzw. klasy, do których są zaliczane przykłady charakteryzujące się pewnym podobieństwem cech (w tym przypadku ceny i atrybutów wartościotwórczych). W każdym kolejnym węźle drzewa, na podstawie testu wartości kolejnego, wybranego przez algorytm atrybutu, następuje podział danych na dwa podzbiory (gałęzie) w taki sposób, że przykłady zaliczone do różnych podzbiorów muszą być jak najbardziej zróżnicowane pod względem ceny [Cichosz 2000].

Algorytm M5.Prime pozwala na wykonywanie obliczeń za pomocą dwóch modeli: model tree oraz regression tree. W obydwu przypadkach jest budowane drzewo decyzyjne według zasad opisanych wyżej, po czym następuje tzw. jego cięcie, zapobiegające zbyt silnemu rozrastaniu się drzewa (głębokość cięcia drzewa określa parametr pruning factor). Podzbiory uzyskane na końcach gałęzi w wyniku cięcia drzewa noszą nazwę liści lub klas. Obliczenie wartości atrybutu, ze względu na który wykonywana jest klasyfikacja (w tym przypadku ceny), może odbywać się w dwojaki sposób: bądź jako średnia arytmetyczna wartości tego atrybutu w przykładach zaliczonych do klasy (regression tree) lub na podstawie równania regresji liniowej (model tree) [Witten, Frank 2000]. Mając na uwadze cel analizy oraz przejrzystość wyników, do obliczeń wybrano model regression tree z cięciem pruning factor = 2. Metodę, opcję testowania oraz wartości parametru pruning factor wybrano po przeprowadzeniu wstępnych analiz, przy czym jako kryteria

przyjęto jak najlepsze dopasowanie modelu (miarą dopasowania jest tutaj współczynnik korelacji między ceną transakcyjną a ceną predykowaną), czasochłonność obliczeń oraz przejrzystość wyników.

Wynikiem obliczeń jest schemat drzewa decyzyjnego zawierający:

- nazwy atrybutów, umieszczone w węzłach drzewa, ze względu na które następowały kolejne jego podziały;
- wartości atrybutów stanowiące kryterium podziału w gałęziach wychodzących z poszczególnych węzłów;
- przedstawione w liściach drzewa, predykowane wartości atrybutu, ze względu na który jest wykonywana klasyfikacja wraz z liczbą przykładów zaklasyfikowanych do poszczególnych liści.

