Gabriela Surowiec

Zastosowanie drzew decyzyjnych do analiz rynku nieruchomości w zakresie predykcji cen na przykładzie rynku lokali mieszkalnych w Krakowie

Acta Scientiarum Polonorum. Administratio Locorum 3/1, 101-111

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.

MUZEUM HISTORII POLSKI

ACTA SCIENTIARUM POLONORUM *Administratio Locorum 3 (1) 2004, 101–112*

ZASTOSOWANIE DRZEW DECYZYJNYCH DO ANALIZ RYNKU NIERUCHOMOŚCI W ZAKRESIE PREDYKCJI CEN NA PRZYKŁADZIE RYNKU LOKALI MIESZKALNYCH W KRAKOWIE

Gabriela Surowiec
Katedra Geodezji Wyższej, Akademia Rolnicza w Krakowie

Streszczenie. Artykuł przedstawia propozycję wykorzystania drzew decyzyjnych – popularnej techniki eksploracji danych – do analizy rynku nieruchomości na przykładzie rynku lokali mieszkalnych w Krakowie. Materiałem badawczym były dane transakcyjne i ofertowe jednostki ewidencyjnej Krowodrza. Celem badań było przedstawienie możliwości analitycznych, jakie dają drzewa decyzyjne w zakresie predykcji wartości nieruchomości.

Słowa kluczowe: drzewa decyzyjne, analiza rynku nieruchomości, wartość nieruchomości, eksploracja danych, KDD

WSTĘP

Tam, gdzie istnieją duże zbiory danych opisujących pewne zjawiska za pomocą bardzo wielu rozmaicie skorelowanych ze sobą elementarnych informacji, niezmiernie trudno ustalić reguły rządzące tymi zjawiskami. Zadanie jest jeszcze trudniejsze, gdy zjawiska te zmieniają się w czasie.

Rynek nieruchomości jest przykładem takiego właśnie miejsca – poligonem, na którym warto wypróbować metody umożliwiające dotarcie do niewidocznych gołym okiem reguł, tzw. metody odkrywania wiedzy w bazach danych (ang. knowledge discovery in databases), znane też pod nazwą eksploracji danych (ang. data mining).

Adres do korespondencji – Corresponding author: mgr inż. Gabriela Surowiec, Akademia Rolnicza w Krakowie, Katedra Geodezji Wyższej, ul. Balicka 253a, 30-198 Kraków, e-mail: rmkoster@cyf-kr.edu.pl

Metody te rozwinęły się wraz z techniką komputerową, ponieważ do ich realizacji, ze względu na złożoność algorytmów i długi czas obliczeń, potrzeba dużej mocy obliczeniowej.

EKSPLORACJA DANYCH

Przez eksplorację danych rozumiemy proces automatycznego odkrywania znaczących, pożytecznych, dotychczas nieznanych i wyczerpujących informacji w dużych bazach danych, informacji ujawniających ukrytą wiedzę o badanym przedmiocie. Wiedza ta przyjmuje postać reguł, prawidłowości, tendencji, korelacji i jest następnie przedstawiana przygotowanemu do jej spożytkowania użytkownikowi w celu rozwiązania stojących przed nim problemów i podjęcia istotnych decyzji [Muraszkiewicz 2003].

Podstawowym celem eksploracji danych jest sięgnięcie możliwie najgłębiej do dostępnych zasobów informacyjnych po to, aby poznać regularności i prawidłowości istniejące w świecie reprezentowanym przez te zasoby i aby skutecznie prognozować.

