

Recenzja
rozprawy doktorskiej mgra Piotra Artiemjewa
pt. "O strategiach granulacji wiedzy i zastosowaniach do systemów
decyzyjnych".

Wstęp

Przedstawiona do recenzji rozprawa składa się ze Wstępu, dwóch części, Aneksu, skorowidzu i bibliografii (91 pozycji, w tym 14 autorstwa bądź współautorstwa doktoranta). Objętość rozprawy - 326 stron.

Praca jest poświęcona zagadnieniom obliczeń granularnych i ich wykorzystaniu, przede wszystkim w procesach podejmowania decyzji, ale także w zadaniach odkrywania wiedzy w zbiorach danych. W konstrukcji pracy wyróżnić można dwie zasadnicze części:

1. wprowadzenie w problematykę systemów decyzyjnych Pawlaka i ich związków z pojęciami obliczeń granularnych – w tym zdefiniowanie podstawowych pojęć dziedziny wnioskowania aproksymatywnego oraz granulacji wiedzy w systemach decyzyjnych; wprowadzone pojęcia są zilustrowane przykładami obliczeń granularnych (część 1 rozprawy);
2. opracowanie metod granulacji zbiorów w oparciu o opisane w części pierwszej inkluzje przybliżone; w części tej autor przedstawia metody granulacji dla czterech przypadków: (1) dla obiektów treningowych, (2) dla reguł wyliczanych z obiektów treningowych, (3) dla granularnych refleksji zbioru treningowego, oraz (4) dla reguł wyliczonych z refleksji granularnych. W tej części autor wprowadza ponadto pewien wariant ε -granulacji, dla którego opracowuje metody obliczeniowe. Dla wszystkich opracowanych metod autor przeprowadza szereg eksperymentów, wskazujących zakres stosowalności opracowanych metod oraz ich jakość i przydatność w odniesieniu do innych znanych metod.

W części III (Aneks) autor dodatkowo prowadzi badania nad wpływem relatywnej wielkości klas decyzyjnych na jakość klasyfikatorów granularnych.

Praca ma charakter teoretyczno-praktyczny. Autor wychodzi od teorii granulacji obiektów w systemach decyzyjnych, weryfikując koncepcje teoretyczne licznymi eksperymentami, co pozwala praktycznie analizować możliwości zaproponowanych metod i porównać je z metodami znanymi z literatury oraz ocenić ich przydatność praktyczną.

Uwagi ogólne

Problematyka

Jednym z bardziej aktualnych kierunków badań w dziedzinie sztucznej inteligencji, a w szczególności dotyczących problemów reprezentacji i odkrywania wiedzy jest kierunek obliczeń granularnych. Nawiązanie do granulacji wiedzy zostało zaproponowane przez L. Zadeha i bierze początek w drugiej połowie lat 70-tych ubiegłego stulecia. Zadeh w swojej teorii zbiorów rozmytych wprowadza możliwość opisu pojęcia w sposób niepewny za pomocą rozmytej funkcji przynależności. Innym niezwykle ważnym narzędziem w zakresie niepewności i nieścisłości w reprezentacji wiedzy jest teoria zbiorów przybliżonych, zaproponowana przez Pawlaka na początku lat 80-tych ubiegłego stulecia. Teoria ta szybko znalazła i wciąż znajduje zastosowania w wielu dziedzinach nauki i techniki. Jest też wciąż dynamicznie rozwijana, wciąż też stawia nowe wyzwania badawcze. Opiniowana rozprawa leży w nurcie badań związanych z teorią zbiorów przybliżonych. W rozwoju teorii zbiorów przybliżonych okazało się celowym rozważenie relacji ogólniejszej od oryginalnej Pawlaka relacji nieodróżnialności.

Cel rozprawy

W rozprawie autor nawiązuje do mechanizmów granulacji wiedzy wprowadzonych przez L. Polkowskiego, wynikających z inkluzji przybliżonych. Mechanizmy te zostają przez autora zmodyfikowane, a następnie dla zdefiniowanych pojęć opracowane są nowe mechanizmy granulacji.

Głównym problemem badawczym, któremu poświęcona jest rozprawa mgra Artiemjewa są strategie granulacji wiedzy w systemach decyzyjnych w kontekście teorii zbiorów przybliżonych. Autor wychodzi z pięciu następujących strategii:

1. standardowej,
2. relatywnej do klas decyzyjnych
3. hierarchicznej
4. hybrydowej (tj. obejmującej (2)+(3) powyżej)
5. epsilonowej, którą definiuje w oparciu o zmodyfikowaną relację podobieństwa.

