

Andrzej Muczyński

Podział zasobu nieruchomości wspólnot mieszkaniowych z wykorzystaniem drzewa klasyfikacyjnego

Acta Scientiarum Polonorum. Administratio Locorum 9/4, 73-84

2010

Artykuł został 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.

PODZIAŁ ZASOBU NIERUCHOMOŚCI WSPÓLNOT MIESZKANIOWYCH Z WYKORZYSTANIEM DRZEWA KLASYFIKACYJNEGO

Andrzej Muczyński

Uniwersytet Warmińsko-Mazurski w Olsztynie

Streszczenie. W pracy przedstawiono wyniki badań dotyczące podziału zasobu nieruchomości wspólnot mieszkaniowych z wykorzystaniem metodyki tworzenia drzew klasyfikacyjnych. Przedmiot badań stanowi populacja nieruchomości wspólnot mieszkaniowych z udziałem miasta Olsztyna. Procedury badawcze zostały tak ukierunkowane, aby wyniki klasyfikacji uwzględniały kryteria istotne z punktu widzenia racjonalnej gospodarki nieruchomościami komunalnymi. Uzyskane drzewo klasyfikacyjne reprezentuje hierarchiczny układ maksymalnie jednorodnych podzbiorów obiektów z minimalną liczbą węzłów generujących zbiór prostych reguł klasyfikacji. Z analiz wynika, że w badanej populacji zdecydowanie dominują duże wspólnoty mieszkaniowe, co sprzyja poprawie efektywności gospodarowania. Stwierdzono ponadto, że w przeważającej liczbie badanych wspólnot gmina utraciła większość udziałów wskutek rozproszonej prywatyzacji mieszkań komunalnych. Stan ten znacznie komplikuje procesy decyzyjne w obszarze zarządzania nieruchomościami komunalnymi.

Słowa kluczowe: podział, wspólnoty mieszkaniowe, drzewa klasyfikacyjne, gospodarka nieruchomościami komunalnymi

WPROWADZENIE

Podział danego zbioru obiektów czy zjawisk na w miarę jednorodne podzbiory z punktu widzenia określonego kryterium opartego na cechach analizowanych przedmiotów nazywa się **klasyfikacją**. Dzięki redukcji entropii i możliwości formułowania wniosków uogólniających proces ten odwzorowuje podstawowe czynności poznawcze wykonywane przez człowieka. Do rozwiązywania zadań klasyfikacji wykorzystywane są metody taksonomiczne, taksonomia jest bowiem dyscypliną naukową zajmującą się problemami podziału, porządkowania, grupowania lub dyskryminacji. W ramach tej dyscypliny wyróżnić można **taksonomię numeryczną**

Adres do korespondencji – Corresponding author: Andrzej Muczyński, Katedra Gospodarki Nieruchomościami i Rozwoju Regionalnego, Uniwersytet Warmińsko-Mazurski w Olsztynie, 10-724 Olsztyn, ul. Romana Prawocheńskiego 15, e-mail: amucz@uwm.edu.pl

(ilościową) i **symboliczną** (jakościową). Klasyczna taksonomia numeryczna (taksonometria) posługuje się obecnie prawie wyłącznie metodami ilościowymi, w których cechy obiektów mierzone są na skalach mocnych, a miary podobieństwa oparte na odległości wyrażonej z reguły funkcjami rzeczywistymi. Z kolei taksonomia symboliczna obejmuje grupę metod klasyfikacji charakteryzujących się tym, że ich przedmiotem są przede wszystkim **obiekty symboliczne**, których cechy mają charakter jakościowy i są reprezentowane przez zmienne mierzone w skalach słabych: nominalnej i porządkowej. Podejście symboliczne pozwala po prostu rozszerzyć przedmiot zainteresowań taksonomii na obiekty o złożonej strukturze, do których zaliczyć należy nieruchomości wspólnot mieszkaniowych. W zbiorze metod taksonomii symbolicznej jedną z podstawowych metod klasyfikacji takich obiektów jest metoda oparta na budowaniu drzew klasyfikacyjnych (decyzyjnych).

Celem pracy jest dokonanie podziału zasobu nieruchomości wspólnot mieszkaniowych z wykorzystaniem metodyki tworzenia drzew klasyfikacyjnych. Opracowanie zasad hierarchicznego podziału badanego zbioru może być wykorzystane do wspomagania decyzji w procesach gospodarowania zasobami gminnymi, w szczególności w zakresie polityki prywatyzacyjnej i inwestycyjnej. Przedmiot badań stanowi populacja nieruchomości wspólnot mieszkaniowych z udziałem miasta Olsztyna. Procedury badawcze zostały tak ukierunkowane, aby wyniki klasyfikacji uwzględniały kryteria istotne z punktu widzenia racjonalnej gospodarki nieruchomościami komunalnymi. Uzyskane drzewo winno reprezentować hierarchiczny układ maksymalnie jednorodnych podzbiorów obiektów – z punktu widzenia przyjętych kryteriów – przy minimalnej liczbie węzłów decyzyjnych prowadzących do zbioru możliwie prostych reguł klasyfikacji.

