

prof. dr hab. inż. Marzena Kryszkiewicz
Politechnika Warszawska
Instytut Informatyki
ul. Nowowiejska 15/19
00-665 Warszawa
Marzena.Kryszkiewicz@pw.edu.pl

Warszawa, 5.11.2023

Recenzja rozprawy doktorskiej
mgr. Sebastiana Stawickiego

zatytułowanej

Ensembles of Classifiers Based on Decision Bireducts

1. Zagadnienie naukowe rozpatrzone w pracy

Opiniowana rozprawa wpisuje się w nurt prac poświęconych rozwojowi teorii zbiorów przybliżonych. W szczególności, w rozprawie przedstawiono związki pomiędzy własnościami pewnych wariantów reduktów decyzyjnych (będących pewnymi minimalnymi podziorami zadanego zbioru atrybutów warunkowych) i pewnych wariantów bireduktów decyzyjnych (które oprócz podziorów atrybutów niosą informację o powiązanych obiektach), sposoby dostosowania metod odkrywania reduktów decyzyjnych do odkrywania bireduktów decyzyjnych, sposoby tworzenia zespołów klasyfikatorów opartych na bireduktach decyzyjnych. Ponadto przedstawiono pojęcie bireduktu decyzyjnego w kontekście strumienia danych oraz sposób wyznaczania takiego bireduktu. Przedstawiono także opracowaną metodę istotności atrybutów, uzyskane wyniki eksperymentalne i przykładowe zastosowanie praktyczne.

Autor sformułował cel rozprawy jako spójne i przejrzyste zaprezentowanie kompletnego zestawienia wyników dotyczących bireduktów decyzyjnych. Ponadto wymienił jako główne wyniki rozprawy:

- W1. *Wprowadzenie pojęć: ε -bireduktu decyzyjnego oraz bireduktu γ -decyzyjnego, związanego z pojęciem obszaru pozytywnego z teorii zbiorów przybliżonych.*
- W2. *Wyrowadzenie funkcji rozróżnialności, której implikanty proste odpowiadają odpowiednio bireduktom decyzyjnym i bireduktom γ -decyzyjnym, poprzez analogię do funkcji rozróżnialności, której implikanty proste odpowiadają reduktom.*
- W3. *Porównanie bireduktów decyzyjnych ze standardowymi i przybliżonymi reduktami decyzyjnymi.*
- W4. *Wykazanie NP-trudności problemu wyszukiwaniu minimalnego ε -bireduktu decyzyjnego.*

- W5. *Zaproponowanie algorytmów i heurystyki, których celem jest efektywne wyszukiwanie bireduktów decyzyjnych oraz ich wariantów.*
- W6. *Wprowadzenie pojęcia bireduktu decyzyjnego dla strumienia danych.*
- W7. *Przedstawienie idei zespołów bireduktów decyzyjnych i zbadanie ich właściwości.*
- W8. *Wykazanie NP-trudności problemu optymalizacji poszukiwania najprostszego prawidłowego zespołu bireduktów decyzyjnych.*
- W9. *Eksperymentalne zweryfikowanie zaproponowanych metod na zestawach danych syntetycznych i referencyjnych.*
- W10. *Przedstawienie studium przypadku demonstrującego zastosowanie zespołów bireduktów do problemu decyzyjnego, który został zdefiniowany podczas realizacji projektu dla firmy HR specjalizującej się w rekrutacji ekspertów IT.*
- W11. *Zaprezentowanie metody oceny istotności atrybutów użytych w zespołach klasyfikatorów opartych na bireduktach decyzyjnych.*
- W12. *Przygotowanie biblioteki programistycznej scikit-rough (<https://github.com/sebov/scikit-rough>).*

Wyniki przedstawione w rozprawie obejmują osiągnięcia przedstawione w następujących współautorskich publikacjach mgr. S. Stawickiego (które zostały wykazane w bibliografii rozprawy):

