Proces wyjaśniania ocen kondycji przedsiębiorstw w układach potencjał-ryzyko – studium eksperymentalne

Krzysztof Pancerz[1,2], Aleksandra Tu-Van[1], Norbert Tu-Van[1]
[1]Wyższa Szkoła Informatyki i Zarządzania w Rzeszowie
[2]Wyższa Szkoła Zarządzania i Administracji w Zamościu

Streszczenie: *Nowe podejście do prognozowania kondycji przedsiębiorstw wykorzystujące metodologię Case-Based Reasoning (CBR) zostało przedstawione w pracach J. Andreasika. Podejście oparte o zaproponowaną oryginalną ontologię przedsiębiorstwa koncentruje się na ocenie przedsiębiorstwa w zakresie wzrostu potencjału i ograniczaniu ryzyka działalności. Agregacja ocen ekspertów prowadzi do wyznaczenia pozycji przedsiębiorstwa w pięciu układach potencjał-ryzyko. Celem niniejszej pracy jest przedstawienie procesu wyjaśniania ocen kondycji przedsiębiorstw w układach potencjał-ryzyko. Proces wyjaśniania składa się z dwóch etapów. W pierwszym z etapów, w każdym z układów, dokonywane jest grupowanie przypadków (pozycji przedsiębiorstw) w odpowiednie klastry. W drugim etapie, dla każdego z układów, generowane są reguły wyjaśniające wpływ dokonanych ocen przez ekspertów na znalezienie się przedsiębiorstwa w określonym klastrze otrzymanym w pierwszym etapie. Do realizacji klasteryzacji i generowania reguł wyjaśniających wykorzystane zostały algorytmy znane z literatury dotyczącej drążenia danych oraz odkrywania wiedzy z danych. Eksperymenty zostały przeprowadzone na danych o ponad dwustu przedsiębiorstwach sektora MŚP z regionu Południowo-Wschodniej Polski zebranych w ramach realizacji projektu EQUAL nr F0086 prowadzonego przez Wyższą Szkołę Zarządzania i Administracji w Zamościu.*

Tło problemu

Nowe podejście do prognozowania kondycji przedsiębiorstw wykorzystujące metodologię Case-Based Reasoning (CBR) [Aamodt, Plaza 1994] zostało przedstawione w pracach J. Andreasika [Andreasik 2006, 2007]. Podejście oparte o zaproponowaną oryginalną ontologię przedsiębiorstwa koncentruje się na ocenie przedsiębiorstwa w zakresie wzrostu potencjału i ograniczaniu ryzyka działalności. Każde przedsiębiorstwo (E) reprezentowane jest przez zestaw dziesięciu funkcji kryterialnych:

E=<PK, PI, PS, PL, PG, RK, RI, RS, RL, RG>,

gdzie:

PK – potencjał kapitałowy,
PI – potencjał inwestycyjny i innowacyjny,
PS – potencjał interesariuszy,
PL – potencjał rynkowy o zasięgu lokalnym,
PG – potencjał rynkowy o zasięgu globalnym,
RK – ryzyko kapitałowe,
RI – ryzyko działalności inwestycyjnej i innowacyjnej,
RS – ryzyko wzrostu wydatków dla interesariuszy,
RL – ryzyko kosztów działalności na rynku lokalnym,
RG – ryzyko kosztów działalności na rynku globalnym.

Wartość każdej funkcji kryterialnej obliczona jest zmodyfikowaną metodą MPCS oraz metodą EUCLID. Metoda MPCS [Rozenes et al. 2004] pozwala na uzyskanie hierarchicznej struktury każdego potencjału i ryzyka:

POZIOM 1: główne rodzaje odpowiednio potencjału i ryzyka,
POZIOM 2: składowe rodzajów odpowiednio potencjału i ryzyka,
POZIOM 3: oceny składowych odpowiednio potencjału i ryzyka w trzech wymiarach:

- ocena dotycząca identyfikacji podstawowych instrumentów (lub polityki) zastosowanych w przedsiębiorstwie do podnoszenia określonego rodzaju potencjału i ograniczania analizowanego rodzaju ryzyka,
- ocena dotycząca stosowania w przedsiębiorstwie typowych zasad zarządzania potencjałem i ryzykiem,

- ocena dotycząca tendencji w analizowanych rodzajach potencjału i ryzyka.