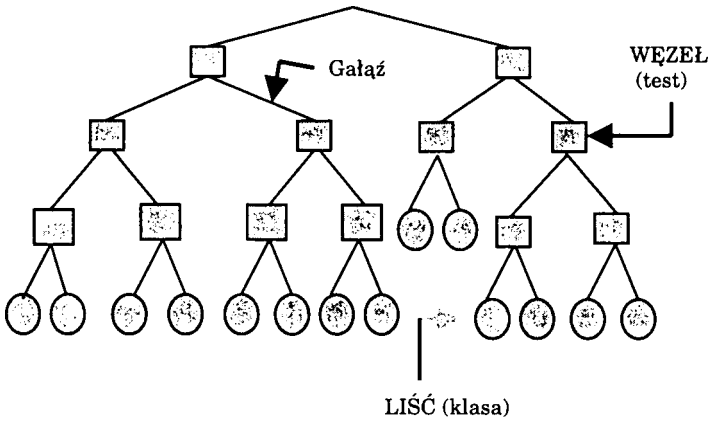
Praktyczne pożytki drążenia danych ujawniają się w dwóch dziedzinach, którymi są: prognozowanie (ang. prediction, forecasting) i opis (ang. description). Prognozowanie polega na wykorzystaniu znanych wartości interesujących nas zmiennych (atrybutów) w celu przewidywania wartości tych lub innych zmiennych w przyszłości.

Opis polega na tworzeniu czytelnej i zrozumiałej dla człowieka reprezentacji wiedzy wydobytej z danych w postaci wykresów, wzorów, reguł, tabel [Muraszkiewicz 2003].

Klasyfikacja jest procesem uczenia się, którego celem jest określenie reguły, służącej do przyporządkowania (zaklasyfikowania), jeśli już zostanie zaakceptowana, branego pod uwagę elementu do jednej lub więcej wcześniej zdefiniowanych klas (zbiorów). Proces ten korzysta ze zbioru wcześniej poklasyfikowanych przykładów, po to, by określić sposób (model) klasyfikowania całej dostępnej populacji elementów [Muraszkiewicz 2003]. Klasyfikacja jest szczególnie przydatna do prognozowania.

DRZEWA DECYZYJNE

Do klasyfikacji często wykorzystuje się algorytmy działające na zasadzie drzew decyzyjnych. Drzewo decyzyjne to struktura złożona z węzłów, z których wychodzą gałęzie prowadzące do innych węzłów lub liści, oraz z liści, z których nie wychodzą żadne gałęzie (rys. 1.). Węzły odpowiadają testom przeprowadzanym na wartościach atrybutów przykładów, gałęzie odpowiadają możliwym wynikom tych testów, liście zaś etykietom kategorii (klasom). Klasyfikacja za pomocą drzewa decyzyjnego polega na przejściu ścieżki od ko-



Rys. 1. Drzewo decyzyjne – schemat ideowy
Fig. 1. Decision tree – conceptual scheme

rzenia do liścia drzewa wzdłuż gałęzi wyznaczanych przez wyniki testów związanych z odwiedzanymi kolejno węzłami. Osiągnięcie liścia wyznacza kategorię (klasę) [Cichosz 2000].

W kolejnych węzłach drzewa, na podstawie testów wartości kolejnych, wybranych przez algorytm atrybutów, następuje podział danych na podzbiory (gałęzie). Podział odbywa się w taki sposób, by osiągnąć maksymalny zysk informacji, czyli możliwie obniżyć różnorodność. Oznacza to, że elementy zaklasyfikowane do jednego podzbioru charakteryzują się jak największym wzajemnym podobieństwem, zaś elementy przypisane do różnych podzbiorów są wzajemnie jak najbardziej zróżnicowane. Kryteria poprawności podziału zależą od algorytmu.

Za pomocą drzewa decyzyjnego można obliczyć nieznane wartości atrybutów na podstawie przykładów pochodzących ze zbioru danych, tzw. zbioru uczącego.

Trzeba podkreślić, że w procesie odkrywania wiedzy niezwykle istotną rolę odgrywa człowiek, analityk problemu, którego umiejętności, doświadczenie i praca mają główne znaczenie w otrzymaniu znaczących rezultatów. Jego rola polega na stałej krytycznej ocenie każdego kroku w procesie odkrywania, swoistym „cenzurowaniu" otrzymywanych rezultatów cząstkowych i sterowaniu całym procesem [Muraszkiewicz 2003].