DRZEWY KLASYFIKACYJNE JAKO METODA TAKSONOMII SYMBOLICZNEJ

Taksonomia symboliczna dostarcza komplementarnych metod klasyfikacji w stosunku do metod klasycznej taksonometrii będących ich rozszerzeniem i uzupełnieniem. Jej metody są stosowane wtedy, gdy charakter danych statystycznych nie pozwala na wykorzystanie metod numerycznych. Podobieństwo pary obiektów w tym ujęciu uwzględnia własności pozostałych obiektów w zbiorze (tzn. wykorzystuje miarę kontekstową). Ważną cechą metod symbolicznych jest generowanie informacji w postaci deklaratywnej, a powstałe w efekcie podziału skupienia są łatwiejsze do interpretacji dzięki zastosowaniu do ich opisu wyrażań koniunkcyjnych opartych na wartościach cech należących do nich obiektów. Klasy obiektów uzyskane metodami symbolicznymi mają charakter naturalny, gdyż większość z nich opiera się na wynikach badań nad kategoryzacją dokonywaną przez ludzi. Metody symboliczne są poza tym mniej wymagające od metod numerycznych, dopuszczają bowiem nieznaną część niektórych cech lub brak precyzji ich pomiaru, jak również sekwencyjne włączanie obiektów do struktury klas. W przeciwieństwie do taksonomii numerycznej, której główne zadanie polega na znalezieniu struktury klas, w ramach taksonomii symbolicznej realizowane są jednocześnie dwie grupy czynności [Gatnar 1998]:

podział na skupienia – wyodrębnienie jednorodnych grup obiektów (klas), i znalezienie charakterystyki wydzielonych klas – opis klas za pomocą odpowiednich pojęć (zbiorów reguł przynależności obiektów do wyróżnionych klas).

Podstawy teoretyczne metodyki symbolicznego opisu obiektów opracowano na początku lat osiemdziesiątych ubiegłego stulecia pod nazwą *konceptualnej analizy skupień* [Michalski 1980]. Metodyka ta pozwala otrzymać zbiór skupień, z których każde opisane jest przez koniunkcję wartości cech należących do nich obiektów. Uzyskany zbiór reguł przynależności obiektów do klasy nazywany jest *pojęciem*. Natomiast obiekty i klasy opisuje się za pomocą wyrażeń koniunkcyjnych nazywanych **kompleksami**. Na przykład kompleks:

$$[\text{typ nieruchomości} = \text{lokalowa}] \wedge [\text{funkcja} = \text{mieszkalna}] \wedge [\text{piętro} = 2] \wedge [\text{powierzchnia} = 52]$$

reprezentuje nieruchomość lokalową o funkcji mieszkalnej położoną na drugim piętrze o powierzchni 52 m². W takim ujęciu nie muszą być znane wartości wszystkich cech obiektu jak w przypadku taksonometrii, w której dane mają postać macierzy. Dodatkową zaletą tego podejścia jest stosowanie podobieństwa bardziej zrozumiałego przez ludzi.

Aktualnie znanych jest wiele algorytmów taksonomii symbolicznej w takich dziedzinach, jak: sztuczna inteligencja, statystyka czy też teoria rozpoznawania obrazów. Zostały one stworzone przez analogię do swoich odpowiedników opracowanych na gruncie taksonomii numerycznej. Algorytmy te dzielą się na trzy grupy [Gatnar 1998]:

- a) **hierarchiczne** – tworzące drzewo klas, w którym liście reprezentują poszczególne obiekty, a węzły ich zbiory; klasy na wyższym poziomie drzewa zawierają w sobie zbiory obiektów z poziomu niższego (np. algorytmy: EPAM, COBWEB, CLASSIT, RUMMAGE itp.),
- b) **optymalizacyjno-iteracyjne** – tworzące strukturę k klas, która jest optymalna ze względu na pewne kryterium jakości podziału (np. algorytmy: CLUSTER, WITT),
- c) **tworzące skupienia nierozłączne** – w których klasy mają części wspólne, tzn. ten sam obiekt może należeć do więcej niż jednej kategorii (np. algorytmy: UNIMEM i HIERARCH).

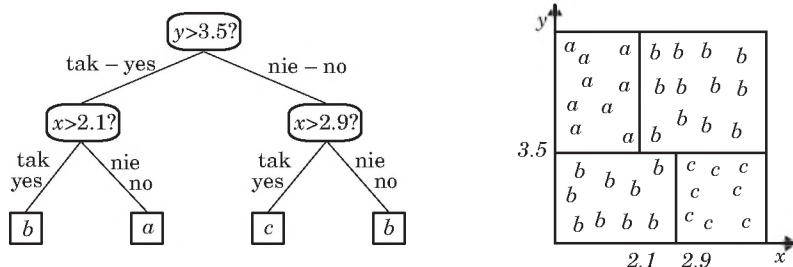
Większość metod taksonomii symbolicznej należy do metod hierarchicznych związanych z tworzeniem drzew klas, w których konstruowanie reguł klasyfikacji obiektów symbolicznych polega na dzieleniu w kolejnych krokach zbioru uczącego na podzbiory tak długo, aż zostanie osiągnięta ich jednorodność ze względu na przynależność do klas.