Wynikiem zastosowania metody MPCS jest obliczona wartość oceny dla każdej składowej potencjału i ryzyka na drugim poziomie układu hierarchicznego. Każda wartość oceny mieści się w przedziale [0,1]. Metoda EUCLID służy do obliczenia wartości dziesięciu funkcji kryterialnych dla każdego przedsiębiorstwa. Metoda ta określa pozycję danego przedsiębiorstwa w układzie potencjał-ryzyko (z punktu widzenia matematycznego pozycja danego przedsiębiorstwa reprezentowana jest przez punkt w przestrzeni dwuwymiarowej). Utworzonych zostaje pięć układów:

- potencjał kapitałowy – ryzyko kapitałowe (PK-RK),
- potencjał inwestycyjny i innowacyjny – ryzyko działalności inwestycyjnej i innowacyjnej (PI-RI),
- potencjał interesariuszy – ryzyko wzrostu wydatków dla interesariuszy (PS-RS),
- potencjał rynkowy o zasięgu lokalnym – ryzyko kosztów działalności na rynku lokalnym (PL-RL),
- potencjał rynkowy o zasięgu globalnym – ryzyko kosztów działalności na rynku globalnym (PG-RG).

Cel i zakres badań

Celem wykonanych badań jest przeprowadzenie procesu wyjaśniania ocen kondycji przedsiębiorstw w układach potencjał-ryzyko. Proces wyjaśniania składa się z dwóch etapów. W pierwszym z etapów, w każdym z układów, dokonywana jest klasteryzacja (grupowanie) przypadków (przedsiębiorstw) w odpowiednie klastry. W drugim etapie, dla każdego z układów, generowane są reguły wyjaśniające wpływ dokonanych ocen przez ekspertów na znalezienie się przedsiębiorstwa w określonym klastrze otrzymanym w pierwszym etapie. Do realizacji klasteryzacji wykorzystano jeden z algorytmów klasteryzacji generujący ostry podział przestrzeni przypadków, tj. podział, w którym dany przypadek należy dokładnie do jednego klastra. Generowanie reguł wyjaśniających oparte zostało o model drzewa decyzyjnego. Zaproponowany schemat postępowania nie ogranicza możliwości zastosowania innych algorytmów klasteryzacji oraz generowania reguł lub innych reprezentacji wiedzy. Eksperymenty zostały przeprowadzone na danych o ponad dwustu przedsiębiorstwach sektora MŚP z regionu Południowo-Wschodniej Polski zebranych w ramach realizacji projektu EQUAL nr F0086 prowadzonego przez Wyższą Szkołę Zarządzania i Administracji w Zamościu (zob. www.e-barometr.pl).

Wykorzystane metodologie

Klasteryzacja danych

Modele prognozowania mogą być oparte o dwa podejścia klasyfikacyjne: klasyfikację wzorcową i klasyfikację bezwzorcową, które są odpowiednio utożsamiane z uczeniem nadzorowanym (ang. supervised learning) i uczeniem nienadzorowanym (ang. unsupervised learning). Z tą drugą klasyfikacją może być utożsamiana analiza skupisk (klastrów). Klasteryzacja przestrzeni przypadków jest jednym z istotnych problemów rozważanych w takich dziedzinach jak drążenie (eksploracja) danych (ang. Data Mining), odkrywanie wiedzy z baz danych (ang. Knowledge Discovery in Databases) oraz uczenie maszynowe (ang. Machine Learning) [Cios et al. 2007]. Proces klasteryzacji realizowany jest przy użyciu różnorodnych teorii i technik z szeroko pojętej sztucznej inteligencji.

Podstawowym zadaniem klasteryzacji (grupowania) jest dokonanie podziału zbioru przypadków C znajdujących się w bazie na grupy C1, C2, ..., Ck, nazywane klastrami, stanowiące podzbiory przypadków podobnych do siebie, przy czym pojęcie podobieństwa może być definiowane w różny sposób [Rutkowski 2005]. Podział zbioru C powinien być dokonany w taki sposób, aby przypadki z danej grupy były bardziej podobne do siebie niż do jakichkolwiek przypadków z pozostałych grup. Istotnym zagadnieniem jest ustalenie liczby k grup, na które zbiór przypadków ma zostać podzielony, gdyż zazwyczaj liczba ta nie jest z góry zadana. Jak do tej pory nie opracowano jednego uniwersalnego algorytmu klasteryzacji sprawdzającego się dla każdych danych. Kryteria klasteryzacji dotyczą interpretacji semantycznej klastrów. Istotna jest odpowiedź na pytanie dlaczego dwa przypadki przypisywane są do tego samego klastra. W tej kwestii odpowiedź może być udzielona na podstawie dostępnej wiedzy. W wielu sytuacjach przypadki grupowane są razem ze względu na istniejące pomiędzy nimi zależności takie jak np. noeodróżnialność, podobieństwo, bliskość, funkcjonalność, zgodność. Istotnym zagadnieniem w procesie klasteryzacji zbioru przypadków jest ustalenie struktury klastrów. Klastry mogą być parami