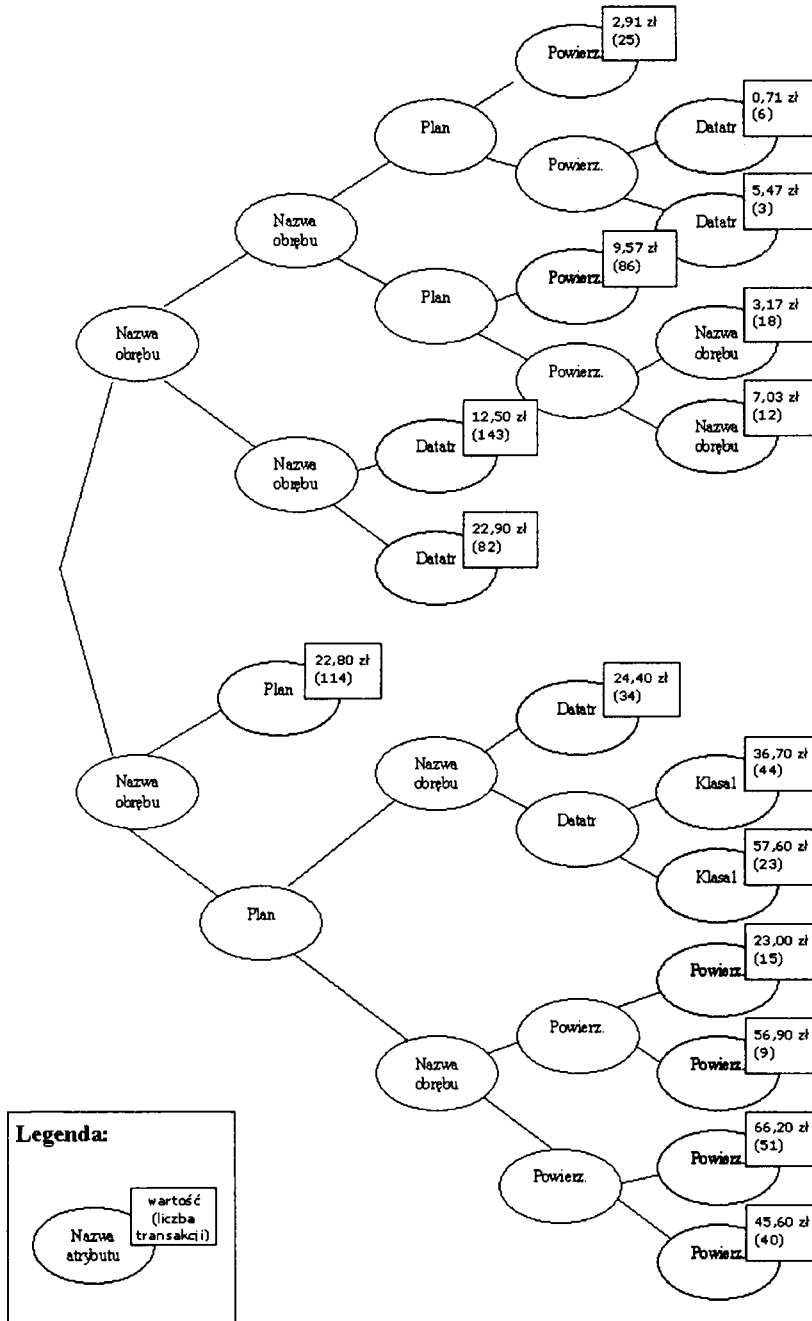
Pozyskanie danych. Dane wejściowe do analizy stanowią opisy transakcji dokonanych w latach 1997-2002 na terenie obszaru badawczego. Bazę danych zbudowano na podstawie następujących materiałów źródłowych: aktów notarialnych, numerycznej mapy ewidencji gruntów, mapy zasadniczej, amatorskich fotografii planu zagospodarowania przestrzennego, numerycznego planu zagospodarowania przestrzennego oraz barwnych zdjęć lotniczych w skali 1: 26 000. Zbiór liczy 1150 rekordów.

Analiza danych. Celem analizy danych było ustalenie, które z cech nieruchomości reprezentowanych przez poszczególne atrybuty w zbiorze danych wejściowych mają istotny wpływ na cenę. Analiza polega na budowie drzew decyzyjnych dla kolejnych podzbiorów zbioru wejściowego, powstałych w wyniku redukcji kolejnych atrybutów, podobnie jak się postępuje w przypadku regresji krokowej wstecznej. Brak zmian struktury drzewa oraz wyników predykcji oznacza, że zredukowany atrybut nie ma wpływu na cenę, natomiast przejawem istotności atrybutu jest jego wystąpienie w węzłach drzewa. Obliczenia przeprowadzono w ramach dwóch podzbiorów zbioru wejściowego, tzn. oddzielnie dla gruntów o przeznaczeniu dopuszczającym zabudowę, oddzielnie dla gruntów położonych poza wyznaczonymi terenami budowlanymi.