DANE OPISUJĄCE RYNEK NIERUCHOMOŚCI

W świetle możliwości analitycznych jakie dają techniki eksploracji danych oraz niewyczerpanych zasobów informacji dostępnych w Internecie warto przyjrzeć się źródłom danych z nieco innej perspektywy. Podstawowym

źródłem danych opisujących zjawiska na rynku nieruchomości są dane transakcyjne zawarte w umowach kupna-sprzedaży nieruchomości. W formie nie przetworzonej i nie usystematyzowanej są niestety mało atrakcyjnym materiałem do analizy.

Drugim, w praktyce mniej docenianym rodzajem danych, są dane ofertowe, które wprawdzie nieużyteczne, jeśli brać pod uwagę obwarowaną rozporządzeniem i standardami zawodowymi sztukę wyceny nieruchomości, mogą okazać się źródłem bezcennych informacji, jeśli potraktujemy je jako dane opisujące trendy i zjawiska, a zatem służące do rozpoznania i zrozumienia zachowań rynku, a nie predykcji wartości pojedynczych elementów tego rynku.

Oprócz prasy branżowej, bogatym źródłem nienajgorzej usystematyzowanych danych ofertowych jest Internet. W Polsce działa kilka dobrych systemów, w których są publikowane ogłoszenia dotyczące sprzedaży nieruchomości. Liczba ofert w każdej kategorii sięga od kilku do kilkunastu tysięcy, a zatem są to na tyle duże zbiory danych, że uzasadniona jest ich eksploracja, w celu poznania pewnych ukrytych w nich reguł i prawidłowości.

W artykule przedstawiono, jak wykorzystać te ogromne zasoby danych i jak na ich podstawie rozpoznawać i interpretować zjawiska zachodzące na rynku nieruchomości.

MATERIAŁ I METODY

Algorytmy odkrywania wiedzy można m.in. stosować do rozwiązywania następujących problemów związanych z analizą rynku nieruchomości:
- oceny istotności cech wartościotwórczych,
- wyodrębnianie stref cenowych,
- predykcji cen – przybliżonego określania wartości (metody statystyczne).

Możliwości analityczne, jakie dają drzewa decyzyjne, przedstawiono na przykładzie rynku mieszkaniowego w Krakowie.

MATERIAŁ

Analizie poddano dwa zbiory: zbiór transakcji dotyczących nieruchomości lokalowych oraz własnościowych spółdzielczych praw do lokali mieszkalnych (315 transakcji zanotowanych przez kolejne 9 miesięcy), zanotowanych w dzielnicy Krowodrza oraz zbiór danych ofertowych w tym samym zakresie, pochodzących z tego samego obszaru, publikowanych w internetowym serwisie www.nieruchomosci.onet.pl (845 ofert zanotowanych w ciągu kolejnych 30 dni).

Transakcje i oferty opisano za pomocą następujących atrybutów:

Dane transakcyjne: data transakcji, obręb ewidencyjny, forma władania, powierzchnia użytkowa, kondygnacja, wiek, stan techniczny, liczba pokoi, technologia.

Dane ofertowe: obręb ewidencyjny, forma władania, powierzchnia użytkowa, kondygnacja, stan techniczny, liczba pokoi, technologia.

Metody. Jako technikę eksploracji danych zastosowano klasyfikację, a do analizy posłużono się algorytmem M5'. Obliczenia wykonano za pomocą programu WEKA [Witten, Frank 1999]. W wyniku analiz, zarówno dla danych transakcyjnych, jak i opisowych, zbudowano modele predykcyjne. Algorytm M5' 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. Podzbiory uzyskane na końcach gałęzi w wyniku cięcia drzewa noszą nazwę liści lub klas. Obliczenie wartości atrybutu, ze względu na który jest wykonywana 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), bądź na podstawie równania regresji liniowej (model tree) [Witten, Frank 2000]. Mając na uwadze cel badań i przejrzystość wyników, analizy wykonano, korzystając z modelu regression tree.