rozłączne lub też zachodzące na siebie. W przypadku klastrów rozłącznych mówi się o tzw. podziale ostrym. W takiej sytuacji dany przypadek należy tylko do jednego klastra. W przypadku klastrów zachodzących na siebie mówi się o tzw. podziale rozmytym. Przy tym podziale dany przypadek może należeć do wielu klastrów. Dodatkowo określany jest stopień przynależności przypadku do danego klastra. Stopień ten ma wartość rozmytą z przedziału [0, 1]. Wynika z tego, że przypadek może należeć do grupy tylko w pewnym stopniu. Dla podziałów rozmytych możliwe są dwie sytuacje. W pierwszej z nich suma stopni przynależności danego przypadku do każdego z klastrów jest zawsze równa 1 (tzw. podział probabilistyczny). W drugiej sytuacji warunek sumowania się stopni przynależności do 1 nie obowiązuje (tzw. podział posybilistyczny). Z algorytmicznego punktu widzenia metody klasteryzacji służą do rozwiązania problemu, który można sprowadzić do następującego pytania: w jaki sposób umieścić dwa przypadki w tym samym klastrze. Istotną rzeczą jest opracowanie algorytmów efektywnego wyznaczania klastrów przy uwzględnieniu zadanego kryterium klasteryzacji. Jednym z podejść przy ustalaniu przynależności przypadków do klastrów jest minimalizacja pewnej funkcji celu. Podstawowym problemem w tym podejściu jest zdefiniowanie odpowiedniej postaci funkcji celu.

Drzewa decyzyjne

Metody uczenia się drzew decyzyjnych to najbardziej znane i najczęściej stosowane w praktyce algorytmy indukcji symbolicznej reprezentacji wiedzy z przykładów. Koncepcja reprezentacji sekwencji warunków wpływających na podejmowaną decyzję z wykorzystaniem struktury drzewa jest stosowana w wielu dziedzinach. Ponadto struktura drzewa jest dość czytelna dla człowieka. Drzewo decyzyjne składa się z korzenia, z którego wychodzą co najmniej dwie gałęzie do węzłów leżących na niższym poziomie. Z każdym węzłem związany jest test sprawdzający wartości atrybutu opisującego przykłady. Dla każdego z możliwych wyników testu odpowiadająca mu gałąź prowadzi do węzła na niższym poziomie drzewa. Węzłom, z których nie wychodzą żadne gałęzie (nazywane liśćmi) przypisane są odpowiednie klasy decyzyjne. Ścieżki prowadzące od korzenia do liścia drzewa reprezentują koniunkcję pewnych testów zdefiniowanych na wartościach atrybutów opisujących przykłady uczące. Drzewo decyzyjne może więc posłużyć do określenia zbioru reguł określających przydział przykładów do klas decyzyjnych. Każda ścieżka drzewa od korzenia do liścia odpowiada jednej regule.

Przykładowe wyniki eksperymentu

Rozdział ten zawiera opis poszczególnych etapów przeprowadzonego eksperymentu na danych o ponad dwustu przedsiębiorstwach sektora MŚP z regionu Południowo-Wschodniej Polski. Jako przykład wybrano dane w układzie potencjał kapitałowy – ryzyko kapitałowe (PK-RK). Dla tego układu, hierarchiczne struktury ocen potencjału i ryzyka mają postać:

Tab. 1. Struktura potencjału kapitałowego

Potencjał kapitałowy	
Rodzaj	Składowe
Kapitał własny	Kapitał właścicielski
	Zysk przedsiębiorstwa
	Inne kapitały
Kapitał obcy	Kapitał obcy
Kapitał obrotowy	Kapitał obrotowy
Kapitał amortyzacyjny	Kapitał amortyzacyjny
Kapitał organizacyjny	Systemy administracyjne
	Systemy logistyczne
	Systemy technicznego przygotowania produkcji
	Systemy kontroli jakości
	System marketingu