Pierwsze pomysły wykorzystania konstrukcji typu drzewo do reprezentowania procesu tworzenia *pojęć* (reguł przynależności obiektów do klas) powstały w latach sześćdziesiątych ubiegłego stulecia. Wówczas stworzono pierwszy algorytm budowy drzew klasyfikacyjnych (decyzyjnych) o nazwie „Concept Learning System” (CLS), w którym za *pojęcie* przyjęto regułę decyzyjną, która zastosowana do charakterystyki obiektu mówi o tym, czy należy on do klasy o określonej nazwie [Hunt i in. 1966]. Kluczowym momentem w rozwoju tego typu metod było stworzenie algorytmu ID3 („Induction of Decision Trees”), którego udane zastosowania praktyczne

zwróciły uwagę na drzewa klasyfikacyjne jako na wygodne narzędzie klasyfikacji danych [Quinlan 1986].

Idea drzewa klasyfikacyjnego polega na stopniowym podziale zbioru obiektów na podzbiory tak długo, aż zostanie osiągnięta ich jednorodność ze względu na przynależność do klas. Wygląd drzewa jest zbliżony do grafu. Drzewo klasyfikacyjne składa się z **korzenia**, z którego wychodzą co najmniej dwie krawędzie do węzłów leżących na niższym poziomie. Z każdym węzłem związane jest pytanie o wartości cech, jeśli pewien obiekt je posiada, to przenosi się go w dół odpowiednią krawędzią. Węzły, z których nie wychodzą już żadne krawędzie, są określane jako *liście*, reprezentują klasy. W formalnym ujęciu, drzewo jest *grafem bez cykli* (pętli), w którym istnieje tylko jedna ścieżka między dwoma różnymi węzłami. Oznacza to, że drzewo jest grafem spójnym, czyli nie można go przedstawić w postaci sumy grafów. Zgodnie z drugą definicją: **drzewo klasyfikacyjne jest drzewem reprezentującym proces podziału zbioru obiektów na jednorodne klasy**. Jego wewnętrzne węzły opisują sposób dokonania tego podziału w oparciu o wartości cech obiektów, a liście odpowiadają klasom, do których należą obiekty. Z kolei krawędzie drzewa reprezentują wartości cech, na podstawie których dokonano podziału [Gatnar 1998].

Z geometrycznego punktu widzenia drzewa klasyfikacyjne są podobne do liniowych funkcji dyskryminacyjnych, ponieważ także dzielą wielowymiarową przestrzeń cech hiperpłaszczyznami, po to by wyodrębnić homogeniczne skupiska obiektów. Na przykład posiadając zbiór obiektów opisanych przez dwie cechy ilościowe x , y należące do jednej z trzech klas: a , b , c , można zbudować jednowymiarowe drzewo klasyfikacyjne. Może ono wyglądać jak na rysunku 1 (obok drzewa zaznaczono dodatkowo linie oddzielające klasy w przestrzeni cech).



Rys.1. Geometryczna interpretacja drzewa klasyfikacyjnego

Fig. 1. Geometrical interpretation of classification tree

Źródło – Source: Gatnar E. 1998. Symboliczne metody klasyfikacji danych. PWN, Warszawa

Drzewa klasyfikacyjne generują w prosty sposób reguły przynależności obiektów do każdej z klas. Mogą być postrzegane wprost jako zbiór reguł klasyfikacji w dysjunkcyjnej postaci normalnej pozwalającej na łączenie rozłącznych zbiorów obiektów w tych samych klasach. Ścieżka prowadząca od korzenia do liścia reprezentuje koniunkcję testów (kompleks), natomiast jeśli do samej klasy prowadzi kilka ścieżek, to można je traktować jako składniki alternatywy. Przykładowe drzewo na rysunku 1 zawiera trzy reguły klasyfikacji:

$$\begin{aligned}
 y > 3,5 \wedge x \leq 2,1 &::> a, \\
 (y > 3,5 \wedge x > 2,1) \vee (y \leq 3,5 \wedge x \leq 2,9) &::> b, \\
 y \leq 3,5 \wedge x > 2,9 &::> c.
 \end{aligned}$$

Wykorzystanie stworzonego drzewa do sklasyfikowania nowego obiektu polega na tym, że poczynając od korzenia przechodzi się od wierzchołka do wierzchołka wzdłuż krawędzi drzewa odpowiadających wartościom cech klasyfikowanego obiektu. W takim ujęciu często nie jest wymagana znajomość wszystkich cech obiektu, co znacznie podnosi praktyczną użyteczność tej metody klasyfikacji. Dotyczy to w szczególności klasyfikacji nieruchomości, gdyż w analizach nie dysponujemy często kompletem informacji o badanym obiekcie oraz posiadamy informacje o zróżnicowanym charakterze (jakościowo-ilościowym).

