Adam Muczyński

Grupowanie nieruchomości wspólnot mieszkaniowych z wykorzystaniem sieci Kohonena

Acta Scientiarum Polonorum. Administratio Locorum 8/4, 5-15

2009

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.

MUZEUM HISTORII POLSKI

ACTA SCIENTIARUM POLONORUM
Acta Sci. Pol., Administratio Locorum 8(4) 2009, 5-15

GRUPOWANIE NIERUCHOMOŚCI WSPÓLNOT MIESZKANIOWYCH Z WYKORZYSTANIEM SIECI KOHONENA

Andrzej Muczyński
Uniwersytet Warmińsko-Mazurski w Olsztynie

Streszczenie. W pracy przedstawiono wyniki badań w zakresie grupowania nieruchomości wspólnot mieszkaniowych z udziałem miasta Olsztyna. Procedurę grupowania zrealizowano z wykorzystaniem sztucznej sieci neuronowej Kohonena. W badaniach posłużono się zestawem cech obiektów obejmujących takie zmienne wejściowe, jak: udział gminy we wspólnocie, walory lokalizacyjne, stan zagospodarowania terenu i stan techniczny zabudowy. Po nauczeniu i sprawdzeniu sieci uzyskano dwuwymiarową mapę topologiczną odwzorowującą osiem jednorodnych grup obiektów poddanych badaniom. Na podstawie obrazu mapy i danych klasyfikacyjnych dokonano szczegółowej analizy i syntezy wyników grupowania. Z badań wynika, że w przeważającej liczbie wspólnot mieszkaniowych z udziałem miasta Olsztyna gmina utraciła większość udziałów, a proces rozproszonej prywatyzacji koncentrował się w obiektach o dobrej lokalizacji i znacznych potrzebach remontowych. Grupowane obiekty cechowały się zróżnicowanymi walorami lokalizacji ogólnej i szczegółowej oraz co najwyżej średnim stanem technicznym. Wytrenowana sieć jest zdolna do oceny nowych obiektów, nieprezentowanych w trakcie uczenia. Zastosowana metoda grupowania okazała się bardziej efektywna od metody *k*-średnich, gdyż doprowadziła do uzyskania skupień obiektów bardziej jednorodnych wewnętrznie. Podział na grupy metodą Kohonena był ponadto bardziej przejrzysty i użyteczny z punktu widzenia przyjętego celu grupowania.

Słowa kluczowe: grupowanie, wspólnoty mieszkaniowe, sieci Kohonena, gospodarka nieruchomościami

WPROWADZENIE

Grupowanie (z ang. clustering) polega na wyodrębnianiu w danym zbiorze przedmiotów, obiektów czy zjawisk w miarę jednorodnych grup (podzbiorów, klas) z punktu widzenia określonego kryterium opartego na cechach analizowanych przedmiotów. Proces

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

ten stanowi działanie pierwotne w stosunku do klasyfikacji, gdyż prowadzi do zdefiniowania wzorców klas (skupień) obiektów, podczas gdy klasyfikacja polega na przyporządkowaniu badanych obiektów – na podstawie ich cech – do określonych klas wzorcowych. Z grupowaniem mamy zatem do czynienia wtedy, gdy przed podziałem zbioru wejściowego nie jest określona ani liczba grup (klas) obiektów, ani nie są znane przedziały wartości cech charakterystycznych. Zagadnienie grupowania zalicza się do podstawowych metod (technik) ekstrapolacji danych i pozyskiwania wiedzy. Pozwala ono na rozpoznanie jednorodnych przedmiotów badania ułatwiających wyodrębnienie ich zasadniczych cech, a także na redukcję dużej liczby danych pierwotnych do kilku głównych kategorii, co sprzyja zmniejszeniu nakładu czasu i kosztów badań. W efekcie grupowania możliwe jest poznanie struktury danych, odkrycie związków między obiektami i ich atrybutami oraz sformułowanie wniosków uogólniających.

Do rozwiązywania zadań grupowania (klasyfikacji) obiektów w przypadkach gdy brak jest *a priori* informacji o wzorcach klas obiektów wykorzystywane są powszechnie metody taksonomii bezwzorcowej, czyli metody rozpoznawania bez nauczyciela, określane tradycyjnie pojęciem analizy skupień [Muczyński, Mrozik 2009]. W jej skład wchodzi zbiór metod statystycznych umożliwiających grupowanie danych w sensowne struktury lub klasy poprzez analizę podobieństw w obszarach poddanych badaniu. Klasyczne metody analizy skupień prowadzą do podziału badanego zbioru na względnie jednorodne klasy obiektów podobnych według określonej miary podobieństwa, identyfikowanej zwykle z odległością między nimi. Dynamicznie rozwijające się technologie pozyskiwania i przetwarzania informacji w ostatnim dziesięcioleciu spowodowały, że w procedurach klasyfikacyjnych obok metod statystycznych wykorzystywane są sztuczne sieci neuronowe. Są one systemami przetwarzania informacji symulującymi działanie rzeczywistych, funkcjonujących w mózgu i uczących się struktur rozpoznających obiekty [Muczyński, Kołek 2007]. Sieci te występują w dwóch formach: jako programy komputerowe uruchamiane na mikrokomputerach typu PC i jako systemy wyspecjalizowane – tzw. neurokomputery – komputery wyposażone w elektroniczne układy scalone odwzorowujące układy nerwowe. W problemach klasyfikacyjnych wykorzystuje się różne typy sztucznych sieci neuronowych, do trenowania których stosuje się zarówno metody uczenia z nauczycielem, jak i bez nauczyciela. W metodach bezwzorcowych liczba neuronów wyjściowych równa się zwykle liczbie skupień, do których mają być zaklasyfikowane badane obiekty. Jednym z najważniejszych modeli neuronowych realizujących procedurę grupowania obiektów wielocechowych bez korekty zewnętrznej (bez wzorca) jest sieć Kohonena.