Tab. 2. Struktura ryzyka kapitałowego

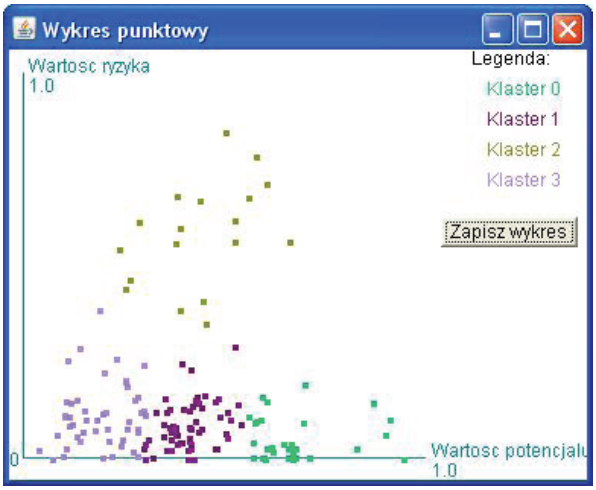
Ryzyko kapitałowe	
Rodzaj	Składowe
Ryzyko pozyskiwania kapitału własnego	Ryzyko pozyskiwania zysku netto
	Ryzyko podwyższania kapitału właścicielskiego
Ryzyko pozyskiwania kapitału obcego	Ryzyko kredytowe
	Ryzyko pozyskiwania kapitału z innych źródeł
Ryzyko zarządzania kapitałem obrotowym	Ryzyko zarządzania należnościami
	Ryzyko zarządzania zapasami
	Ryzyko zarządzania gotówką
	Ryzyko zarządzania zobowiązaniami
Ryzyko zarządzania kapitałem organizacyjnym	Ryzyko wdrażania systemów technicznego przygotowania produkcji
	Ryzyko wdrażania systemów zarządzania logistyką
	Ryzyko wdrażania systemów administracyjnych
	Ryzyko wdrażania systemów zapewnienia jakości
	Ryzyko wdrażania systemu marketingowego

Tablica przypadków dla algorytmu klasteryzacji zawiera wartości funkcji kryterialnych (potencjał kapitałowy oraz ryzyko kapitałowe) dla każdego ocenianego przedsiębiorstwa, wyznaczone za pomocą metod MPCS oraz EUCLID. Wartości te zawierają się w przedziale [0,1]. Fragment takiej tabeli pokazany jest poniżej.

Tab. 3. Tablica wejściowa przypadków (przedsiębiorstw) dla algorytmu klasteryzacji

Przedsiębiorstwo	Potencjał kapitałowy	Ryzyko kapitałowe
1	0.416165	0.235956
2	0.100704	0.118308
3	0.657005	0.031208
4	0.297355	0.031606
5	0.524269	0.565348
...	...	...

Z punktu widzenia matematycznego otrzymujemy punkty w przestrzeni dwuwymiarowej. Współrzędną x jest potencjał kapitałowy, zaś współrzędną y jest ryzyko kapitałowe. Jako algorytm klasteryzacji wybrano powszechnie używany algorytm HCM (ang. Hard C-Means) dokonujący podziału ostrego przestrzeni przypadków [Rutkowski 2005, Stąpor 2005]. Algorytm wykonano przy założeniu, że przypadki zostaną pogrupowane w cztery klastry. Do przeprowadzenia eksperymentu wykorzystano własną aplikację napisaną w języku Java. Wynik klasteryzacji w postaci graficznej ma postać:



Rys. 1. Wynik klasteryzacji przestrzeni przypadków w układzie potencjał kapitałowy – ryzyko kapitałowe

Wyniki klasteryzacji mogą zostać przedstawione w postaci tabelarycznej. Każdemu przedsiębiorstwu opisanemu w układzie potencjał kapitałowy – ryzyko kapitałowe (PK-RK) został przydzielony numer klastra, do którego to przedsiębiorstwo zostało zaliczone w wyniku klasteryzacji. Ponieważ został wykorzystany algorytm klasteryzacji generujący podział ostry, każdemu przedsiębiorstwu jest przypisany dokładnie jeden numer klastra. Fragment tablicy z przypisanymi numerami klastrów (K0, K1, K2, K3) pokazany jest poniżej.