PROCEDURY TWORZENIA DRZEW KLASYFIKACYJNYCH

Tworzenie drzew klasyfikacyjnych odbywa się poprzez rekurencyjny podział zbioru uczącego na podzbiory aż do uzyskania ich jednorodności ze względu na przynależność obiektów do klas. Głównym celem ich tworzenia jest uzyskanie maksymalnie jednorodnych podzbiorów obiektów z punktu widzenia zmiennej zależnej [Łapczyński 2003]. W trakcie ich konstruowania chodzi również o to, by drzewo było jak najmniejsze, a reguły klasyfikacji jak najprostsze. Większość znanych metod tworzenia drzew klasyfikacyjnych ma podobną konstrukcję, gdyż wywodzi się z trzech klasycznych algorytmów, do których zalicza się wspomniane algorytmy *CLS* i *ID3* oraz algorytm o nazwie „Classification and Regression Trees” (C&RT) rozpropagowany przez Briemana i in. [1984]. Metody tworzenia drzew klasyfikacyjnych mogą prowadzić do budowy drzew **binarnych** i **niebinarnych**. W przypadku drzew binarnych z jednego wewnętrznego węzła wychodzą tylko dwie krawędzie, czyli każdy zbiór obiektów dzieli się na dwa rozłączne podzbiory. Drzewa takie najczęściej występują w przypadku klasyfikacji obiektów o cechach ilościowych, których wartości podlegają dyskretyzacji. Drzewa niebinarne wyróżniają się natomiast tym, że z jednego węzła wychodzą więcej niż dwie krawędzie (zbiór obiektów można podzielić na więcej niż dwa rozłączne podzbiory). Najczęściej występują one w przypadku klasyfikacji obiektów o cechach jakościowych, które mają odpowiednio liczne zbiory wartości. Wyróżnia się ponadto metody budowy drzew jednowymiarowych (podział zbioru obiektów następuje na podstawie pojedynczych cech) oraz wielowymiarowych, gdzie podział zbioru obiektów następuje na podstawie kombinacji liniowych analizowanych cech.

Ogólna procedura tworzenia drzewa klasyfikacyjnego – w nawiązaniu do algorytmu C&RT – obejmuje następujące kroki [Gatnar 1998]:

- a) sprawdzenie czy dany zbiór obiektów S należy do tej samej klasy, jeśli tak to na tym etapie kończymy pracę, jeśli nie – przechodzimy do punktu nr b,
- b) określenie wszelkich możliwych podziałów zbioru S na podzbiory $S_1, S_2, \dots, S_n$, tak aby podzbiory te były jak najbardziej jednorodne,

- c) przeprowadzenie oceny jakości każdego z tych podziałów zgodnie z przyjętym kryterium i wybór najlepszego z nich,
- d) dokonanie podziału zbioru S w wybrany sposób,
- e) wykonanie kroków 1–4 rekurencyjnie dla każdego z podzbiorów.

Podział zbioru S dokonywany w czwartym kroku następuje w oparciu o wartości cech charakteryzujących obiekty. Jednak cechy te nie mogą być wybierane losowo, ponieważ mogłoby to doprowadzić do sytuacji, w której liczba liści drzewa klasyfikacyjnego równałaby się liczbie obiektów. Stąd też przy wyborze cech będących podstawą podziału stosowane są różne miary statystyczne lub metody oparte na teorii informacji. Podczas budowania drzewa klasyfikacyjnego można pominąć fragmenty, które mają niewielkie znaczenie dla jakości rezultatów klasyfikacji (porządkowanie drzewa). W algorytmie C&RT najpierw tworzy się pełne drzewo klasyfikacyjne, niezależnie od rozmiarów jakie by osiągnęło, a następnie dokonuje się porządkowania, czyli usuwane są pewne fragmenty drzewa, w taki sposób, aby wraz ze zmniejszaniem się jednorodności klas nie wzrastał błąd klasyfikacji. Szczegółowe problemy do rozwiązania przed każdym algorytmem tworzenia drzewa klasyfikacyjnego obejmują udzielenie odpowiedzi na następujące pytania:

- w jaki sposób wybrać jedną lub kilka cech, w oparciu o które nastąpi podział zbioru obiektów na klasy (problem miar jakości podziału – jednorodności lub różnicowania)?
- w jakim momencie zaprzestać dalszego podziału podzbiorów obiektów, tak aby uzyskać drzewo o minimalnej liczbie węzłów, nie obniżając przy tym jakości klasyfikacji (problem porządkowania drzewa)?
- w jaki sposób przydzielić obiekty znajdujące się w liściu drzewa do pewnej klasy?

Po rozwiązaniu tych problemów powstałe drzewo klasyfikacyjne winno charakteryzować się maksymalną jednorodnością uzyskanych podzbiorów (klas) obiektów i minimalną liczbą węzłów prowadzących do zbioru możliwie prostych reguł klasyfikacji.

MATERIAŁ, METODY I PRZEBIEG BADAŃ

Przedmiot klasyfikacji stanowi zasób nieruchomości wspólnot mieszkaniowych z udziałem miasta Olsztyna. Na podstawie informacji wyjściowych – uzyskanych od jednostki zarządzającej – stwierdzono, że badany zasób wspólnot mieszkaniowych składa się z 686 obiektów charakteryzujących się znacznym zróżnicowaniem takich atrybutów, jak: liczba lokali w poszczególnych obiektach, wielkość udziału własności komunalnej, funkcje i sposób wykorzystania lokali gminnych, jak również walory lokalizacyjne poszczególnych obiektów i stan techniczny zabudowy. Celem procedur klasyfikacji jest wydzielenie w miarę jednorodnych klas badanych nieruchomości wspólnot mieszkaniowych z udziałem gminy z uwzględnieniem kryteriów istotnych z punktu widzenia racjonalnej gospodarki nieruchomościami.