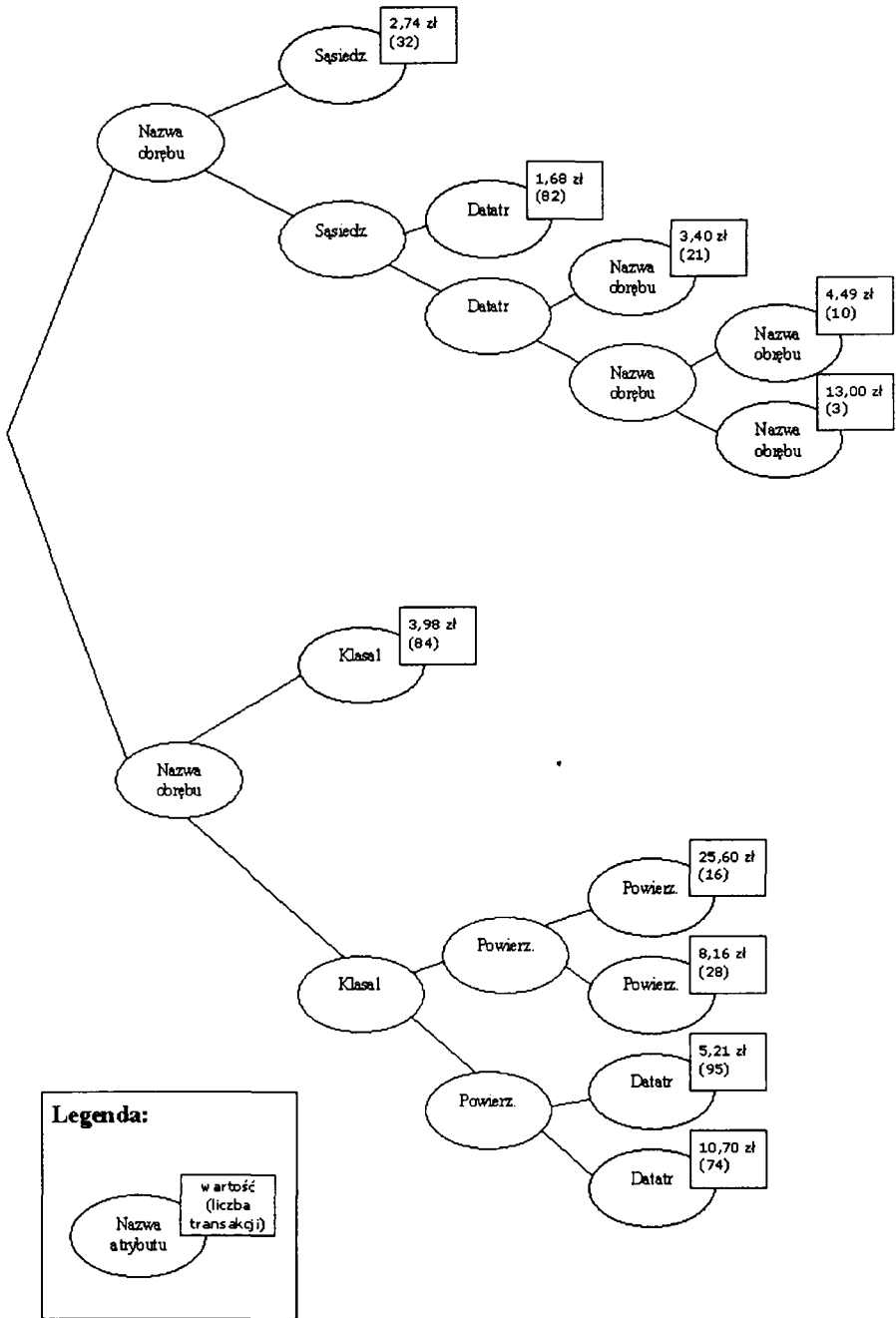
Wyniki analiz wykonanych za pomocą drzew decyzyjnych. Dla badanego lokalnego rynku nieruchomości, w odniesieniu do gruntów o przeznaczeniu dopuszczającym zabudowę (zbiór B), czynnikami wpływającymi na ich cenę są: lokalizacja, przeznaczenie gruntu, powierzchnia oraz klasa bonitacyjna. Jako cechy o mniejszym znaczeniu można uznać uzbrojenie i słuzebności. W przypadku gruntów położonych poza wyznaczonymi terenami budowlanymi (zbiór N) cechami w sposób istotny wpływającymi na cenę są: lokalizacja, klasa bonitacyjna gruntu, data transakcji, powierzchnia oraz uzbrojenie.

Ilustrację wyników, kolejno dla zbioru B i N przedstawiają rysunki 1 i 2.