Dla strategii tych autor weryfikuje hipotezę, że granule złożone z obiektów podobnych powinny prowadzić do bardziej „oszczędnych” granularnych zbiorów danych, które pomimo zastosowanej redukcji zachowują informacje oryginalnych zbiorów danych. Weryfikacja ta jest przeprowadzona przez zaimplementowanie omawianych mechanizmów granulacji i wykonanie szeregu eksperymentów, przy czym za kryterium zachowania wiedzy przez zbiór granularny autor przyjmuje jakość klasyfikatora działającego w oparciu o zbiór granularny w porównaniu do klasyfikatora działającego w oparciu o oryginalny zbiór danych.

Na szczególną uwagę zasługuje udana próba oryginalnego rozszerzenia klasycznych algorytmów klasyfikacji. W szczególności, w celu poprawienia jakości klasyfikacji autor modyfikuje metodę ważenia obiektów treningowych poprzez zmodyfikowanie relacji podobieństwa. Wprowadzona jest mianowicie epsilonowa miara podobieństwa, która działa w ten sposób, iż zmniejszane są wagi małe, a zwiększane wagi duże, co prowadzi do lepszych wyników klasyfikacji.

Bardzo wysoko oceniam przeprowadzone w pracy eksperymenty. Ich celem było zweryfikowanie uzyskanych efektywności. W szczególności na uwagę zasługują eksperymenty dotyczące ε -granulacji (punkty II.4 i 5). Przeprowadzone badania eksperymentalne dostarczyły interesujących wyników pozwalających porównać proponowane algorytmy z rozwiązaniami znanymi z literatury (autor pokazuje to w części Konkluzje do części II).

Uwagi redakcyjne

Uwagi ogólne

Autor sformułował cele rozprawy w sposób jasny i poprawny. Praca jest napisana starannie. Układ pracy jest logiczny. W pracy zamieszczony jest skorowidz, jest on bardzo przydatny przy czytaniu rozprawy. Język rozprawy jest poprawny.

Moja uwaga krytyczna dotyczy przede wszystkim Wstępu, w którym moim zdaniem warto byłoby wyróżnić „tezę rozprawy”. Autor zamieszcza tam, co prawda, „przewodnik” po pracy oraz prezentuje cel pracy, jednak dodanie wyróżnionego paragrafu dotyczącego tezy zdecydowanie poprawiłoby czytelność tego fragmentu. Kolejna uwaga wiąże się z tym, że w niektórych miejscach niezbyt czytelnie autor rozgranicza swoje osiągnięcia od przytaczanych wyników innych autorów. W szczególności, celowe byłoby dodać (we Wstępie lub w Rozdziale 1) dyskusję z innymi podejściami i wyjaśnienia dotyczące własnych rozwiązań i motywacji. Pewne tego typu elementy znajdują się w konkluzjach, jest to jednak moim zdaniem niewystarczające.

Uwagi szczegółowe

Przy czytaniu nasunęło mi się kilka uwag o charakterze redakcyjnym:

1. Czytelność pracy uległa by istotnemu polepszeniu, gdyby autor wyróżniał w tekście definicje, własności, itp.
2. Autor stosuje nazwy rozmaitych parametrów w wersji angielskiej. W kilku miejscach autor skraca zdanie pomijając polskie „kwalifikatory”, co czyni tekst żargonowy i niezbyt elegancki. Np. na stronie 26 autor pisze:
„...liczenie *strength* reguły” zamiast „liczenie parametru *strength* reguły”.
Podobnie jest z „liczeniem *accuracy*”
3. Nie jest czytelne łączenie znaków greckich z polskimi końcówkami deklinacji (w szczególności zamiast pisać skrótowo „granulacje ϵ -we”, należałoby pisać „granulacje epsilonowe”, ewentualnie zdefiniować ϵ -granulacje).
4. Brakuje w pracy wykazu oznaczeń. Pojawienie się takiego wykazu znacznie ułatwiłoby zapoznanie się z prezentowanym materiałem.
5. Drobne usterki „interpunkcyjne” zaznaczyłem w tekście.

Końcowe uwagi merytoryczne

Praca ma oryginalny charakter. Moje uwagi mają przede wszystkim charakter redakcyjny. Wkład autora jest znaczący. Za najważniejsze osiągnięcia uważam

zmodyfikowane algorytmy klasyfikacji i uzyskane z ich pomocą wyniki, prezentowane w częściach II.4 i II.5.

Podsumowanie

Uważam, że opiniowana praca jest dobrym opracowaniem. Niewątpliwie spełnia ona wymagania zawarte w obowiązujących przepisach dotyczących rozpraw doktorskich. Wnoszę zatem o dopuszczenie mgra Piotra Artiemjewa do publicznej obrony.

H. [signature]