- [51] Andrzej Janusz and **Sebastian Stawicki**. Applications of Approximate Reducts to the Feature Selection Problem. In Procs. of RSKT 2011
- [52] Andrzej Janusz, **Sebastian Stawicki**, Michał Drewniak, Krzysztof Ciebiera, Dominik Slezak, and Krzysztof Stencel. How to Match Jobs and Candidates - A Recruitment Support System Based on Feature Engineering and Advanced Analytics. In Procs. of IPMU 2018
- [53] Andrzej Janusz, **Sebastian Stawicki**, Marcin S. Szczuka, and Dominik Slezak. Rough Set Tools for Practical Data Exploration. In Procs. of RSKT 2015
- [59] Andrzej Janusz, Dominik Slezak, **Sebastian Stawicki**, and Krzysztof Stencel. A Practical Study of Methods for Deriving Insightful Attribute Importance Rankings Using Decision Bireducts. Information Sciences, 2023
- [117] **Sebastian Stawicki** and Sebastian Widz. Decision Bireducts and Approximate Decision Reducts: Comparison of Two Approaches to Attribute Subset Ensemble Construction. In Procs. of FedCSIS 2012
- [118] **Sebastian Stawicki** and Dominik Slezak. Recent Advances in Decision Bireducts: Complexity, Heuristics and Streams. In Procs. of RSKT 2013
- [119] **Sebastian Stawicki**, Dominik Slezak, Andrzej Janusz, and Sebastian Widz. Decision Bireducts and Decision Reducts – A Comparison. International Journal of Approximate Reasoning, 84:75–109, 2017
- [142] Dominik Slezak and **Sebastian Stawicki**. The Problem of Finding the Simplest Classifier Ensemble is NP-Hard - A Rough-Set-Inspired Formulation Based on Decision Bireducts. In Procs of IJCRS 2020

2. Struktura rozprawy

Praca liczy 184 strony i składa się z 8 rozdziałów, bibliografii oraz 2 dodatków. W rozdziale 1 przedstawiono cel i wyniki rozprawy, jej układ oraz wstępny 2-stronicowy przegląd literatury powiązanej z jej tematyką. Stosowne odwołania do literatury występują także we wszystkich pozostałych rozdziałach. W rozdziale 2 przedstawiono pojęcia wstępne, a w rozdziale 3

podstawowe pojęcia teorii zbiorów przybliżonych, w tym pojęcia reduktów decyzyjnych i przybliżonych reduktów decyzyjnych. W rozdziałach 4-7 zaprezentowano wyniki rozprawy, w tym: wyniki W1-W4 oraz W7-W8 zostały przedstawione w rozdziale 4, W5-W6 oraz W9 w rozdziale 5, W10 w rozdziale 6, a W11-W12 w rozdziale 7. W rozdziale 8 zawarto podsumowanie i plany przyszłych prac. W dodatku A przedstawiono wybrane wyniki dotyczące istotności atrybutów występujących w zbiorach danych używanych w eksperymentach. W dodatku B zaprezentowano przykłady użycia opracowanej biblioteki programistycznej.

W pracy zawarto odwołania do 144 publikacji, w tym 9 powstałych z udziałem Autora rozprawy, z czego 8 jest ściśle powiązanych z tematyką rozprawy.

3. Oryginalny wkład naukowy rozprawy

Za najbardziej oryginalny wkład naukowy rozprawy uważam:

- wyprowadzone własności bireduktów decyzyjnych, ε -bireduktów decyzyjnych, γ -decyzyjnych bireduktów oraz γ -decyzyjnych ε -bireduktów, w tym ich powiązania ze stosownymi wariantami reduktów decyzyjnych (p. rozdział 4),
- wprowadzenie pojęcia bireduktu decyzyjnego dla strumienia danych oraz zaproponowanie heurystycznej metody wyszukiwania takich bireduktów (p. rozdział 5).

Warto odnotować, że powyższe osiągnięcia zostały opublikowane w następujących pracach, których pierwszym autorem jest Autor rozprawy:

- [117] **Sebastian Stawicki** and Sebastian Widz. Decision Bireducts and Approximate Decision Reducts: Comparison of Two Approaches to Attribute Subset Ensemble Construction. In Procs, of FedCSIS 2012
- [118] **Sebastian Stawicki** and Dominik Slezak. Recent Advances in Decision Bireducts: Complexity, Heuristics and Streams. In Procs. of RSKT 2013
- [119] **Sebastian Stawicki**, Dominik Slezak, Andrzej Janusz, and Sebastian Widz. Decision Bireducts and Decision Reducts – A Comparison. International Journal of Approximate Reasoning, 84:75–109, 2017

Pozostałe teoretyczne wyniki rozprawy i algorytmy postrzegam jako zasadne, użyteczne, ale stanowiące dość oczywiste modyfikacje bądź rozszerzenia istniejących wcześniej rozwiązań. Uzyskane wyniki eksperymentalne postrzegam jako właściwe dopełnienie wyników teoretycznych.

4. Poprawność

Rozprawa jest zredagowana bardzo starannie. Jest napisana w języku angielskim, którym Autor posługuje się na ogół biegle i precyzyjnie. Układ pracy jest dobry, chociaż decyzja o omówieniu pojęcia bireduktu decyzyjnego w strumieniu danych w rozdziale 5, który jest poświęcony algorytmom i eksperymentom, nie wydaje się właściwym wyborem. Prezentacja znanych i proponowanych przez Autora rozwiązań oraz wywody na ich temat (w tym dowody twierdzeń, propozycji i konkluzji) są prowadzone w sposób logiczny, przekonujący i zrozumiały. Zaproponowane rozwiązania zostały dogłębnie i w sposób właściwy przeanalizowane na drodze analitycznej i częściowo zweryfikowane eksperymentalnie.