Tab. 4. Tabelaryczny wynik klasteryzacji przestrzeni przypadków w układzie potencjał kapitałowy – ryzyko kapitałowe

Przedsiębiorstwo	Potencjał kapitałowy	Ryzyko kapitałowe	Numer klastra
1	0.416165	0.235956	K1
2	0.100704	0.118308	K3
3	0.657005	0.031208	K0
4	0.297355	0.031606	K1
5	0.524269	0.565348	K2
...	...	...	

Kolejnym etapem eksperymentu było wygenerowanie reguł wyjaśniających wpływ dokonanych ocen przez ekspertów na znalezienie się przedsiębiorstwa w określonym klastrze otrzymanym w pierwszym etapie. Poniżej przedstawiony będzie przykład, gdzie pod uwagę wzięto oceny związane z potencjałem kapitałowym przedsiębiorstw. Do wygenerowania reguł posłużono się modelem drzewa decyzyjnego. Do eksperymentu użyta została wersja demonstracyjna oprogramowania o nazwie See5 australijskiej firmy RuleQuest Research [zob. www.rulequest.com/see5-info.html]. Oprogramowanie to opiera się o algorytmy generowania drzew decyzyjnych zaproponowane przez J. R. Quinlana [Quinlan 1993]. Oprogramowanie See5 pozwala na wygenerowanie modelu drzewa decyzyjnego oraz reguł postaci JEŻELI-TO. Tablica decyzji dla algorytmu generowania drzewa zawiera 11 atrybutów warunkowych (odpowiadających składowym potencjału kapitałowego) oraz jeden atrybut decyzyjny określający numer klastra, do którego zakwalifikowano przedsiębiorstwo w wyniku algorytmu klasteryzacji. Fragment tej tablicy przedstawiony jest poniżej:

Tab. 5. Tablica decyzji dla algorytmu generowania modelu drzewa decyzyjnego

Przedsiębiorstwo	Kapitał właścicielski	Zysk przedsiębiorstwa	Inne kapitały	Kapitał obcy	Kapitał obrotowy	Kapitał amortyzacyjny	Systemy administracyjne	Systemy logistyczne	Systemy technicznego przygotowania produkcji	Systemy kontroli jakości	System marketingu	Numer klastra
1	0,5	0,5	0,9	0	0,8	0,5	0,5	0,3	0	0	0,3	K1
2	0,7	0,7	0,6	0,7	0,8	0,5	0,6	0,7	0,7	0,8	0,8	K3
3	0,8	0,8	0	0	0,5	0	0,4	0,3	0	0	0,3	K0
4	0,9	0,999	0,9	0,6	0,6	0	0,5	0	0	0	0,8	K1
5	0,5	0,1	0,1	0	0,4	0,5	0,5	0,6	0,9	0,1	0,1	K2
...	...	...	...	...	...	...	...	...	...	...	...	...

Fragment wygenerowanego modelu drzewa decyzyjnego za pomocą programu See5 ma postać:

```
Decision tree:

kapitał właścicielski > 0.78:
:...inne kapitały > 0.4:
:   :...systemy administracyjne <= 0.7: K1 (9/3)
:   :   systemy administracyjne > 0.7: K2 (3/1)
:   inne kapitały <= 0.4:
:   :...systemy tech przyg prod > 0.8: K2 (3/1)
:       systemy tech przyg prod <= 0.8:
:       :...kapitał właścicielski > 0.9: K0 (13/1)
:           kapitał właścicielski <= 0.9:
:           :...zysk przedsiębiorstwa > 0.6:
:               :...systemy tech przyg prod > 0.6: K1 (4/2)
:               :   systemy tech przyg prod <= 0.6:
:               :   :...systemy kontroli jakości <= 0.1: K0 (7/2)
:               :       systemy kontroli jakości > 0.1: K3 (8/1)
:               zysk przedsiębiorstwa <= 0.6:
:               :...systemy logistyczne > 0.5:
:                   :...systemy administracyjne <= 0.54: K2 (3)
:                   :   systemy administracyjne > 0.54: K0 (2)
:                   systemy logistyczne <= 0.5:
:                   :...zysk przedsiębiorstwa <= 0.3: K0 (4)
:                       zysk przedsiębiorstwa > 0.3:
:                       :...systemy logistyczne <= 0.3: K1 (2)
:                           systemy logistyczne > 0.3: K0 (3)
```