Algorytm drzewa decyzyjnego dzieli kolejno zbiór danych wejściowych na podzbiory (w przypadku M5' na każdym kolejnym poziomie są to dwa podzbiory), wybierając kolejno atrybuty, które najbardziej różnicują cenę. Oznacza to, że mieszkania zaklasyfikowane do jednego podzbioru są wzajemnie możliwie jak najbardziej do siebie podobne oraz jak najbardziej różnią się od tych, które zaliczono do drugiego podzbioru. W każdym węźle podział następuje na podstawie wartości jednego z atrybutów – tego, który daje największy zysk informacji.

WYNIKI

Ocena istotności cech wartościotwórczych

Najprostszym sposobem ustalenia hierarchii atrybutów za pomocą drzewa decyzyjnego jest odczytanie, które cechy występują w górnych węzłach drzewa. Przejawem istotności atrybutu jest jego wystąpienie w węzłach drzewa [Malczewska 2003]

W analizowanych przykładach są to odpowiednio:

- dla zbioru danych transakcyjnych – lokalizacja (wyrażona numerem obrębu ewidencyjnego), stan techniczny i wiek budynku, powierzchnia;

```
Obręb = 3006, 3028, 3011, 3003, 3034, 3002, 3014, 3005, 3007, 3013, 3009, 3010 - fałsz:
| Stan = rowe, dobry, bardzo dobry - fałsz:
| | Obręb = 3042, 3012, 3029, 3004, 3045, 3046 - fałsz: LM1 (44 / 2350 zł)
| | Obręb = 3042, 3012, 3029, 3004, 3045, 3046 - prawda: LM2 (64 / 2630 zł)
| Stan = rowe, dobry, bardzo dobry - prawda:
| | Stan = dobry, bardzo dobry - fałsz:
| | | Pow <= 46.8: LM3 (33 / 2800 zł)
| | | Pow > 46.8: LM4 (35 / 2570 zł)
| | Stan = dobry, bardzo dobry - prawda:
| | | Pow <= 51.8: LM5 (11 / 2860 zł)
| | | Pow > 51.8:
| | | | Pow <= 64.5: LM6 (3 / 3940 zł)
| | | | Pow > 64.5: LM7 (11 / 2900 zł)
Obręb = 3006, 3028, 3011, 3003, 3034, 3002, 3014, 3005, 3007, 3013, 3009, 3010 - prawda:
| Wiek = lata1995-2004, lata 1985-1994 - fałsz: LM8 (58 / 2650 zł)
| Wiek = lata1995-2004, lata 1985-1994 - prawda:
| | Obręb = 3014, 3005, 3007, 3013, 3009, 3010 - fałsz: LM9 (43 / 3390 zł)
| | Obręb = 3014, 3005, 3007, 3013, 3009, 3010 - prawda: LM10 (13 / 4340 zł)
```

Rys. 2. Drzewo decyzyjne (zbiór danych transakcyjnych)
Fig. 2. Decision tree (sale prices data set)

– dla zbioru danych ofertowych – stan techniczny, lokalizacja (wyrażona numerem obrębu ewidencyjnego), powierzchnia mieszkania i kondygnacja;