5. Uwagi do rozprawy

Pytania:

- Pojęcie bireduktów decyzyjnych zostało zaproponowane przez D. Ślęzaka i A. Janusza z myślą o wykorzystaniu ich jako zespół klasyfikatorów. W rozprawie zaproponowano jako modyfikację tych bireduktów pojęcia ε -bireduktów decyzyjnych, γ -decyzyjnych bireduktów oraz γ -decyzyjnych ε -bireduktów. Czy przeprowadzono eksperymentalną weryfikację użyteczności tych nowych wariantów bireduktów decyzyjnych jako zespołów klasyfikatorów? Jeśli tak, to czy okazało się, że klasyfikacja przy ich użyciu jest skuteczniejsza niż w przypadku zespołu klasyfikatorów opartych na bireduktach decyzyjnych? Jeśli nie, to w jakich zastosowaniach zaproponowane w rozprawie warianty bireduktów decyzyjnych mogłyby okazać się skuteczniejsze od standardowych bireduktów decyzyjnych?
- Rozprawa w dużym stopniu jest oparta na współautorskich publikacjach mgr. S. Stawickiego. Na czym polegał wkład Autora rozprawy w jego publikacjach, w których nie był pierwszym autorem?

Poniżej zamieszczam pomniejsze uwagi dotyczące zauważonych słabości/nieściłości:

- W tekście pojawiają się nieliczne literówki, np. na str. 21 jest: *a quotient set determined be the decision attribute*, a powinno być: *a quotient set determined by the decision attribute*.
- Niepoprawne sformułowanie *equals to* występuje kilku miejscach w rozprawie, np. na str. 26 jest: *value equals to 1*, a powinno być: *value equals 1* lub *value is equal to 1*.
- W rozprawie używane jest pojęcie *support of a rule*, ale nie wprowadzono jego definicji. Przypuszczam, że Autor rozprawy pisząc *support of a rule*, miał na myśli (wszystkie) obiekty wspierające regułę lub zbiór obiektów wspierających regułę. Jeśli tak, to chyba lepiej byłoby pisać np. *(all) objects supporting a rule*, albo *a set of objects supporting a rule* zamiast *support of a rule*. Moja uwaga wynika stąd, że *support of a rule* jest zwykle rozumiany jako liczba obiektów wspierających regułę.
- Na str. 55 jest: *Sort sequences ... in a descending order*. Chyba powinno być *Sort sequences ... in a non-ascending order*.
- Na str. 56 jest: $|C_{j1}|, |C_{j1}|, |C_{j1}|$, a powinno być: $|C_{j1}|, |C_{j2}|, |C_{j3}|$.
- W def. 42 na str. 78 pojawia się zapis: *is not true for neither... nor*. Przypuszczam, że miało być: *is true for neither... nor* lub *is not true for either... or*.

Powyższe uwagi nie wpływają na moją pozytywną ocenę osiągnięć naukowych Autora, które przedstawił w rozprawie.

6. Przydatność rozprawy

W rozprawie przedstawiono potencjalnie użyteczne rozszerzenia teorii zbiorów przybliżonych. Oprócz wkładu teoretycznego, przedstawiono zastosowanie praktyczne klasyfikatora opartego na bireduktach decyzyjnych, które wykorzystano podczas realizacji projektu dla firmy HR specjalizującej się w rekrutacji ekspertów IT. Opracowano też otwarte oprogramowanie <https://github.com/sebov/scikit-rough>.

7. Podsumowanie

Opiniowana rozprawa zawiera oryginalne rozwiązanie problemu naukowego. Jej Autor wykazał się posiadaniem ogólnej wiedzy teoretycznej w dyscyplinie Informatyka, a także umiejętnością twórczego, samodzielnego prowadzenia pracy naukowej. Na drodze teoretycznej Autor wyprowadził szereg własności zaproponowanych rozszerzeń teorii zbiorów przybliżonych. Ponadto, mgr Sebastian Stawicki jest współautorem 8 publikacji tematycznie powiązanych z rozprawą (w tym publikacji w *International Journal of Approximate Reasoning* oraz publikacji w *Information Sciences*).

Stwierdzam, że recenzowana rozprawa doktorska mgr. Sebastiana Stawickiego spełnia wymagania stawiane rozprawom doktorskim przez obowiązującą ustawę i wnoszę o dopuszczenie jej do publicznej obrony.