Celem pracy jest przeprowadzenie grupowania nieruchomości wspólnot mieszkaniowych z wykorzystaniem sztucznej sieci neuronowej Kohonena. Przedmiot badań stanowi zasób nieruchomości wspólnot mieszkaniowych z udziałem miasta Olsztyna. Procedury badawcze zostały tak ukierunkowane, aby wyniki grupowania dostarczyły użytecznych informacji do podejmowania decyzji w obszarze racjonalnej gospodarki nieruchomościami. Uzyskane grupy (skupienia) obiektów jednorodnych powinny odzwierciedlać rzeczywiste zróżnicowanie badanego zasobu nieruchomości wspólnot mieszkaniowych z udziałem gminy i wskazywać występujące zależności między obiektami i ich atrybutami.

SIECI KOHONENA JAKO NARZĘDZIE GRUPOWANIA

Sieci Kohonena stanowią specyficzną klasę sztucznych sieci neuronowych wzorowanych na właściwościach topologicznych ludzkiego mózgu. Należą do klasy sieci samouczących i samoorganizujących się z wbudowaną konkurencją i mechanizmem sąsiedztwa. Ich nazwa pochodzi od nazwiska fińskiego uczonego Teuvo Kohonena, który je pierwszy skonstruował i badał w latach 1979–1982 [Osowski 1996]. Sieci te występują w literaturze także pod nazwą samoorganizujących się map cech (*SOFM* – self-organizing feature maps). Odwzorowanie n-wymiarowych danych wejściowych sieci na m-wymiarowy wektor wyjściowy realizowane jest w trybie uczenia bez nauczyciela, co oznacza, że użytkownik ma do dyspozycji jedynie wzorce wejściowe, natomiast nie posiada żadnych wzorców wyjścia. Podczas procesu uczenia w tym trybie dane uczące, zawierające wyłącznie zmienne wejściowe, pozwalają na wykrycie i wykorzystanie pewnych cech zawartych w sposób ukryty w danych (na przykład ich wzajemnego podobieństwa), nie ma natomiast z góry zdefiniowanego sygnału wyjściowego, oczekiwanego jako pożądana reakcja sieci. Bazując na obserwacjach danych wejściowych, sieci tego typu próbują nauczyć się zatem samodzielnie „naturalnej" struktury danych, w tym rozpoznawania skupień (grup obiektów) występujących w zbiorze nieskategoryzowanych danych uczących. Skupienie stanowi grupę danych (obiektów) podobnych do siebie w jakimś stopniu, natomiast dane należące do różnych skupień różnią się między sobą. Można przyjąć, że poszczególne neurony będą rozpoznawały poszczególne skupienia danych. Analizowane sieci mogą ponadto nauczyć się kojarzenia podobnych klas danych – bez względu na to, ile i jakie są kryteria podobieństwa danych. Dzięki projekcji wykonanej przez sieć Kohonena użytkownik może zatem lepiej zrozumieć dane, co umożliwia ulepszenie procesu ich dalszej analizy i wykorzystania – za pomocą tej samej sieci neuronowej lub z wykorzystaniem innych narzędzi [StatSoft 2001]. W związku z tym głównym zastosowaniem sieci Kohonena jest grupowanie i klasyfikowanie danych (obiektów) wejściowych w sensowne skupienia zgodnie z wewnętrzną ich logiką (bez wzorców i korekt zewnętrznych). Podobne dane wejściowe są klasyfikowane do tej samej grupy, natomiast powstałe grupy rozpoznawane są przez sieć na podstawie korelacji danych wejściowych [Witkowska 2002]. Dzieje się tak dzięki procesowi samoorganizacji sieci, w wyniku której zbliżone do siebie skupienia danych reprezentowane są przez położone blisko siebie neurony warstwy wyjściowej, co powoduje, że neurony te tworzą mapę topologiczną danych wejściowych, która stanowi narzędzie wizualizacji stopnia podobieństwa pomiędzy klasami [Tadeusiewicz 1998].