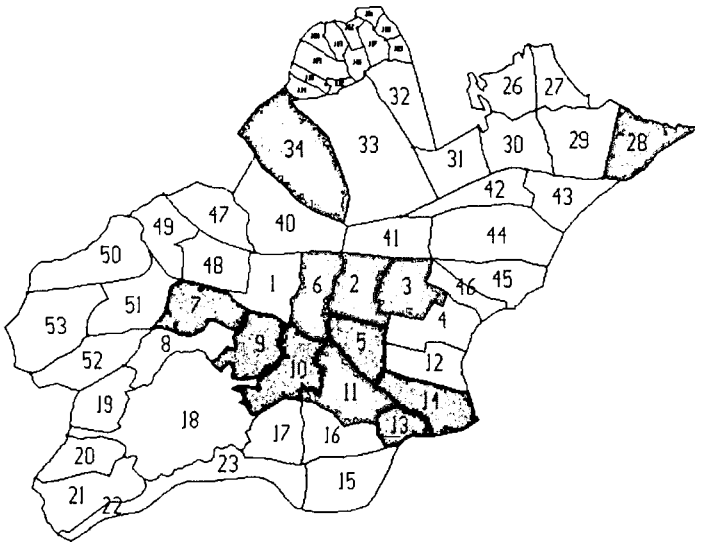
```
Stan = nowe, bardzo dobry - fałsz:
| Obręb = 3005, 3012, 3046, 3031, 3011, 3001, 3004, 3002, 3014, 3013, 3009, 3007, 3045, 3010,
3008, 3003, 3006 - fałsz:
| | Stan = dobry - fałsz: LM1 (160 / 3070 zł)
| | Stan = dobry - prawda:
| | | Kondyg <= 3.5: LM2 (36 / 3630 zł)
| | | Kondyg > 3.5: LM3 (30 / 3160 zł)
| Obręb = 3005, 3012, 3046, 3031, 3011, 3001, 3004, 3002, 3014, 3013, 3009, 3007, 3045, 3010,
3008, 3003, 3006 - prawda:
| | Stan = do indywidualnego wykończenia, dobry - fałsz: LM4 (257 / 3540 zł)
| | Stan = do indywidualnego wykończenia, dobry - prawda: LM5 (170 / 3840 zł)
Stan = nowe, bardzo dobry - prawda:
| Obręb = 3004, 3002, 3014, 3013, 3009, 3007, 3045, 3010, 3008, 3003, 3006 - fałsz:
| | Pow <= 73.5:
| | | Stan = bardzo dobry - fałsz: LM6 (25 / 3700 zł)
| | | Stan = bardzo dobry - prawda: LM7 (30 / 4190 zł)
| | Pow > 73.5: LM8 (40 / 3520 zł)
| Obręb = 3004, 3002, 3014, 3013, 3009, 3007, 3045, 3010, 3008, 3003, 3006 prawda:
| | Stan = bardzo dobry - fałsz: LM9 (32 / 4330 zł)
| | Stan = bardzo dobry - prawda:
| | | Kondyg <= 2.5:
| | | | Pow <= 94: LM10 (16 / 4850 zł)
| | | | Pow > 94: LM11 (8 / 6270 zł)
| | | Kondyg > 2.5: LM12 (41 / 4530 zł)
```

Rys. 3. Drzewo decyzyjne (zbiór danych ofertowych)
Fig. 3. Decision tree (advertisement prices data set)

Wyodrębnianie stref cenowych

Łączenie obszarów w strefy jest uzasadnione szczególnie tam, gdzie liczba transakcji jest zbyt mała, by wykonywać obliczenia odrębnie dla każdego z obrębów. Oto ilustracja propozycji podziału dzielnicy Krowodrza na strefy wykonana za pomocą drzewa decyzyjnego.

Analiza tych samych dwóch zbiorów – danych ofertowych i transakcyjnych w Krowodrzy, umożliwiła wyodrębnienie stref cenowych, których przestrzenne rozmieszczenie przedstawiają rysunki 4 i 5.

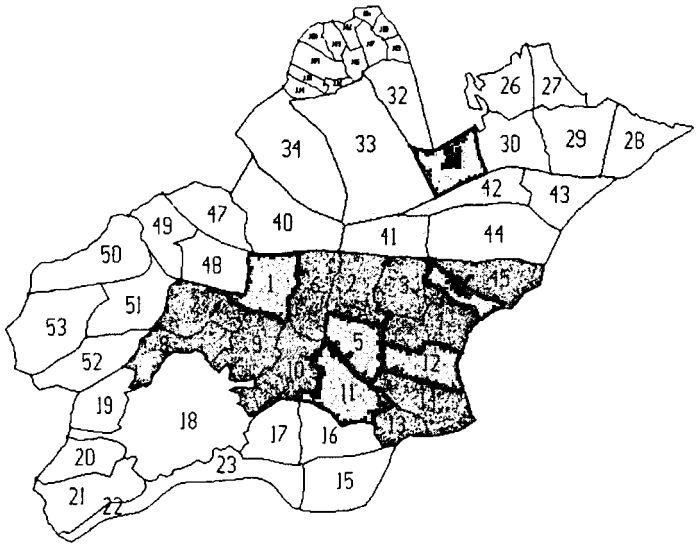


Rys. 4. Strefy lokalizacyjne mieszkań w Krowodrzy na podstawie danych transakcyjnych
Fig. 4. Location zones of residential property in the Krowodrza district according to sale prices