Podział badanego zasobu nieruchomości wspólnot mieszkaniowych przeprowadzono za pomocą drzewa klasyfikacyjnego jako jednej z hierarchicznych metod klasyfikacji wyróżnianych w ramach taksonomii symbolicznej. Analizy wykonano

z wykorzystaniem programu komputerowego Statistica.pl 8.0 Firmy StatSoft, w tym szczególnie aplikacji metody „C&RT wyczerpującego poszukiwania podziałów jednowymiarowych” [StatSoft, 2001]. Proces budowy drzewa klasyfikacyjnego przeprowadzono metodą interaktywną polegającą na połączeniu wiedzy eksperckiej badacza z automatycznymi procedurami algorytmu numerycznego. W trakcie budowy drzewa klasyfikacyjnego początkowy wybór zmiennych i pierwsze (podziały) gałęzie zostały eksperymentalnie zaprojektowane przez badacza. Dalsze podziały i gałęzie drzewa uzyskano automatycznie.

W pierwszym etapie badań dobrano i opisano cechy diagnostyczne niezbędne do klasyfikacji nieruchomości wspólnot mieszkaniowych z udziałem miasta Olsztyna. Następnie zebrano niezbędne dane źródłowe, wykorzystując informacje z rejestrów prowadzonych przez jednostki gminne i z wywiadu terenowego. Wybierając cechy, wzięto pod uwagę cel klasyfikacji, względy merytoryczne wynikające z gospodarki nieruchomościami, przesłanki statystyczne oraz kompletność i wiarygodność informacji źródłowych. W efekcie tych zabiegów sformułowano zestaw cech opisujących badany zbiór obiektów obejmujący następujące zmienne:

- a) „rodzaj wspólnoty mieszkaniowej” – X_1 ,
- b) „udział własności gminy we wspólnocie mieszkaniowej” – X_2 ,
- c) „typ funkcjonalny nieruchomości wspólnoty mieszkaniowej” – X_3 .

Przyjęto, że cecha X_1 przyjmuje dwie wartości: „mała wspólnota mieszkaniowa” (w przypadku gdy łączna liczba lokali l_i we wspólnocie jest mniejsza niż osiem) oraz „duża wspólnota mieszkaniowa” (jeśli łączna liczba lokali l_i we wspólnocie wynosi osiem lub więcej). Cecha ta jest kluczową zmienną klasyfikacyjną, gdyż w obu rodzajach wspólnot występują odmienne procedury zarządzania nieruchomością wspólną. Z kolei cechy opisującej „udział własności gminy we wspólnocie mieszkaniowej” – X_2 nadano dwa główne poziomy wartości:

- mniejszościowy (gdy udział gminy we wspólnocie u_i jest mniejszy niż 50%),
- większościowy (gdy udział gminy we wspólnocie u_i wynosi 50% lub więcej).

Poziom większościowy cechy X_2 na potrzeby klasyfikacji dodatkowo podzieleno – w ujęciu hierarchicznym – na podpoziom przeważający (gdy udział gminy we wspólnocie u_i wynosi 50% lub więcej, ale nie przekracza 80%) i podpoziom dominujący (gdy udział gminy u_i wynosi 80% lub więcej). Taki sposób dyskretyzacji cechy X_2 wynika ze zróżnicowania reguł podejmowania decyzji gospodarowania we wspólnocie mieszkaniowej z punktu widzenia gminy jako współwłaściciela. Ostatnia cecha opisuje „typ funkcjonalny nieruchomości wspólnoty mieszkaniowej”. Jest to klasyczna cecha jakościowa, której wartości nominalne przypisano w zależności od układu (kombinacji) funkcji lokali gminnych w poszczególnych nieruchomościach wspólnot. Praktycznie cecha X_3 może przyjmować w nieruchomościach wspólnot mieszkaniowych z udziałem miasta Olsztyna dziesięć poziomów wartości:

- 1) „lokalne mieszkalne gminy” – typ m ;
- 2) „lokalne użytkowe gminy o funkcjach komercyjnych” – typ k ;
- 3) „lokalne użytkowe gminy o funkcjach użyteczności publicznej” – typ p ;
- 4) „lokalne mieszkalne z lokalami socjalnymi” – typ ms ;
- 5) „lokalne mieszkalne z lokalami użytkowymi o funkcjach komercyjnych” – typ mk ;