Spostrzeżenia. Drzewo decyzyjne, będące rezultatem obliczeń, oprócz listy atrybutów istotnych umieszczonych w jego węzłach, zawiera informację



Rys. 1. Drzewo decyzyjne dla zbioru gruntów o przeznaczeniu dopuszczającym zabudowę (B)
 Fig. 1. Decision tree for the data set concerning building plots (B)



Rys. 1. Drzewo decyzyjne dla zbioru gruntów o przeznaczeniu wykluczającym zabudowę (N)
 Fig. 1. Decision tree for the data set concerning non-building plots (N)

o predykowanej cenie przykładów, które na podstawie wartości poszczególnych atrybutów są zaliczane do danej klasy (wartości w liściach). Na podstawie właściwości drzewa można zbudować uproszczoną metodę obliczania wartości katastralnej, przyjmując, że jednostkowe wartości katastralne poddanych analizie działek (nowych, nie należących do zbioru uczącego przykładów) są równe predykowanym cenom w liściach, do których zostały zaliczone.

Badania ankietowe. Eksperckiej oceny wpływu cech nieruchomości na ich wartość dokonano na podstawie wywiadu przeprowadzonego w środowisku rzeczoznawców majątkowych. Formularz ankiety opublikowano w Internecie i ogłoszono na liście dyskusyjnej rzeczoznawców majątkowych. Zadaniem respondentów było określenie, jaki jest, ich zdaniem, wpływ poszczególnych atrybutów na wartość rynkową nieruchomości. Ankietowanym poproszono o listę cech nieruchomości wraz z rangami tych cech w odniesieniu do niezabudowanych nieruchomości gruntowych położonych w obszarach gmin wiejskich i podmiejskich.

Ponieważ obszar badań w zakresie analizy danych transakcyjnych jest ograniczony do jednego lokalnego rynku nieruchomości (podkrakowskiej gminy Zabierzów), uznano za niezbędne przeprowadzenie szczegółowego wywiadu wśród rzeczoznawców działających na terenie Krakowa i okolic. Internetowa forma ankiety charakteryzowała się bardzo niską frekwencją (jedynie jedenastu respondentów), dlatego prośbę o wzięcie udziału w badaniach skierowano bezpośrednio do trzydziestu rzeczoznawców, spośród których odpowiedzi udzieliło szesnastu.

Opracowanie wyników. Ponieważ ankietowani zostali poproszeni nie tylko o ocenę wpływu cech gruntów na wartość, ale także o ich nazwanie (w ankiecie nie wymieniono cech wartościotwórczych, nie chcąc w żaden sposób sugerować odpowiedzi), pierwszym etapem analizy wyników ankiety było sporządzenie tezauryusa. Z uwagi na brak usystematyzowanego nazewnictwa w odniesieniu do cech wartościotwórczych nieruchomości, w celu interpretacji wyników ankiety sporządzono tabele wyrazów bliskoznacznych, w których sklasyfikowano nazwy atrybutów użyte przez ankietowanych.

Forma ankiety pozostawiła respondentom pewną dowolność nadawania cechom rang, w wyniku czego suma rang przypisanych cechom przez poszczególne osoby w kilku przypadkach była różna od 100. W celu dokonania późniejszych porównań przeprowadzono ich normalizację, polegającą na proporcjonalnym zwiększeniu lub zmniejszeniu wartości rang tak, aby ich suma wynosiła 100. Następnie dla każdej cechy (atrybutu) obliczono: sumę rang, średnią rangę, liczbę i stosunek procentowy wystąpień oraz średnią rangę atrybutów niezerowych.

Wyniki badań ankietowych. Zarówno w odniesieniu do gruntów budowlanych, jak i pozostałych ankietowani wymienili w sumie po dziewięć cech. W jednym i w drugim przypadku, o uwzględnieniu opinii wszystkich

ankietowanych, ranga tych samych trzech cech (ograniczeń i służebność, warunków fizycznych oraz zagospodarowania) wynosi mniej niż dwa procent. Należy zatem uznać, że cechy te, jako najmniej istotne, można pominąć.

Najważniejszą, zdaniem ekspertów, cechą jest lokalizacja, zarówno w znaczeniu ogólnym (miejsce), jak i w sensie położenia względem innych obiektów na danym terenie. W przypadku gruntów o przeznaczeniu dopuszczającym zabudowę jako ważne cechy rzeczoznawcy wymieniają przeznaczenie w planie zagospodarowania przestrzennego, uzbrojenie, wielkość i kształt. Dla gruntów położonych w terenach rolnych najważniejsza jest lokalizacja, bonitacja, kształt i powierzchnia oraz możliwość zmiany przeznaczenia terenu w przyszłości. Skądinąd ta ostatnia cecha jest w zasadzie niemierzalna i ma raczej spekulacyjny charakter.