Sieci Kohonena posiadają w swej strukturze dwie warstwy neuronów: warstwę wejściową oraz warstwę wyjściową (zwaną konkurencyjną lub Kohonena) składającą się z neuronów radialnych, która po wytrenowaniu sieci tworzy mapę topologiczną. Warstwa wejściowa zbudowana jest z tylu neuronów, ile zmiennych (cech) opisują wektory danych wejściowych. Neurony tej warstwy nie przetwarzają wprowadzanych do sieci danych, a tylko przekazują sygnał wejściowy do neuronów warstwy wyjściowej. Z kolei warstwa wyjściowa składa się z pewnej – ustalonej przez projektanta – liczby neuronów przetwarzających wyposażonych w radialną funkcję wyznaczającą potencjał postsynaptyczny. Neurony tego typu realizują klasyfikację sygnałów poprzez pomiar odległości punktów reprezentowanych przez wektory wejściowe od środków hipersfer wyznaczonych przez

wektory wag neuronów warstwy wyjściowej. Cechą wyróżniającą warstwę tworzącą mapę topologiczną jest to, że jest ona dwuwymiarowa (tzn. zbiorowość neuronów wyjściowych ułożona jest w regularną siatkę rozciągającą się w dwóch kierunkach). Neurony tej warstwy nie są ze sobą połączone, natomiast dla wygody interpretacji wyników przyjęto model, w którym neurony te stanowią jak gdyby węzły dwuwymiarowej siatki o kwadratowych oczkach. W procesie uczenia sieci siatka ta może być dowolnie zginana i wykrzywiana w n-wymiarowej przestrzeni wejść w taki sposób, aby mapa topologiczna zachowała oryginalną strukturę danych wejściowych.

Do uczenia sieci Kohonena wykorzystywany jest algorytm iteracyjny, zwany uczeniem konkurencyjnym lub uczeniem typu „zwycięzca bierze wszystko". Uczenie rozpoczyna się od losowej inicjalizacji wag neuronów wyjściowych (centrów radialnych). Następnie na wejścia sieci wprowadzane są wartości zmiennych (cech) objaśniających, opisujących kolejne obiekty. Neurony sieci w odpowiedzi na sygnał wejściowy rywalizują ze sobą i w każdym przypadku zostanie wyszukany taki neuron wyjściowy, którego wagi są w największym stopniu zbliżone do wektora wartości wejściowych. Neuron ten, zwany neuronem zwycięskim, reprezentuje centrum najbardziej zbliżone do prezentowanego na wejściu przypadku uczącego. To wybór neuronu zwycięskiego decyduje o sposobie zakwalifikowania danej wejściowej. Następnie wagi neuronu zwycięskiego modyfikowane są w taki sposób, aby upodobnić jego wzorzec do wektora wejściowego. Wraz z neuronem zwycięskim w podobny sposób modyfikowane są wagi neuronów sąsiednich – wyznaczonych w oparciu o przyjęty wzór topologii sieci. Algorytm iteracyjny działa przez dużą liczbę epok w taki sposób, że w trakcie każdej epoki prezentowany jest każdy z przypadków uczących [StatSoft 2001). W trakcie procesu uczenia stopniowo zmniejszany jest współczynnik uczenia (odpowiadający za skalę każdorazowej modyfikacji neuronów) oraz wielkość sąsiedztwa. Stąd też w początkowej fazie uczenia algorytm tworzy przybliżoną mapę topologiczną, którą następnie precyzyjnie dostraja, tak aby pojedyncze neurony odpowiadały niewielkim skupieniom podobnych do siebie przypadków wejściowych. W tych późniejszych etapach procesu uczenia pojedyncze neurony wchodzące w skład skupienia są precyzyjnie „rozprowadzane" z uwzględnieniem różnych form powiązanych ze sobą przypadków [Tadeusiewicz 1998]. Dzięki temu nawet mały stopień zróżnicowania poszczególnych „podgrup" danych wejściowych wchodzących w skład rozpoznawanej grupy może zostać uwzględniony w formie oddzielnych (ale sąsiadujących ze sobą) neuronów sieci będących ich detektorami. Przedstawiona procedura ucząca prowadzi zatem nie tylko do wyznaczenia centrów skupień w wielowymiarowej przestrzeni wejść, ale także dodatkowo porządkuje neurony w taki sposób, że neurony reprezentujące centra położone blisko siebie w przestrzeni wejść są również usytuowane blisko siebie na modelu dwuwymiarowej mapy topologicznej. Oznacza to w praktyce, że im bardziej dwa obiekty będą do siebie podobne, to tym bliżej siebie znajdować się będą odpowiadające im neurony na mapie topologicznej.