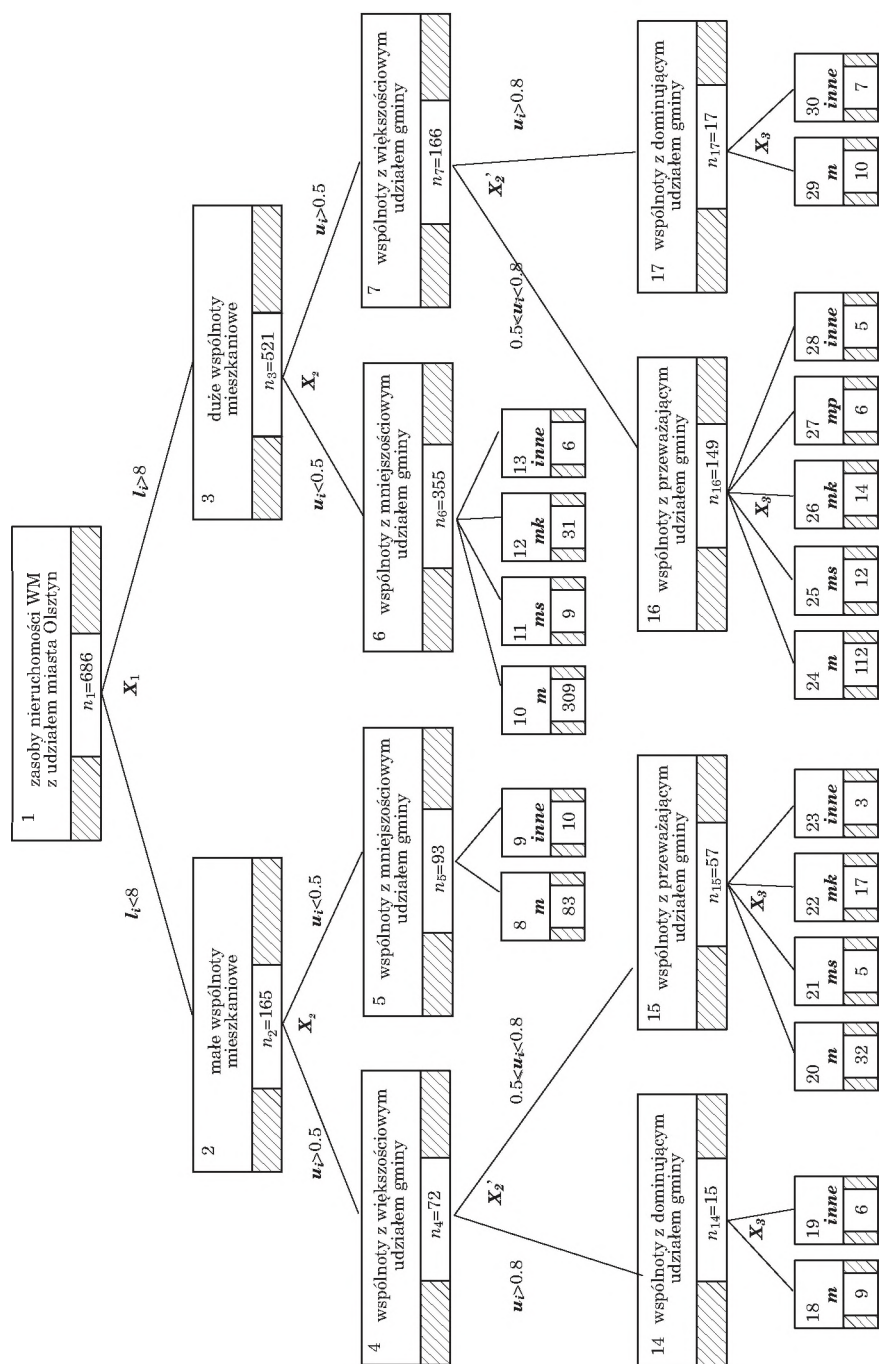
- 6) „lokale mieszkalne z lokalami użytkowymi o funkcjach użyteczności publicznej” – typ *mp*;
- 7) „lokale użytkowe o funkcjach komercyjnych i użyteczności publicznej” – typ *kp*;
- 8) „lokale mieszkalne, socjalne i użytkowe o funkcjach komercyjnych” – typ *msk*;
- 9) „lokale mieszkalne, socjalne i użytkowe o funkcjach użyteczności publicznej – typ” *msp*;
- 10) „lokale mieszkalne, użytkowe o funkcjach komercyjnych i użyteczności publicznej” – typ *mkp*.

Informacje o wartościach cechy X_3 – typach funkcjonalnych lokali stanowiących własność gminy w poszczególnych wspólnotach – są istotne z punktu widzenia celu przeprowadzonej klasyfikacji, gdyż rzutują na społeczno-ekonomiczne aspekty gospodarki nieruchomościami, w tym na politykę czynszową i prywatyzacyjną w zasobach gminnych.

W drugim etapie badań przeprowadzono właściwą klasyfikację zasobu nieruchomości wspólnot mieszkaniowych z udziałem miasta Olsztyna, budując do tego celu odpowiednie drzewo klasyfikacyjne. Proces tworzenia tego drzewa przeprowadzono metodą interaktywną polegającą na połączeniu wiedzy eksperckiej (w pierwszej fazie podziału) z automatycznymi procedurami algorytmu C&RT wyczerpującego poszukiwania podziałów jednowymiarowych. Proces podziału kontynuowano do momentu uzyskania jednorodności wszystkich węzłów końcowych. Te węzły końcowe, których liczebność nie przekraczała pięciu przypadków (nieruchomości) zostały zaagregowane w węzłach końcowych niejednorodnych o nazwie „inne”. Do określenia pożądanej minimalnej liczby przypadków jako kryterium zatrzymania procesu podziałów wykorzystano w algorytmie C&RT opcję „minimalne” n . Wykorzystując proces budowy i uporządkowania drzewa klasyfikacyjnego, uzyskano maksymalnie jednorodne podzbiory nieruchomości wspólnot mieszkaniowych z punktu widzenia przyjętych kryteriów podziału. Stworzone drzewo umożliwia opracowanie w łatwy sposób zbioru możliwie prostych reguł klasyfikacji w postaci koniunkcji kilku warunków logicznych.

