

Dr. hab. Zbigniew Raś, Prof. PJWSTK  
Polsko-Japońska Wyższa Szkoła Technik Komputerowych  
ul. Koszykowa 86, 02-008 Warszawa

Warszawa, 2009-04-30

### **Recenzja**

Rozprawy doktorskiej mgra Piotra Artiemjewa pt.

*„O Strategiach Granulacji Wiedzy i Zastosowaniach do Systemów Decyzyjnych”*

#### **Wstęp:**

Przedstawiona rozprawa doktorska składa się ze wstępu i trzech rozdziałów o następujących tytułach: (1) Wprowadzenie, (2) W poszukiwaniu logiki wewnętrznej systemu decyzyjnego – klasyfikatory granularne, (3) Aneks – Studium wpływu wielkości relatywnej klas decyzyjnych na jakość klasyfikatorów granularnych. W bibliografii podanych jest 91 prac, w tym 14 autorstwa albo współautorstwa Pana Artiemjewa.

Rozprawa ma charakter teoretyczno-praktyczny i dotyczy pewnej klasy metod aproksymacji w systemach decyzyjnych zaproponowanych wiele lat temu przez Z. Pawlaka. Wyniki przedstawione przez autora opisują jakość różnych strategii granulacji z punktu widzenia redukcji reguł przy równoczesnej minimalnej utracie wiedzy w nich zawartej.

#### **Problematyka:**

Niniejsza rozprawa lokuje się tematycznie w obszarze teorii zbiorów przybliżonych i granulacji wiedzy. Mechanizm granulacji wiedzy przyjęty w niniejszej pracy, będący pewną modyfikacją granulacji wiedzy podaną wcześniej przez L. Polkowskiego, przyjmuje inkluzję przybliżoną jako jej podstawowy składnik i definiuje granule wiedzy jako klasy mereologiczne tej inkluzji. W procedurach granulacji wiedzy wykorzystane są pewne relacje podobieństwa między obiektami.

Teoria granulacji została zapoczątkowana przez L. Zadeha. W dziedzinie teorii zbiorów przybliżonych zaproponował ją T.Y Lin na konferencji ISMIS w Charlotte ponad 20 lat temu. Pokazał on również pewne podobieństwo między granulami a otoczeniami topologicznymi. Od tego czasu tematyka ta jest dynamicznie rozwijana i zapoczątkowała badania w dziedzinie teorii obliczeń granularnych znajdującą silne zastosowania w systemach inteligentnych oraz w wydobywaniu wiedzy z baz danych.

### **Cel Rozprawy:**

Głównym celem rozprawy jest przedstawienie szeregu metod granulacji obiektów w systemach decyzyjnych prowadzących do sporej redukcji reguł wydobywanych standardowymi metodami takimi jak LERS przy równoczesnej minimalnej utracie ukrytej w nich wiedzy.

Strategie granulacji przedstawione w tej rozprawie to: strategia standardowa, strategia relatywna do klas decyzyjnych (concept-dependent), strategia hierarchiczna (layered granulation), strategia hybrydowa (połączenie strategii concept dependent ze strategią hierarchiczną), oraz tak zwane strategie  $\epsilon$ -we. Przy strategii „concept-dependent” wewnątrz granul pokrycia następuje głosowanie obiektów przez „majority voting” z założeniem że konflikty rozwiązywane są losowo. System decyzyjny granularny zbudowany jest tu z nowo stworzonych granularnych reprezentantów. Ważną cechą granulacji jest to, że daje ona zbiory danych zredukowane co do liczby obiektów oraz liczby reguł z nich indukowanych.

Wyniki opisane w pierwszym rozdziale pokazują że granulacja oparta o naturalne inkluzje i „strategię relatywną do klas decyzyjnych” oraz „strategię hierarchiczną”, prowadzi do struktur granularnych które w dużym stopniu zachowują wiedzę ukrytą w oryginalnych zbiorach danych. Pokazane zostało że przy optymalnych wynikach „accuracy” dostaje się duży procent redukcji ilości reguł.

W drugim rozdziale rozważane są struktury granularne otrzymane za pomocą inkluzji przybliżonych opisanych wcześniej w rozdziale pierwszym. Zakłada się tutaj że każdy zbiór posiada swoistą logikę wewnętrzną. Celem autora jest ich aproksymacja z wykorzystaniem w tym celu struktur granularnych indukowanych przez pewne rodzaje logik reprezentowanych

przez inkluzje przybliżone. Użyte są standardowe miary zaufania i pokrycia klasyfikatorów do weryfikacji jakości aproksymacji.