WYNIKI

Porównanie wyników analiz wykonanych za pomocą drzew decyzyjnych z wynikami badań ankietowych. W przypadku gruntów o przeznaczeniu dopuszczającym zabudowę różnica między zdaniem ekspertów a wynikami analizy danych dotyczy głównie kształtu działki. Cecha nie pojawiła się w zestawie wygenerowanym przez drzewo decyzyjne, ale obok powierzchni, wymienili ją rzeczoznawcy. Podobnie w przypadku uzbrojenia, które pojawia się na liście ekspertów, a w węzłach drzewa ulokowane jest tak nisko, że widoczne staje się dopiero po zmianie głębokości przycinania drzewa. Trzecia różnica – między wynikami analizy KDD a wynikami ankiety – niewspomniana przez ekspertów klasa bonitacyjna gruntu pojawia się w węzłach drzewa. Uznanie klasy bonitacyjnej gruntu jako czynnika wpływającego na wartość gruntów dopuszczających zabudowę wydaje się nieuzasadnione z punktu widzenia ekonomicznego interesu inwestora, jednak analiza danych pokazała, że usunięcie tego atrybutu ze zbioru uczącego powoduje zmiany wartości w liściach, a zatem w tym konkretnym przypadku ma wpływ na uzyskane wyniki predykcji. Sytuacje takie mogą być spowodowane wystąpieniem silnej zależności między atrybutami, np. w omawianym przypadku może to być powiązanie klasy bonitacyjnej z lokalizacją. Wyniki ilustruje tabela 1.

Porównanie wyników analizy danych dotyczących gruntów o przeznaczeniu wykluczającym zabudowę ze zdaniem ekspertów pokazuje, że obok lokalizacji ogólnej i bonitacji istotnym czynnikiem decydującym o wartości jest możliwość zmiany przeznaczenia gruntu, co w wynikach analizy obrazuje występująca również na trzecim miejscu cecha „lokalizacja szczegółowa – sąsiedztwo”. To właśnie sąsiedztwo w dużej mierze decyduje o ewentualnej przyszłej zmianie przeznaczenia gruntu. Ponieważ dane wejściowe nie zawierały atrybutu „możliwość zmiany przeznaczenia”, czego jednym z powodów jest fakt, że jest to cecha subiektywna, trudna do oceny, a zbiór danych

Tabela 1. Porównanie wyników ankietowych z wynikami analizy KDD dla gruntów budowlanych
 Table 1. Comparison of the poll results and the results of KDD analysis for building plots

Według rzeczoznawców According to appraisers	Na podstawie analizy za pomocą drzewa decyzyjnego On the basis of a decision-tree analysis
1. Lokalizacja ogólna General location	1. Lokalizacja ogólna General location
2. Przeznaczenie w PZP Land function/zoning according to a spatial management plan	2. Przeznaczenie w PZP Land function/zoning according to a spatial management plan
3. Lokalizacja szczegółowa – sąsiedztwo Precise location - neighborhood	3. Powierzchnia Area
4. Uzbrojenie Provision of technical infrastructure	4. Klasa bonitacyjna gruntu Soil quality class
5. Powierzchnia i kształt Area and shape	

wejściowych powinien być zbiorem faktów, a nie ocen, za cechą najbliższą znaczeniowo można przyjąć lokalizację szczegółową, czyli sąsiedztwo. Można założyć też istnienie związku między cechami „możliwość zmiany przeznaczenia” i „uzbrojenie”, gdyż grunt uzbrojony lub sąsiadujący z terenami uzbrojonymi ma większe szanse na zmianę przeznaczenia (obecnie obowiązujące przepisy, regulujące zasady wydawania decyzji WZ w przypadku braku planu zagospodarowania przestrzennego, [Ustawa... 2003] potwierdzają te przypuszczenia). W tabeli 2 zestawiono listę cech uzyskaną w wyniku ankiety oraz analizy za pomocą drzew decyzyjnych.