Po nauczeniu sieci Kohonena poprawnego rozpoznawania struktury prezentowanych danych wejściowych można jej użyć jako narzędzia przeprowadzającego grupowanie obiektów i wizualizację danych w celu ich lepszego poznania. Przedtem należy jednak sprawdzić, czy sieć jest już wystarczająco wytrenowana. Służyć do tego może statystyka opisująca częstość zwycięstw neuronów, ukazująca ile razy każdy z nich zwyciężył podczas prezentacji danych uczących [StatSoft 2001]. Pozwala ona na obserwowanie, gdzie

na mapie topologicznej tworzą się skupienia o największej częstości występowania obiektów wejściowych. Wysokie częstości zwycięstw neuronów wskazują na wyraźnie ukształtowane skupienia na mapie topologicznej, natomiast neurony z zerową częstością zwycięstw świadczą o niecałkowitym powodzeniu uczenia, gdyż nie zostały one wykorzystane podczas uczenia, a zatem sieć nie wykorzystała wszystkich dostępnych w niej zasobów. Do oceny poprawności działania sieci istotne jest, aby statystyki częstości zwycięstw neuronów w zbiorze uczącym odpowiadały podziałowi na skupienia w zbiorze walidacyjnym. Ważnym elementem przygotowania sieci do użytkowania jest ponadto właściwe opisanie mapy topologicznej przez identyfikację uzyskanych skupień i ich znaczeń. Na tej podstawie następuje przypisanie określonych etykiet (nazw) do neuronów wchodzących w skład mapy topologicznej, co pozwala na dokonanie pogłębionej analizy jakościowej wszelkich danych podawanych na wejścia sieci. Tak przygotowana sieć jest zdolna do klasyfikowania nowych przypadków, dzięki swej zdolności do uogólniania wiedzy zdobytej w trakcie uczenia.

MATERIAŁ, METODY I PRZEBIEG BADAŃ

Przedmiot grupowania stanowi zbiór nieruchomości wspólnot mieszkaniowych z udziałem miasta Olsztyna. Na podstawie informacji wstępnych stwierdzono, że przeznaczony do badań zasób wspólnot mieszkaniowych składa się z 680 obiektów, które charakteryzuje znaczne zróżnicowanie walorów lokalizacyjnych, sposobu zagospodarowania i stanu technicznego budynków. Celem procedur grupowania jest przeprowadzenie nienadzorowanego podziału badanego zbioru nieruchomości na w miarę jednorodne grupy obiektów z uwzględnieniem potrzeb wynikających z gospodarki nieruchomościami. Proces grupowania nieruchomości wspólnot mieszkaniowych zrealizowano metodą analizy skupień z zastosowaniem sztucznej sieci neuronowej Kohonena. Analizy wykonano z wykorzystaniem programu komputerowego Statistica.Pl 8.0 Firmy StatSoft, w tym szczególnie aplikacji specjalistycznej zawartej w tym programie o nazwie Sieci Neuronowe. W trakcie konstruowania i testowania sieci wykorzystano narzędzie wspomagające, jakim jest Automatyczny Projektant Sieci. Narzędzie to ocenia automatycznie dużą liczbę różnych architektur sieci o różnym stopniu złożoności, wybierając zbiór tych, które najbardziej odpowiadają danemu problemowi.

W pierwszym etapie badań przeprowadzono wstępny dobór i kwantyfikację cech diagnostycznych we wszystkich nieruchomościach wspólnot mieszkaniowych z udziałem gminy Olsztyn stanowiących przedmiot grupowania. Następnie przystąpiono do gromadzenia danych surowych o stanach tych cech w poszczególnych obiektach. Zbierając dane źródłowe, wykorzystano informacje zawarte w rejestrach prowadzonych przez jednostki gminne oraz dane z wywiadu terenowego. Wybierając cechy wzięto pod uwagę cel grupowania, względy merytoryczne wynikające z gospodarki nieruchomościami, przesłanki statystyczne oraz dostępność i wiarygodność informacji źródłowych. Do wstępnie wybranego zestawu cech opisujących badany zbiór obiektów zaliczono zestaw następujących dziewięciu zmiennych:

a) udział własności gminy we wspólnocie mieszkaniowej – X_{UG}
b) walory lokalizacji ogólnej nieruchomości – X_{LO}

c) walory lokalizacji szczegółowej nieruchomości – X_{LS}
d) stan zagospodarowania terenu wokół budynku – X_{ZT}
e) stan techniczny: elewacji budynku – X_{EB}, dachu budynku – X_{DB}, klatek schodowych – X_{KS}, schodów i drzwi zewnętrznych – X_{SD} oraz stolarki okiennej – X_{OK}.

Sposób kwantyfikacji wytypowanych cech oparto na skalach porządkowych. Kwantyfikacja polegała na przypisaniu badanym cechom określonych wartości liczbowych. Opisując cechy ilościowe przyjęto zasadę, że wartość zmiennej odwzorowuje przedziały zmienności wartości cechy. Cechy jakościowe odwzorowano z kolei za pomocą skal interwałowych, w których liczbę interwałów powiązano z zakresem zmienności cechy w populacji. W wyniku tych analiz stany poszczególnych cech diagnostycznych wyrażono odpowiednio na skalach trzy- lub pięciostopniowych, a wyróżnionym stopniom przypisano liczby naturalne według przesłanek wynikających z gospodarki nieruchomościami. W rezultacie przyjęto następujący sposób kwantyfikacji stanów wyróżnionych cech diagnostycznych:

- udział własności gminy we wspólnocie mieszkaniowej (X_{UG}) opisano skalą trzystopniową: wysoki (powyżej 80%) – 3, średni (50–80%) – 2 oraz niski (poniżej 50%) – 1.
- walory lokalizacji ogólnej (X_{LO}) i szczegółowej (X_{LS}) opisano skalą trzystopniową: lokalizacja bardzo dobra – 3, lokalizacja średnia – 2 oraz lokalizacja zła – 1.
- pozostałe cechy X_{ZT}, X_{EB}, X_{DB}, X_{KS}, X_{SD}, X_{OK} opisano skalą pięciostopniową: stan bardzo dobry – 5, dobry – 4, średni –3, zły –2 i bardzo zły – 1.

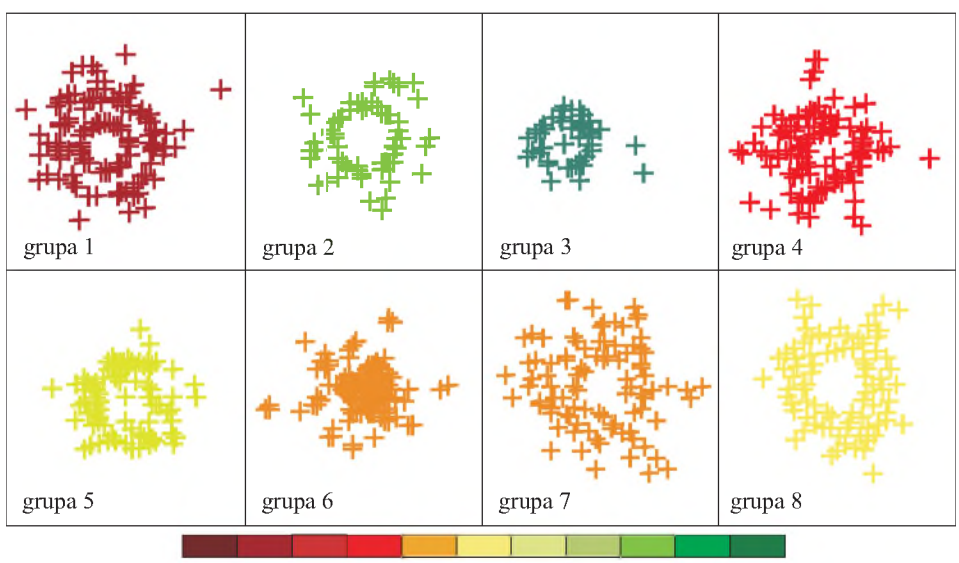
W drugim etapie badań dokonano przeglądu i sprawdzenia zebranych danych surowych pod względem ich wiarygodności i kompletności. Następnie przeprowadzono weryfikację wstępnie wytypowanego zestawu cech diagnostycznych polegającą na eliminacji zmiennych silne skorelowanych między sobą i wprowadzeniu w ich miejsce zmiennych agregatowych. Zabieg ten wykonano w odniesieniu do pięciu cech stanu technicznego: X_{EB}, X_{DB}, X_{KS}, X_{SD} i X_{OK}, które zastąpiono zmienną agregatową X_{BM} opisującą stan techniczny elementów budynku wspólnoty mieszkaniowej. W efekcie tych działań uzyskano wektor pięciu cech diagnostycznych – opisanych symbolami: X_{UG}, X_{LO}, X_{ZT}, X_{KS} i X_{BM}, którym nadano status zmiennych wejściowych. Wartości tych cech jako parametry poszczególnych przypadków wejściowych wykorzystano w procedurze konstruowania i uczenia sieci neuronowej rozwiązującej zadanie grupowania.

W trzecim etapie badań przeprowadzono właściwy proces grupowania nieruchomości wspólnot mieszkaniowych z udziałem miasta Olsztyna za pomocą sieci Kohonena. Jako typ problemu w aplikacji komputerowej wybrano analizę skupień z opcją samoorganizującej się mapy cech. Na wstępie zbiór 680 przypadków podzielono losowo na trzy podzbiory: uczący (do uczenia sieci: 70% zbioru), walidacyjny (do śledzenia błędu działania sieci: 15% zbioru) oraz testowy (do niezależnej oceny działania sieci: 15% zbioru). W dalszej kolejności zrealizowano liczne próby samouczenia sieci neuronowej w powiązaniu z wariantowym poszukiwaniem najlepszych wymiarów mapy topologicznej do rozwiązania postawionego zadania w dostępnym zbiorze uczącym. Podczas eksperymentowania mapie topologicznej nadawano zróżnicowane formy i rozmiary. Do przeprowadzenia ostatniego trenowania sieci wykorzystano sieć o wymiarach 2 x 4, tzn. z 8 neuronami na wyjściu. Czas trwania uczenia wyniósł 10 000 epok, a jego szybkość na poziomie od 0,001 do 0,0001. Parametr sąsiedztwa zmieniał się od 1 (na początku) do 0 (na końcu). Analizując błędy działania sieci neuronowej uznano, że tym razem sieć została nauczona

w sposób poprawny [Kozicki 2009]. Liczba grup (skupień) obiektów również wydaje się być odpowiednia z punktu widzenia przyjętego celu grupowania. Architektura sieci Kohonena ostatecznie uformowana do rozwiązania zadania grupowania nieruchomości wspólnot mieszkaniowych z udziałem miasta Olsztyna składa się zatem z pięciu neuronów w warstwie wejściowej i ośmiu neuronów na wyjściu sieci.