WYNIKI BADAŃ I DYSKUSJA

Zastosowane procedury badawcze doprowadziły do uzyskania drzewa klasyfikacyjnego nieruchomości wspólnot mieszkaniowych z udziałem miasta Olsztyna (rys. 2). Analizując architekturę tego drzewa, można ogólnie stwierdzić, że jest ono niebinarne. Wynika to stąd, że ostatnie kryterium podziału – opisane cechą X_3 – miało charakter jakościowy o liczebności zbioru wartości przekraczającej liczbę 2. Poprawnie przeprowadzona klasyfikacja badanego zbioru obiektów doprowadziła do powstania drzewa składającego się z jednego węzła źródłowego na początku drzewa (na rysunku oznaczonego nr 1), dziesięciu węzłów (podziałów) wewnętrznych (nr: 2, 3, 4, 5, 6, 7, 14, 15, 16 i 17) oraz dziewiętnastu węzłów końcowych (liści) – nr: 8, 9, 10, 11, 12, 13, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 i 30. Węzeł źródłowy (nadrzędny o liczebności 686 obiektów) dzieli zasób nieruchomości wspólnot na dwa węzły podrzędne (potomków), z których lewy (nr 2) reprezentuje podzbiór małych



Rys. 2. Drzewo klasyfikacyjne nieruchomości wspólnot mieszkaniowych z udziałem miasta Olsztyn: np. 8 – numer węzła, m – nazwa węzła, np. 83 – liczebność obiektów w węźle

Źródło: Opracowanie własne

Fig. 2. Classification tree of housing community properties with the Olsztyn municipality share: ex. 8 – nodes number, np. m – nodes name, ex. m – number of elements in the node

Source: Own elaboration

wspólnot mieszkaniowych (165 obiektów), a prawy (nr 3) podzbior dużych wspólnot mieszkaniowych (521 obiektów). W dalszym ciągu realizacji procesu klasyfikacji oba analizowane węzły wewnętrzne uzyskały po dwa węzły podrzędne oznaczone odpowiednio numerami: 4 i 5 oraz 6 i 7. Jako kryterium podziału obu podzbiorów przyjęto wielkość udziału gminy we wspólnocie mieszkaniowej. Ostatnia faza podziałów wewnętrznych polegała na dalszym rozwarstwieniu podzbiorów wspólnot z większościowym udziałem gminy, w wyniku czego węzły wewnętrzne nr 4 i 7 uzyskały swoich potomków odpowiednio o numerach 14 i 15 oraz 16 i 17. Utworzenie tych węzłów wewnętrznych zamknęło proces rozrostu gałęzi drzewa. Następnie wygenerowano węzły końcowe (liście), które kończą proces klasyfikacji obiektów, gdyż poza nimi nie podejmuje się dalszych podziałów (decyzji). Z punktu widzenia kryterium wyodrębniania liści, jakim były wartości cechy X_3 , można stwierdzić, że trzynaście spośród wszystkich dziewiętnastu węzłów końcowych jest tzw. węzłami czystymi – zawierającymi przypadki maksymalnie jednorodne (pochodzące z dokładnie jednej klasy). Natomiast węzły końcowe typu „inne” – oznaczone nr: 9, 13, 19, 23, 28 i 30 – są węzłami niejednorodnymi (zanieczyszczonymi) skupiającymi kilkuelementowe podzbiory obiektów w różnych typach funkcjonalnych.

Ze schematu drzewa klasyfikacyjnego wynika, że spośród 686 nieruchomości wspólnot mieszkaniowych z udziałem miasta Olsztyna 521 obiektów (ponad 75%) stanowią duże wspólnoty, natomiast jedynie 165 obiektów (niepełna 25%) zalicza się do małych wspólnot. Z punktu widzenia gminy jako współwłaściciela sytuacja taka jest korzystna, gdyż w dużych wspólnotach procesy gospodarowania są łatwiejsze do realizacji i na ogół efektywniejsze. Procesy decyzyjne zarządzania nieruchomościami komunalnymi komplikuje jednakże stale zmniejszający się udział gminy we wspólnotach wskutek prywatyzacji. Badane drzewo pokazuje, że miasto Olsztyn posiadało większościowe udziały jedynie w 44% populacji małych i 32% populacji dużych wspólnot mieszkaniowych. Zaledwie w 15 małych (9%) i 17 dużych (3%) wspólnotach gmina mogła podejmować decyzje w pełni samodzielnie. Z analizy struktury liści drzewa wynika ponadto, że badany zasób nieruchomości wspólnotowych wykazuje relatywnie małe zróżnicowanie funkcjonalne lokali komunalnych. Wyniki pokazują bowiem, że w 555 obiektach (81% populacji) udziały gminne we wspólnotach składały się wyłącznie z mieszkań komunalnych (typ *m*). Na drugim miejscu pod względem liczebności uplasowały się wspólnoty, w których udział gminy składał się z lokali mieszkalnych i użytkowych komercyjnych, w tym głównie garaży (typ *mk*) – 71 obiektów (10% populacji). Kolejne pozycje zajęły wspólnoty mieszkaniowe typu *ms* (30 obiektów – 4% populacji) oraz *mp* (11 obiektów – niepełna 2% populacji). Pozostałe typy funkcjonalne nieruchomości odegrały marginalną rolę w klasyfikacji, gdyż ich liczebność nie przekroczyła 5 obiektów.