Tabela 2. Porównanie wyników ankietowych z wynikami analizy KDD dla gruntów niebudowlanych
 Table 2. Comparison of the poll results and the results of KDD analysis for non-building plots

Według rzeczoznawców According to appraisers	Na podstawie analizy KDD On the basis of a KDD analysis
1. Lokalizacja ogólna General location	1. Lokalizacja ogólna General location
2. Klasa bonitacyjna gruntu Soil quality class	2. Klasa bonitacyjna gruntu Soil quality class
3. Możliwość zmiany przeznaczenia Possibility of changing land function	3. Lokalizacja szczegółowa – sąsiedztwo Precise location – neighborhood
4. Lokalizacja szczegółowo – sąsiedztwo Precise location – neighborhood	4. Powierzchnia Area
5. Powierzchnia i kształt Area and shape	5. Uzbrojenie Provision of technical infrastructure

DYSKUSJA

Wyniki badań potwierdzają, że przy masowej wycenie gruntów można, a nawet z powodów czysto technicznych i ekonomicznych należy zredukować liczbę uwzględnianych przy szacowaniu cech do minimum. Potwierdzony badaniami fakt, że lokalizacja jest najistotniejszą cechą wartościotwórczą, pozwala twierdzić, że granice obrębów mogą stanowić pierwsze przybliżenie granic stref taksacyjnych. Spośród pozostałych istotnych cech, strefowy, choć znacznie mniej ciągły charakter ma przeznaczenie w planie zagospodarowania przestrzennego. Pozostałe atrybuty nie układają się strefami i mają charakter indywidualny. Wydaje się więc zasadne zastosowanie modelu mieszanego – strefowo atrybutowego, w którym wartość wyjściową będzie stanowić wartość ustalona na podstawie atrybutów strefowych (głównie lokalizacji), natomiast wartość katastralna będzie skorygowana (na podstawie wartości pozostałych atrybutów, uznanych na terenie danej gminy za istotne) wartością wyjściową. Ustalenie, które z cech wartościotwórczych należy uwzględnić, a które pomijać, może się odbywać na podstawie zbudowanego na danych transakcyjnych drzewa decyzyjnego. Oto schemat postępowania:

1. Zebranie informacji o transakcjach.
2. Kodowanie danych i zapis danych zawierających atrybutowy opis transakcji.
3. Podział danych na podzbiory (grunty o przeznaczeniu dopuszczającym zabudowę oraz grunty rolne i leśne).
4. Budowa modelu za pomocą algorytmu M5.Prime
5. Odczytanie, które z atrybutów mają znaczący wpływ na wartość (występują w węzłach drzewa).

Zbudowany za pomocą algorytmu M5.Prime model może służyć również do predykcji cen. Postępowanie prowadzące do określenia pierwszego przybliżenia wartości katastralnej na terenie danej jednostki taksacyjnej przebiega według następującego schematu:

1. Zebranie informacji o transakcjach.
2. Podział danych na podzbiory (grunty o przeznaczeniu dopuszczającym zabudowę oraz grunty o przeznaczeniu wykluczającym zabudowę).
3. Usunięcie z podzbiorów rekordów zawierających skrajne ceny.
4. Obliczenie trendu czasowego i przeliczenie cen transakcyjnych tak, by odnosiły się do jednej daty.
5. Kodowanie danych i zapis danych zawierających atrybutowy opis transakcji.
6. Wygenerowanie modeli.
7. Ustalenie listy atrybutów istotnych.
8. Zebranie informacji umożliwiających określenie wartości atrybutów istotnych dla wszystkich działek w jednostce taksacyjnej.
9. Zaklasyfikowanie (na podstawie przypisanych im wartości atrybutów istotnych) szacowanych działek do klas wygenerowanych przez model i nadanie im wartości przypisanych poszczególnym klasom.