WYNIKI BADAŃ I DYSKUSJA

Szczegółowe wyniki badań empirycznych dotyczących podziału zbioru przedmiotowych nieruchomości na jednorodne grupy przedstawiono na podstawie danych uzyskanych przez Kozickiego [2009]. Graficzny obraz wykonanego grupowania zaprezentowano za pomocą mapy topologicznej, której nadano formę dwuwymiarowej siatki kwadratów (rys. 1). Mapa ta, przedstawiając obraz warstwy wyjściowej sieci nauczonej grupowania sygnałów (obiektów) wejściowych, składa się z dwóch wierszy i czterech kolumn. W węzłach siatki znajduje się osiem wyraźnych skupień odzwierciedlających „ślady" zwycięstw każdego z neuronów radialnych będącego jednocześnie detektorem podobnych do siebie przypadków (obiektów) prezentowanych na wejściu sieci. Mapa topologiczna, poprzez wizualizację zależności między obiektami w pięciowymiarowej przestrzeni wejść, pozwala na zbadanie stopnia podobieństwa obiektów położonych wewnątrz każdej z grup, jak również na ocenę relacji podobieństwa zachodzącego między wyróżnionymi grupami obiektów jednorodnych.



Rys. 1. Mapa topologiczna sieci Kohonena w wyniku grupowania nieruchomości wspólnot mieszkaniowych z udziałem miasta Olsztyna [na podstawie Kozickiego 2009]

Fig. 1. Self-organized feature map of housing community properties that the Olsztyn Commune have a share in [on the basis of Kozicki 2009]

Z obrazu mapy topologicznej wynika, że badana zbiorowość 680 nieruchomości wspólnot mieszkaniowych z udziałem miasta Olsztyna została podzielona na osiem jednorodnych grup obiektów podobnych do siebie z uwzględnieniem stanów przyjętych zmiennych wejściowych. Grupy te są podzbiorami rozłącznymi o różnym stopniu wewnętrznego podobieństwa obiektów jednorodnych i zróżnicowanej liczebności. Najbardziej liczna okazała się grupa 1, w której neuron zwycięski wykazał swoją aktywność 118 razy. Z kolei najmniej liczna jest grupa 3, gdzie neuron zwycięski był aktywowany zaledwie 46 razy. Szczegółowe dane odnośnie liczby obiektów w poszczególnych grupach zawarto w tabeli 1. W tabeli tej przedstawiono ponadto wyniki zaagregowanych danych klasyfikacyjnych o cechach obiektów wewnątrz grup na podstawie dominujących (najczęściej występujących) stanów wartości zmiennych diagnostycznych.

Z danych zestawionych w tabeli 1 wynika, że w każdej z grup występują po 2–3 cechy przyjmujące wartości w tym samym przedziale zmienności, co sprawia, że zmienne te determinują przynależność obiektu do określonej grupy. Do cech mających decydujący wpływ na uzyskane wyniki grupowania należy zatem zaliczyć: udział własności gminy X_{UG}, lokalizację ogólną X_{LO} i szczegółową X_{LS} oraz zmienną agregatową stanu technicznego X_{BM} (grupa 6 i 7). Dane z tabeli 1 pokazują, że w badanym zbiorze obiektów przeważają wspólnoty o niskim udziale własności komunalnej, w których gmina utraciła

Tabela 1. Zestawienie jednorodnych grup nieruchomości wspólnot mieszkaniowych z udziałem miasta Olsztyna (ujęcie syntetyczne)
Table 1. Homogeneous housing community properties clustering with the share of Olsztyn Commune (synthetic deception)

Nr grupy Number of group	Liczba obiektów Quantity of objects	Dominujące stany zmiennych wejściowych w poszczególnych grupach obiektów Dominating features in particular groups of objects				
		X_{UG}	X_{LO}	X_{LS}	X_{ZT}	X_{BM}
1	118	niski low	zły/średni bad/medium	zły bad	zły/średni bad/medium	zły/średni bad/medium
2	63	niski low	bdb very good	zły/średni bad/medium	db/bdb good/very good	średni/db medium/good
3	46	niski low	średni medium	średni medium	średni/db medium/good	średni/db medium/good
4	103	niski/średni low/medium	zły bad	średni medium	średni/db medium/good	średni medium
5	81	niski low	bdb very good	zły/średni bad/medium	zły/średni bad/medium	zły/średni bad/medium
6	95	niski low	średni medium	średni medium	zły/średni bad/medium	zły bad
7	91	wysoki high	bdb very good	średni medium	średni/db medium/good	db good
8	83	niski/średni low/medium	średni/zły medium/bad	bdb very good	db good	średni/db medium/good