Z powyższego fragmentu drzewa widać, że pierwszym atrybutem, który jest brany pod uwagę do zakwalifikowania przedsiębiorstwa do odpowiedniej klasy jest Kapitał właścicielski. Dla tego atrybutu podział dokonywany jest względem wartości 0,78. W drugiej kolejności brany jest pod uwagę atrybut Inne kapitały. Dla tego atrybutu podział dokonywany jest względem wartości 0,4. Jeżeli wartość atrybutu Inne kapitały jest większa od 0,4, to trzecim testowanym atrybutem jest atrybut Systemy administracyjne. W tym przypadku podział dokonywany jest względem wartości 0,7 tego atrybutu. Po przetestowaniu tego atrybutu zostały w drzewie utworzone węzły – liście z przypisanymi klasami decyzyjnymi (numerami klastrów). Dla każdego liścia (węzła etykietowanego numerem klastra) podane są wartości: n lub n/m. Wartość n oznacza liczbę przypadków spełniających wszystkie warunki leżące na ścieżce prowadzącej od korzenia do danego liścia, dla których numer klastra zgadza się z numerem klastra etykietującego dany liść (tj. liczbę przypadków prawidłowo zaklasyfikowanych przez dany liść). Wartość m, jeśli występuje, oznacza liczbę przypadków spełniających wszystkie warunki leżące na ścieżce prowadzącej od korzenia do danego liścia, dla których numer klastra nie zgadza się z numerem klastra etykietującego dany liść (tj. liczbę przypadków nieprawidłowo zaklasyfikowanych przez dany liść). Wobec powyższego, przykładowo, dla następującej ścieżki: wartość atrybutu Kapitał właścicielski większa od 0,78 oraz wartość atrybutu Inne kapitały większa od 0,4 oraz wartość atrybutu Systemy administracyjne mniejsza lub równa 0,7 otrzymujemy, że przedsiębiorstwo spełniające takie warunki powinno zostać zakwalifikowane do klastra K1, bowiem dziewięć przypadków w tablicy decyzji, dla których wymienione atrybuty spełniały podane warunki należało do klastra K1, a trzy przypadki, dla których wymienione atrybuty spełniały podane warunki należało do innych klastrów. Analogicznie należy rozumieć pozostałe fragmenty drzewa.

Na podstawie modelu drzewa decyzyjnego można wygenerować reguły decyzyjne. Fragment wygenerowane zbioru reguł przy użyciu programu See5 ma postać:

```
Rules:

Rule 1: (13/1, lift 3.9)
   kapitał właścicielski > 0.9
   systemy tech przyg prod <= 0.8
   ->  class K0  [0.867]

Rule 2: (4, lift 3.8)
   kapitał właścicielski > 0.78
   zysk przedsiębiorstwa <= 0.6
   inne kapitały <= 0.4
   systemy logistyczne > 0.3
   systemy logistyczne <= 0.5
   ->  class K0  [0.833]

Rule 3: (2, lift 3.4)
   kapitał obrotowy > 0.1
   kapitał obrotowy <= 0.2
   systemy administracyjne <= 0.1
   ->  class K0  [0.750]
```

Dla każdej reguły wyznaczony zostaje tzw. współczynnik dokładności reguły (ang. accuracy factor) nazywany także współczynnikiem zaufania do reguły lub współczynnikiem pewności reguły. W programie See5 współczynnik dokładności reguły wyznaczany jest jako (n-m+1)/(n+2), gdzie n i m mają znaczenie takie jak podano wcześniej. Wartość tego współczynnika mieści się w przedziale [0,1]. Im większa wartość współczynnika, tym reguła jest dokładniejsza. Dla wartości równej 1 otrzymujemy regułę dokładną. Oznacza to, że wszystkie przypadki spełniające warunki w poprzedniku reguły (JEŻELI) spełniają także następnik reguły (TO).

Przykładowo, regułę Rule 1 możemy przedstawić w zapisie JEŻELI-TO następująco:

JEŻELI (kapitał właścicielski > 0,9) i (systemy tech przyg prod <= 0,8) TO class K0

W poprzednikach reguł może występować koniunkcja odpowiednich warunków. Przedstawiona reguła dostarcza następującego wyjaśnienia przynależności danego przedsiębiorstwa do klastra K0: jeśli ocena kapitału właścicielskiego danego przedsiębiorstwa będzie większa od 0,9 i jednocześnie ocena systemów technicznego przygotowania produkcji będzie mniejsza lub równa 0,8, to przedsiębiorstwo będzie klasyfikowane do klastra K0. Współczynnik dokładności dla takiej reguły wynosi 0,867. Nie jest to więc reguła w pełni dokładna, ale stopień jej dokładności jest duży.