Należy pamiętać, że algorytmy odkrywania wiedzy w bazach danych rozpoznają reguły i prawidłowości (ang. patterns) występujące w zbiorach danych. Wykorzystane w niniejszych badaniach drzewa decyzyjne potrafią uczyć się takich reguł, a „zdobyta” przez nie wiedza może być wykorzystana do klasyfikacji nowych danych (nie biorących udziału w procesie uczenia), a także do predykcji nieznanymi wartościami atrybutów na podstawie atrybutów znanych. Ważne jest, by zbiór uczący był reprezentatywny, to znaczy, by wystąpiły w nim możliwie wszystkie przypadki charakterystyczne dla badanego zjawiska. Ze względu na specyfikę rynku nieruchomości, przejawiającą się między innymi tym, że rynek ten niechętnie poddaje się prawdom statystyki, do wyników analiz nie wolno podchodzić bezkrytycznie. W procesie interpretacji wyników udział eksperta z dziedziny rynku nieruchomości wydaje się być konieczny.

WNIOSKI

1. Redukcja liczby atrybutów wartościotwórczych pozwala na uproszczenie metod powszechnej taksacji poprzez uniknięcie konieczności określania, dla nieruchomości stanowiących przedmiot masowej wyceny, wartości atrybutów nieistotnych, których pozyskanie nie odbywa się w sposób automatyczny i powoduje znaczne zwiększenie pracochłonności całego procesu masowej wyceny.

2. Drzewa decyzyjne mogą być z powodzeniem wykorzystywane do ustalania hierarchii atrybutów wartościotwórczych niezabudowanych nieruchomości gruntowych w celu pominięcia cech nieistotnych. Nie jest natomiast możliwa ilościowa ocena istotności poszczególnych atrybutów bezpośrednio na podstawie struktury drzewa i wartości w liściach.

3. Przejrzysta struktura drzew decyzyjnych pozwala na wyodrębnianie klas lokalizacyjnych, co może stanowić ułatwienie procesu wyznaczania stref taksacyjnych. W przypadku rynków charakteryzujących się niewielką liczbą transakcji mogą posłużyć do agregacji stref.

4. Lista atrybutów istotnych zbudowana na podstawie analizy danych na terenie obszaru badawczego pokrywa się w dużej mierze z listą ekspercką, reprezentującą prawidła na różnych rynkach lokalnych. Pozwala to na sformułowanie wniosku, że cechy wartościotwórcze, zaliczone w niniejszej pracy do istotnych, można uznać za istotne dla innych wiejskich i podmiejskich rynków nieruchomości. Należy jednak pamiętać, że fakt ten nie wyklucza pojawienia się innych atrybutów, których wpływu na wartość nie będzie można pominąć na danym rynku. Techniki KDD pozwalają na stwierdzenie ich istnienia na podstawie danych transakcyjnych reprezentatywnych dla analizowanego lokalnego rynku nieruchomości.

PIŚMIENNICTWO

Cichosz P., 2000. Systemy uczące się. WNT, Warszawa 2000.

Rozporządzenie Rady Ministrów z dnia 19 października 2001 r. w sprawie powszechnej taksacji nieruchomości (DzU nr 135, poz. 1514).

Ustawa z dnia 21 sierpnia 1997 r. o gospodarce nieruchomościami. (DzU nr 115, poz. 741), tekst jednolity DzU 46 poz. 543 z 2000 r.

Witten I. H., Frank E., 1999. Data mining. Morgan Kaufmann Publishers.

Witten I. H., Frank E., 2000. WEKA Machine Learning Algorithms in Java. Morgan Kaufmann Publishers.

Ustawa z dnia 27 marca 2003 r. o planowaniu i zagospodarowaniu przestrzennym (DzU nr 80, poz. 717).

APPLICATION OF DECISION TREES FOR THE ESTIMATION OF THE EFFECTS OF ATTRIBUTES OF UNBUILT PLOTS ON THEIR VALUE

Abstract: The paper presents the possibilities of the application of decision trees for the elimination of land plot attributes of low-importance, taken into account during mass appraisal. The influence of various real estate features has been discussed and a method for the elimination of attributes of low-importance has been presented.

Keywords: value-creating attributes, real-estate appraisal, mass appraisal, decision trees, artificial intelligence

Zaakceptowano do druku: 2004.04.14
Accepted for print