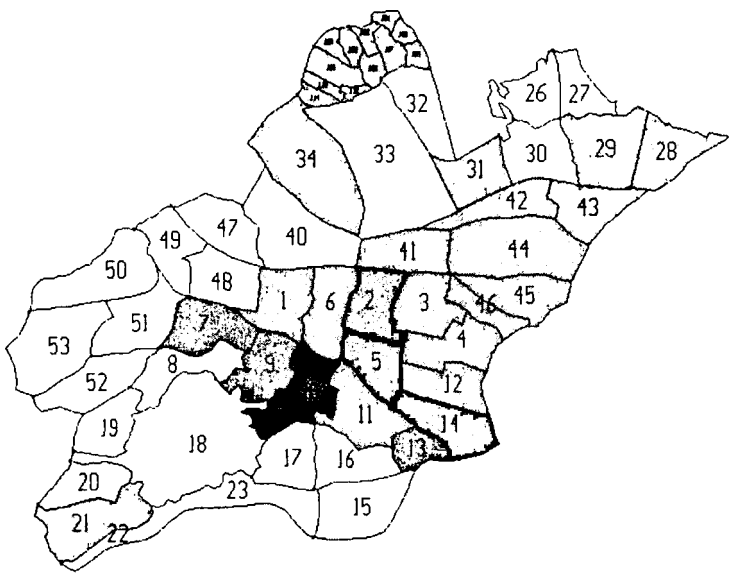
W tym miejscu Czytelnik może zadać pytanie, czym różni się wyodrębnianie stref przy zastosowaniu drzew decyzyjnych od zwykłego podziału obszaru na strefy, bezpośrednio na podstawie cen. Rysunek 6 przedstawia strefy lokalizacyjne wyznaczone wprost na podstawie cen transakcyjnych.

Wyodrębnienie stref bezpośrednio na podstawie transakcji (np. średnich cen notowanych w poszczególnych obrębach) oznacza, że jedynym czynnikiem branym pod uwagę jest cena. Wyodrębnianie stref powinno obrazować wpływ lokalizacji na cenę, ale czy powinno odbywać się jedynie na podstawie ceny?

Wyobraźmy sobie sytuację, kiedy w danym obrębie ewidencyjnym zanotowano transakcje dotyczące wyłącznie mieszkań w złym stanie technicznym w budynkach do remontu. Średnia cena takich transakcji będzie zatem stosunkowo niska, nawet gdy pod względem lokalizacyjnym miejsce jest atrakcyjne. Gdyby w tej lokalizacji zaistniała podaż mieszkań nowych lub wyremontowanych, cena byłaby prawdopodobnie dużo wyższa. Drzewo decyzyjne



Rys. 5. Strefy cenowe mieszkań w Krowodrzy na podstawie danych ofertowych
Rys. 5. Location zones of residential property in the Krowodrza district according to sale advertisements



Rys. 6. Strefy cenowe wyznaczone na podstawie średnich cen w obrębach, obliczonych na podstawie danych transakcyjnych
Fig. 6. Location zones based on average sale prices

umożliwia uwzględnienie takich właśnie relacji i poprawne rozpoznanie najlepszych lokalizacji na danym obszarze badań. Potwierdzają to rezultaty badań zilustrowane na rysunkach 4 (na podstawie wyników uzyskanych za pomocą drzewa decyzyjnego) oraz 6 (na podstawie cen średnich w poszczególnych obrębach).