Analogicznie można przeprowadzić eksperymenty dla pozostałych układów potencjał – ryzyko biorąc pod uwagę wszystkie oceny występujące w hierarchicznych strukturach ocen odpowiednich rodzajów potencjału i ryzyka. Otrzymane zbiory reguł pozwalają wyjaśnić wpływ dokonanych ocen przez ekspertów na znalezienie się przedsiębiorstwa w określonym klastrze. Istotną rzeczą jest ekonomiczna interpretacja otrzymanych w wyniku klasteryzacji klastrów przedsiębiorstw. Przedstawiony powyżej proces pozwala na klasyfikację nowych przedsiębiorstw. Znając poszczególne składowe oceny danego nowego przedsiębiorstwa dokonane przez ekspertów, na podstawie wygenerowanego drzewa możemy zakwalifikować to przedsiębiorstwo do odpowiedniego klastra. Taka klasyfikacja pozwala na przypuszczenie, że nowe przedsiębiorstwo będzie charakteryzowało się cechami (np. w przyszłości) podobnymi do cech tych przedsiębiorstw, które na etapie klasteryzacji zostały zakwalifikowane do danego klastra.

Podsumowanie

W pracy przedstawiono opis prostego procesu wyjaśniania ocen kondycji przedsiębiorstw w układach potencjał-ryzyko. W procesie tym wykorzystano znane z literatury algorytmy klasteryzacji oraz generowania drzew decyzyjnych. Jednak w dalszych pracach w ramach tej tematyki przewidziane jest zaproponowanie własnych metod wyjaśniania opartych o nowe algorytmy. Przedstawiony proces różni się od procesu definiowania klas przedsiębiorstw przedstawionego w pracach J. Andreasika, gdzie w układach potencjał – ryzyko określano cztery obszary (wysoki potencjał i wysokie ryzyko, niski potencjał i wysokie ryzyko, niski potencjał i niskie ryzyko, wysoki potencjał i niskie ryzyko). Obszary te wyznaczane były poprzez odpowiednie średnie wartości potencjału i wartości ryzyka. Przy zastosowaniu algorytmów klasteryzacji możliwe jest ustalenie różnej liczby klas przedsiębiorstw (klasy te wyznaczane są przez klastry). W przedstawionym eksperymencie założono także cztery klasy przedsiębiorstw, ale układ tych klas w przestrzeni różni się od układu obszarów wyznaczanych przez średnie.

Literatura

Aamodt, A., Plaza, E. 1994: *Case-based Reasoning: Foundational Issues, Methodological Variations, and System Approach*. AI Communications, vol. 7, no. 1, pp. 39–59.

Andreasik, J. 2006: *System oceny kondycji przedsiębiorstwa z wykorzystaniem metod wielokryterialnego podejmowania decyzji*. Barometr Regionalny, nr 6, Wyższa Szkoła Zarządzania i Administracji w Zamościu.

Andreasik, J. 2007: *A Case-Base Reasoning System for Predicting the Economic Situation on Enterprises – Tacit Knowledge Capture Process (Externalization)*. [w:] M. Kurzynski, E. Puchala, M. Wozniak, A. Zolnierek (Eds.): *Computer Recognition Systems 2*. Advances in Soft Computing, vol. 45, pp. 718-730, Springer-Verlag, Berlin Heidelberg.

Cios, K. J., Pedrycz, W., Swiniarski, R. W., Kurgan, L. A. 2007: *Data Mining. A Knowledge Discovery Approach*. Springer-Verlag, Berlin Heidelberg.

Quinlan, J. R. 1993: *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo.

Rozenes, S., Vitner, G., Spraggett, S. 2004: *MPCS: Multidimensional Project Control System*. International Journal of Project Management, vol. 22, pp. 109–118.

Rutkowski, L. 2005: Metody i techniki sztucznej inteligencji. Wyd. Naukowe PWN, Warszawa.

Stąpor, K. 2005: *Automatyczna klasyfikacja obiektów*. Akademicka Oficyna Wydawnicza EXIT, Warszawa.

www.e-barometr.pl

www.rulequest.com/see5-info.html