PODSUMOWANIE I WNIOSKI

Przeprowadzone badania pozwalają stwierdzić, że procedury taksonomii symbolicznej realizowane z wykorzystaniem metodyki tworzenia drzew klasyfikacyjnych mogą dostarczać efektywnych narzędzi do hierarchicznego podziału złożonych

obiektów symbolicznych, którymi są nieruchomości wspólnot mieszkaniowych. Cechą charakterystyczną metody drzew klasyfikacyjnych jest jednoczesna realizacja czynności podziału zbioru obiektów na klasy oraz określania charakterystyki wydzielonych klas za pomocą relatywnie prostych reguł przynależności obiektów do klas. Ujęcie takie czyni je strukturami użytecznymi w praktyce klasyfikowania nowych obiektów, w tym szczególnie w sytuacjach, gdy nie dysponujemy o nich kompletem informacji. Do zalet drzew klasyfikacyjnych zalicza się ponadto: łatwość interpretacji uzyskanych wyników i ich hierarchiczną naturę, jak również elastyczność techniki postępowania, która umożliwia badanie wpływu różnych zmiennych wyjściowych (kryteriów) na wyniki klasyfikacji. Zastosowane procedury badawcze łączące wiedzę ekspercką badacza z algorytmem komputerowym doprowadziły do utworzenia drzewa klasyfikacyjnego reprezentującego maksymalnie jednorodne klasy nieruchomości wspólnot mieszkaniowych z punktu widzenia przyjętych kryteriów podziału.

Wyniki badań umożliwiają sformułowanie następujących wniosków empirycznych:

1. Efekty zastosowania drzewa, jako narzędzia hierarchicznego podziału wielocechowych obiektów symbolicznych, dowodzą, że konstrukcje tego typu mogą być z powodzeniem stosowane do rozwiązywania zadań klasyfikacji nieruchomości wspólnot mieszkaniowych z udziałem gminy z uwzględnieniem kryteriów istotnych z punktu widzenia racjonalnej gospodarki nieruchomościami komunalnymi.

2. W badanej populacji nieruchomości wspólnot mieszkaniowych z udziałem miasta Olsztyna zdecydowanie dominują duże wspólnoty mieszkaniowe. Z punktu widzenia gminy jest to sytuacja korzystna, gdyż w obiektach dużych wspólnot gospodarowanie nieruchomościami nie wymaga jednomyślności, przez co jest skuteczniejsze i na ogół efektywniejsze.

3. W przeważającej liczbie wspólnot mieszkaniowych z udziałem miasta Olsztyna gmina utraciła większość udziałów. Jest efekt sukcesywnej sprzedaży mieszkań komunalnych metodą prywatyzacji rozproszonej. Taki stan rzeczy mocno komplikuje procesy decyzyjne w obszarze zarządzania nieruchomościami komunalnymi, ściśle uzależniając gminę od decyzji właścicieli uprzednio sprywatyzowanych lokali.

4. Analiza struktury węzłów końcowych utworzonego drzewa pokazuje, że udziały gminy Olsztyn we wspólnotach mieszkaniowych cechują się relatywnie małym zróżnicowaniem funkcjonalnym. Badana gmina posiada we wspólnotach prawie wyłącznie lokale mieszkalne w podstawowym standardzie z niewielkim zasobem lokali użytkowych (głównie garaży) oraz jeszcze mniejszą grupą lokali socjalnych.

PIŚMIENNICTWO

- Breiman L., Friedman J.H., Olshen R.A., Stone C.J., 1984. Classification and regression trees. Wadsworth, Belmonth.
- Gatnar E., 1998. Symboliczne metody klasyfikacji danych. PWN, Warszawa.
- Hunt E.B., Marin J., Stone P.J. 1966. Experiments in induction. Academic Press, New York.

- Łapczyński M., 2003. Drzewa klasyfikacyjne w badaniach satysfakcji i lojalności klientów. Materiały szkoleniowe firmy StatSoft, Warszawa-Kraków.
- Michalski R.S., 1980. Knowledge acquisition through conceptual clustering: a theoretical framework and an algorithm for partitioning data into conjunctive concepts. [W:] International Journal of Policy Analysis and Information Systems, t. 4.
- Quinlan J.R., 1986. Induction of decision trees. [W:] Machine learning, t. 1.
- StatSoft, 2001. Drzewa klasyfikacyjne i regresyjne. Statsoft Polska Sp. z o.o., Kraków, www.statsoft.pl.

THE CLUSTERING OF HOUSING COMMUNITY PROPERTIES WITH THE USE OF CLASSIFICATION TREE

Abstract. The paper presents results of research on clustering housing community properties with the use of classification tree. The subject of the research was the population of housing community properties with the Olsztyn community share. The research procedures were used in particular way that enabled acquired classification tree include all the essential criteria in terms of the rational community properties management. The classification tree presents hierarchical order of maximum homogenous subclusters with the minimum number of nodes generating clusters of simple classification rules. The analysis have shown that big housing communities were dominant in the population and that it contributed to improved management efficiency. One have noticed that in the overwhelming majority of analyzed housing communities Olsztyn community lost most of the shares in a result of shattered privatization. This situation greatly complicates the decision-making process in terms of community properties management.

Key words: clustering, housing community, classification trees, community properties management

Zaakceptowano do druku – Accepted for print: 30.11.2010