Predykcja cen – przybliżone określanie wartości (metody statystyczne)

Wartość generowana w liściu drzewa w zależności od wybranego modelu jest średnią arytmetyczną wartości atrybutu, ze względu na który jest wykonywana analiza przykładów (rekordów) zaliczonych do liścia (regression tree) lub jest wynikiem równania regresji liniowej wygenerowanego dla tych przykładów (model tree). Zbudowany w ten sposób model pozwala na określenie wartości analizowanego atrybutu (w tym wypadku ceny) dla nowych, nieznanych przykładów, które nie były wykorzystane przy jego budowie. W jednym i drugim przypadku predykowana cena (czyli wartość) nieruchomości jest wynikiem zastosowania metod statystycznych. Jak już wspomniano, każdemu liściowi (klasie) przypisana jest wartość (w przypadku regression tree – średnia arytmetyczna cen przykładów pochodzących ze zbioru uczącego, zaliczonych do danej klasy). Zaklasyfikowanie nowego przykładu jest równoznaczne z przypisaniem mu predykowanej wartości. Liście są oznaczone kolejno L1, L2, L3, ... (rys. 7).

```
Obręb = 3006, 3028, 3011, 3003, 3034, 3002, 3014, 3005, 3007, 3013, 3009, 3010 - fałsz:
| Stan = nowe, dobry, bardzo dobry - fałsz:
| | Obręb = 3042, 3012, 3029, 3004, 3045, 3046 - fałsz: LM1 (44 / 2350 zł)
| | Obręb = 3042, 3012, 3029, 3004, 3045, 3046 - prawda: LM2 (64 / 2630 zł)
| Stan = nowe, dobry, bardzo dobry - prawda:
| | Stan = dobry, bardzo dobry - fałsz:
| | | Pow <= 46.8: LM3 (33 / 2800 zł)
| | | Pow > 46.8: LM4 (35 / 2570 zł)
| | Stan = dobry, bardzo dobry - prawda:
| | | Pow <= 51.8: LM5 (11 / 2860 zł)
| | | Pow > 51.8:
| | | | Pow <= 64.5: LM6 (3 / 3940 zł)
| | | | Pow > 64.5: LM7 (11 / 2900 zł)
Obręb = 3006, 3028, 3011, 3003, 3034, 3002, 3014, 3005, 3007, 3013, 3009, 3010 - prawda:
| Wiek = lata1995-2004, lata 1985-1994 - fałsz: LM8 (58 / 2650 zł)
| Wiek = lata1995-2004, lata 1985-1994 - prawda:
| | Obręb = 3014, 3005, 3007, 3013, 3009, 3010 - fałsz: LM9 (43 / 3390 zł)
| | Obręb = 3014, 3005, 3007, 3013, 3009, 3010 - prawda: LM10 (13 / 4340 zł)
```

Rys. 7. Drzewo decyzyjne (zbiór danych transakcyjnych)
Fig. 7. Decision tree (sale prices data set)

Mieszkaniom w bardzo dobrym i dobrym stanie technicznym, o powierzchni mniejszej od 51,8 m^2, położonym w obrębach 42, 12, 29, 4, 45, 46 przypisana jest wartość 2860 zł/m^2 (L5), natomiast takim samym lokalom o powierzchni większej niż 64,5 m^2 – 2900 zł/m^2 (L7). Mieszkania zlokalizowane w obrębach 6, 28, 11, 3, 34, 2, 14, 5, 7, 13, 9 i 10, wybudowane przed 1985 rokiem, mają cenę 2650 zł/m^2.

DYSKUSJA

W niniejszym artykule pokazano, w jaki sposób drzewa decyzyjne mogą być pomocne w rozpoznawaniu i rozumieniu zjawisk zachodzących na rynku nieruchomości. O ile spośród przedstawionych tu obszarów ich zastosowania dwa pierwsze (ocena istotności cech wartościotwórczych oraz wyodrębnianie stref cenowych) wydają się nie budzić kontrowersji, o tyle trzeci pozostaje niezbadany.