Rozważane są tu trzy struktury granularne:

[1] Pierwsza z nich polega na wyborze wartości parametru  $\varepsilon$  ze zbioru  $[0,1]$  oraz utworzenia dla danego środka  $u$  granuli  $g_\varepsilon(u) = \{v: (\forall a \in A) [ |a(v) - a(u) | < \varepsilon ]\}$  z normalizacją outlierów.

[2] Druga metoda jest relatywizacją metody pierwszej i definiuje granulę zakładając obok parametru  $\varepsilon$  dodatkowy parametr  $r_{catch}$  nazwany „promieniem chwytania”. Granula dla danego środka  $u$  jest tu zdefiniowana jako zbiór  $g_{\varepsilon, r_{catch}}(u) = \{v: \text{card}\{a \in A: |a(v) - a(u) | < \varepsilon\} / \text{card}(A) \geq r_{catch}\}$ . Obydwie metody granulacji są tu stosowane dla obiektów treningowych, dla reguł wyliczonych z obiektów treningowych, dla granularnych refleksji zbioru treningowego, oraz dla reguł wyliczonych z tych granularnych refleksji.

[3] W trzeciej metodzie granulacja odbywa się na podstawie wag uzyskanych przez obiekty, reguły, i refleksje granularne przy czym obiekty z tej samej klasy decyzyjnej tworzą jedną granulę biorącą udział w głosowaniu nad wartością decyzji na danym obiekcie testowym. Wagi przypisywane są obiektom na podstawie wartości inkluzji przybliżonych indukowanych z residuów ciągłych t-norm.

W ostatnim rozdziale pracy (Aneks) przedstawiona jest bardzo ciekawa analiza jakości klasyfikatorów granularnych w przypadku zbiorów danych które nie są zrównoważone ze względu na wielkość klas decyzyjnych. Do tego celu został wybrany Australian Credit. Otrzymane wyniki pokazują tu dobrą jakość (zaufanie, wsparcie) klasyfikatorów granularnych.

Dla wszystkich strategii granulacji opisanych w niniejszej pracy autor przeprowadza wiele eksperymentów z użyciem danych publicznie dostępnych w repozytorium University of Irvine. Są one bardzo ciekawe i w dużym stopniu podwyższają całościową wartość pracy i przedstawionych w niej wyników. Szczególnie dotyczy to eksperymentów związanych z  $\varepsilon$ -granulacją.

### **Uwagi Redakcyjne i Merytoryczne:**

Praca napisana jest bardzo starannie i przedstawiony w niej podział na rozdziały i podrozdziały jest prawidłowy. Sugerowałbym jednak poprawić jej Wstęp w którym to powinien być jasno opisany problem którego dotyczy ta rozprawa, wytłumaczenie dlaczego jest on ciekawy i ważny, co zostało w tej tematyce już zrobione na świecie, jakich metod i narzędzi autor proponuje tu użyć aby problem ten przekonywująco i możliwie prosto rozwiązać. Podobna informacja powinna znaleźć się również na początku pierwszego i drugiego rozdziału.

Jest wiele publikacji dotyczących budowania klasyfikatorów przy zbiorach danych które nie są zrównoważone ze względu na wielkość klas decyzyjnych. Autor wymienia tu tylko pracę J. Stefanowskiego. Sugerowałbym dodanie do bibliografii kilku innych publikacji (na przykład wyników uzyskanych przez Nathalię Japkowicz). Ciekawe byłoby sprawdzenie czy metoda przesuwania granicy dzielącej klasy decyzyjne zaproponowana przez Japkowicz daje gorsze wyniki od metod balansowania opartych na granulacji danych.

Praca traci minimalnie na czytelności ze względu na to że „Definicje” i „Fakty” w niej przedstawione nie są wyróżnione w tekście rozprawy.

### **Podsumowanie:**

Opiniowania praca jest niewątpliwie bardzo ciekawa i wyniki w niej przedstawione są oryginalne. Uważam że spełnia ona wymagania zawarte w obowiązujących przepisach dotyczących rozpraw doktorskich. Wnoszę więc o dopuszczenie mgra Piotra Artiemjewa do publicznej obrony.