Źródło: Opracowanie własne na podstawie Kozickiego [2009]
Source: Authors' own study on the basis of Kozicki [2009]

większość udziałów. Jest to konsekwencją szerokiej wyprzedaży mieszkań komunalnych, zrealizowanej przez gminę prawie wyłącznie metodą prywatyzacji rozproszonej. Wprawdzie sprzedawano lokale w budynkach położonych w całym mieście, można jednak zauważyć, że proces prywatyzacji koncentrował się przede wszystkim na obiektach o dobrej lokalizacji (grupy: 2, 3 i 5) i/lub na obiektach, które były w złym stanie technicznym (grupy 1 i 6). Z kolei w budynkach o relatywnie dobrym stanie technicznym i starannym zagospodarowaniu terenu gmina zachowała w dalszym ciągu wysokie udziały (grupy: 4, 7 i 8). Stwierdzono także, że większość obiektów znajduje się albo w strefie centralnej miasta, albo w niewielkim oddaleniu od centrum. Badane obiekty charakteryzują się bardzo zróżnicowanymi walorami lokalizacji szczegółowej wykazującymi dodatnią korelację ze stanem zagospodarowania terenu. Należy ponadto podkreślić, że ogólny stan techniczny analizowanych nieruchomości wspólnot mieszkaniowych kształtuje się – poza wyjątkami (w grupie 7) – na poziomie średnim i poniżej niego. Stan ten nieznacznie się poprawia w obiektach młodszych – położonych dalej od centrum. Wskazuje to na pasywną gospodarkę remontową realizowaną we wspólnotach mieszkaniowych z udziałem gminy.

W ostatniej fazie badań przeprowadzono ocenę jakości grupowania z wykorzystaniem sieci Kohonena poprzez porównanie zmienności obiektów wewnątrz poszczególnych skupień (grup) ze zmiennością występującą po zastosowaniu innej niezależnej metody grupowania. Do porównania wybrano niehierarchiczną metodę *k*-średnich, natomiast jako miernik stopnia zmienności obiektów przyjęto wariancję wewnątrzgrupową. Analizując empiryczne wartości wariancji w obu metodach grupowania stwierdzono, że w każdej z badanych grup obiektów znacznie niższe wartości wariancji uzyskano w metodzie wykorzystującej sieć Kohonena niż w metodzie *k*-średnich. Wynika stąd, że grupy obiektów otrzymane w efekcie uczenia sieci są wewnętrznie bardziej jednorodne. Na podstawie ich analizy można ponadto zaobserwować bardziej wyraziste i użyteczne zależności z punktu widzenia gospodarki nieruchomościami.

PODSUMOWANIE I WNIOSKI

Specyficzne właściwości sieci Kohonena, konstruowanych na wzór sieci neuronowych w przyrodzie, umożliwiają ich zastosowanie w procedurach grupowania i klasyfikacji danych pochodzących z różnych źródeł. Sieci tego rodzaju są predysponowane do rozwiązywania złożonych zadań klasyfikacji, w tym szczególnie w przypadkach dużej liczby danych i braku prostych reguł klasyfikacji. Warunkiem prawidłowego uczenia sieci jest istnienie redundancji informacji w danych wejściowych, która służy sieciom do rozpoznania prawidłowości w ciągach uczących. Samoczynnie uformowane skupienia obiektów reprezentowane są przez neurony warstwy konkurencyjnej, natomiast uporządkowanie topologiczne osiągane jest przez zastosowanie koncepcji sąsiedztwa. Szczególna użyteczność metody Kohonena wyraża się w tworzeniu mapy topologicznej. Na podstawie wizualnej oceny stopnia podobieństwa skupień obiektów na mapie możliwe jest lepsze zrozumienie danych i ich pogłębiona analiza. Należy podkreślić, że podział badanego zbioru na grupy obiektów jednorodnych zastosowaną metodą nie jest jedynie aktem jednorazowym. Wytrenowana sieć jest bowiem przydatna do oceny nowych obiektów, dzięki

zdolności do uogólniania wiedzy nabytej w trakcie uczenia. Może zatem spełniać funkcje detektora sygnalizującego pojawianie się nowości.

Przeprowadzone analizy i badania pozwoliły na sformułowanie następujących wniosków:

1. Efekty zastosowania sieci Kohonena, jako narzędzia analizy skupień obiektów w zbiorze obiektów wielocechowych, dowodzą, że sieci tego typu mogą być z powodzeniem stosowane do rozwiązywania zadań grupowania i klasyfikacji nieruchomości wspólnot mieszkaniowych na potrzeby gospodarki nieruchomościami. Zastosowana metoda grupowania doprowadziła do jednoznacznego podziału badanego zbioru na rozłączne skupienia obiektów jak najbardziej podobnych do siebie pod względem przyjętych cech i jednocześnie maksymalnie różniących się od obiektów z innych skupień.