W świetle Standardu III [Standardy Zawodowe... 2003], dopuszczającego metody statystyczne do obliczania wartości rynkowej nieruchomości (metoda analizy statystycznej rynku), wydaje się, że wartość obliczona za pomocą drzewa decyzyjnego może stanowić przybliżoną wartość nieruchomości.

Dotychczasowe próby stosowania drzew decyzyjnych (algorytmu M5') do określania wartości nieruchomości, dotyczące analizy rynku nieruchomości gruntowych [Malczewska 2003], mimo że nie obalają, to jednak nie potwierdzają w sposób przekonujący tej hipotezy.

Podobnie wyniki badań przedstawione w niniejszym artykule, choć w żaden sposób nie obalają hipotezy o możliwości stosowania drzew decyzyjnych do obliczania wartości nieruchomości, nie pozwalają na wyciągnięcie tak daleko idących wniosków. Będzie to możliwe dopiero poprzez porównanie dokładności predykcji wykonanej za pomocą drzew decyzyjnych z dokładnością oszacowania wartości innymi metodami statystycznymi na różnych, pod względem obszarowym i przedmiotowym, rynkach nieruchomości.

WNIOSKI

1. Drzewa decyzyjne, na pozór nieco skomplikowane, są strukturą bardzo prostą i przejrzystą. Skomplikowane są jedynie algorytmy, ale ich budową zajmują się matematycy i informatycy.

2. Rozwiązanie problemu nieczytelnej formy zobrazowania struktury drzewa w programie WEKA ogranicza się do zaprojektowania bardziej przyjaznego interfejsu.

3. Drzewa decyzyjne można stosować do oceny istotności cech wartościotwórczych oraz wyznaczania stref, natomiast możliwości ich zastosowania do

predykcji wartości nie są jeszcze dostatecznie potwierdzone. Uzyskane wyniki dowodzą, że nie ma powodu do odrzucenia drzew decyzyjnych jako metody obliczania przybliżonej wartości nieruchomości, natomiast ze względu na fakt, że badania prowadzone są od niedawna i jedynie na wybranych, ograniczonych obszarowo i przedmiotowo rynkach, należy odnosić się do nich z rezerwą.

4. Celem publikacji jest przedstawienie możliwości, jakie dają drzewa decyzyjne jako narzędzie analityczne i zainteresowanie nimi badaczy rynku nieruchomości. Poznanie i zweryfikowanie wszystkich możliwości zastosowania drzew decyzyjnych do analiz rynku nieruchomości wymaga bowiem jeszcze wielu badań.

PIŚMIENNICTWO

Muraszkiewicz M., 2003. Eksploracja danych dla telekomunikacji. Instytut Informatyki Teoretycznej i Stosowanej PAN. http://www.icie.com.pl/ARCHIVE/dm_tel.doc

Witten I, Frank E., 1999. Data Mining. Morgan Kaufmann, Los Altos/Palo Alto/San Francisco.

Malczewska A., 2003. Ocena wpływu cech niezabudowanych nieruchomości gruntowych na ich wartość w świetle powszechnej taksacji. Rozprawa doktorska. Maszynopis. Olsztyn.

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

Standardy Zawodowe Rzeczoznawców Majątkowych. 2003. Polska Federacja Stowarzyszeń Rzeczoznawców Majątkowych. Warszawa.

APPLICATION OF DECISION TREES TO REAL ESTATE MARKET ANALYSES AND PRICE PREDICTION - THE EXAMPLE OF RESIDENTIAL REAL PROPERTY MARKET IN CRACOW

Abstract: The paper presents a proposal for the application of decision trees – a popular technique of data exploration - for real estate market analyses. The study area was the Krowodrza district, where the market of residential real property (sale prices and advertisements) had been analyzed. In particular the aim of the research was to show analytical possibilities of decision tree applications for the prediction of the real estate value.

Keywords: decision trees, real estate market analyses, real property value, data mining, knowledge discovery in databases

Zaakceptowano do druku: 2004.04.14
Accepted for print