2. Analiza i ocena uzyskanych grup obiektów prowadzi do rozpoznania rzeczywistego zróżnicowania zbioru wspólnot mieszkaniowych z udziałem miasta Olsztyna, jak również do identyfikacji faktycznych zależności między obiektami i ich atrybutami. Na tej podstawie możliwe jest sformułowanie wniosków uogólniających, użytecznych do podejmowania decyzji w gospodarce nieruchomościami. Mogą one dotyczyć takich obszarów jak np.: polityka prywatyzacyjna, gospodarka remontowa i modernizacyjna, czy też przemieszczenia lokatorów mieszkań komunalnych wewnątrz zasobu.

3. Z punktu widzenia efektywnej i skutecznej gospodarki nieruchomościami komunalnymi badany zasób wspólnot mieszkaniowych nie może być traktowany jako zbiór jednorodny. Wyniki badań potwierdzają rozwarstwienie tego zasobu na osiem grup obiektów podobnych. Do cech determinujących wyniki grupowania zaliczono: udział własności gminy, lokalizację ogólną i szczegółową oraz stan techniczny obiektu.

4. W przeważającej liczbie wspólnot mieszkaniowych z udziałem miasta Olsztyna gmina utraciła większość udziałów. Wynika to z konsekwentnej sprzedaży mieszkań komunalnych metodą prywatyzacji rozproszonej. Proces prywatyzacji koncentrował się na ogół w obiektach o dobrej lokalizacji i znacznych potrzebach remontowych. Badane obiekty charakteryzują się bardzo zróżnicowanymi walorami lokalizacji oraz co najwyżej średnim stanem technicznym. Gospodarka remontowa wspólnot z udziałem gminy była pasywna.

5. Metoda grupowania wykorzystująca sieć Kohonena okazała się bardziej efektywna od metody *k*-średnich. Niższe wariancje wewnątrzgrupowe w metodzie Kohonena świadczą o tym, że sieć w efekcie samouczenia precyzyjniej wyznaczyła położenie centroidów w wielowymiarowej przestrzeni cech. Podział zbiorowości na skupienia tą metodą był ponadto bardziej przejrzysty i użyteczniejszy z punktu widzenia przyjętego celu grupowania.

PIŚMIENNICTWO

Kozicki M., 2009. Klasyfikacja wspólnot mieszkaniowych z udziałem gminy Olsztyn z zastosowaniem sztucznych sieci neuronowych (praca magisterska, Kat. Gospodarki Nieruchomościami i Rozwoju Regionalnego, Uniwersytet Warmińsko- Mazurski w Olsztynie, maszynopis).

Muczyński A., Kołek M., 2007. Modelowanie czynszu najmu lokali użytkowych gminy z zastosowaniem sztucznych sieci neuronowych. Acta Sci. Pol., Administratio Locorum 6(2), 55–67.

Muczyński A., Mrozik M., 2009. Klasyfikacja zasobów budynkowych gminy metodą analizy skupień. Studia i Materiały Towarzystwa Naukowego Nieruchomości 17(4), 23–35.
Osowski S., 1996. Sieci neuronowe. Oficyna Wydawnicza PW Warszawa.
StatSoft 2001. Wprowadzenie do sieci neuronowych. Statsoft Polska Sp. z o.o., Kraków, www.statsoft.pl.
Tadeusiewicz R., 1998. Elementarne wprowadzenie do techniki sieci neuronowych z przykładowymi programami. Akademicka Oficyna Wydawnicza PLJ Warszawa.
Witkowska D., 2002. Sztuczne sieci neuronowe i metody statystyczne. Wybrane zagadnienia finansowe. Wydawnictwo C.H. Beck Warszawa.

THE CLUSTERING OF HOUSING COMMUNITY PROPERTIES WITH THE USE OF SOFM (SELF-ORGANIZING FEATURE MAPS) METHOD

Abstract. The following paper presents the results of a housing community clustering research that has been carried out in Olsztyn city. The clustering procedure was realized with the use of the SOFM (self-organizing feature maps) method. The method required particular property features application. These involved: commune share in a housing community, location, land management, condition of the buildings. Having learned and verified the tool a two dimension topological map was made that organized the researched object into eight homogeneous groups. On the basis of the maps and classifying data detailed analysis and cluster synthesis of the given results has been made. The results show that in most of the housing community properties that the Olsztyn commune had a share in, moste of the communes' shares significantly decreased and the process of scattered privatization was concentrated among properties that had good location and a big need of renovation. The clustered objects differed in terms of their location (general and detailed) and at most medium technical condition. The learned tool is capable of object assessment that has not been presented during the process of learning. The adopted method occurred to be much more effective than the *k*-average method because if has given much more internally homogenous clusters of the objects. Moreover the given clusters were more transparent and useful in terms of the assumed aim of clustering.

Key words: clustering, housing community, SOFM (self-organizing feature maps), real estate management

Zaakceptowano do druku – Acctepted for print: 1.11.2009