



Politechnika Poznańska  
Wydział Informatyki

Streszczenie rozprawy doktorskiej

Nowe metody dezagregacji  
preferencji i analizy post factum  
z wykorzystaniem modelu wiedzy  
w postaci addytywnej funkcji  
wartości

**Krzysztof Ciomek**

Promotor: dr hab. inż. Miłosz Kadziński

Poznań, 2019

# Wprowadzenie

Przedmiotem analizy w ramach problemów decyzyjnych jest zazwyczaj zbiór wariantów ocenionych na zbiorze kryteriów. W przypadku istnienia jednego kryterium w całości opisującego jakość analizowanych wariantów, rozwiązanie problemu jest trywialne. W przypadku istnienia wielu konfliktowych kryteriów, problemy decyzyjne stają się złożone. W ich analizie i rozwiązywaniu pomagają komputerowe metody wielokryterialnego wspomaganie decyzji.

Wykorzystywane są one m.in. w ekonomii, ochronie zdrowia, zarządzaniu czy ochronie środowiska. Przykładem ich zastosowania może być klasyfikacja ryzyka kredytowego na podstawie charakterystyki klienta, konstrukcja rankingu leków przeciwwązkowych w zapobieganiu udarów na podstawie danych klinicznych czy ocena budynków mieszkalnych pod kątem jakości powietrza przy wzięciu pod uwagę ich cech oraz stężenia biologicznych i cząsteczkowych zanieczyszczeń.

Wyróżnia się trzy podstawowe klasy problemów decyzyjnych: porządkowanie wariantów od najlepszego do najgorszego (ranking), klasyfikację do znanych kategorii oraz wybór podzbioru najlepszych wariantów. Celem metod wielokryterialnego wspomaganie decyzji jest rozwiązywanie takich problemów poprzez wypracowywanie rekomendacji możliwie najbliższych preferencjom decydenta. Te modeluje się przy pomocy narzędzi matematycznych w oparciu o informację preferencyjną, którą mogą stanowić przykłady decyzji czy

subiektywna ocena istotności kryterium lub inne wymaganie dotyczące problemu decyzyjnego takie jak np. oczekiwana liczność podzbioru najlepszych wariantów, które mają zostać wybrane.

Metody komputerowego wspomaganie decyzji różnią się między sobą typem akceptowanej informacji, rodzajem modelu preferencji oraz charakterystyką zwracanych wyników. Obsługują deterministyczne lub stochastyczne oceny na kryteriach, współpracują z precyzyjną lub nieprecyzyjną informacją preferencyjną oraz dostarczają rekomendacje w formie probabilistycznej lub bezpośredniej. Preferencje są modelowane z wykorzystaniem funkcji, relacji lub reguł. Różne metody oferują różne właściwości i dlatego powinny być odpowiednio dobierane w zależności od natury i wymagań konkretnego problemu decyzyjnego.

Badania opisane w niniejszej rozprawie skupiły się na adytywnej funkcji wartości jako modelu reprezentacji wiedzy o preferencjach. Model ten wyrażony jest jako suma nieujemnych wartości monotonicznych funkcji cząstkowych – po jednej dla każdego kryterium. W ten sposób oceny z wielu kryteriów, wykorzystujących potencjalnie różne skale, agregowane są do pojedynczej wartości globalnej.

Motywacją do przeprowadzenia badań były następujące obserwacje. Dostarczana przez decydenta informacja preferencyjna nie zawsze przybliża do rozwiązania problemu. Może to prowadzić do niepotrzebnej złożoności modelu oraz wydłużyć proces wypracowywania rekomendacji, a w praktyce potencjalnie oznaczać zwiększenie kosztów analizy. Ponadto sama rekomendacja może być niewystarczająca, aby podjąć decyzję. Użytkownik może wymagać uzasadnienia wyniku lub określenia wielkości jego marginesu poprawności. Czasami mała modyfikacja oceny wariantu decyzyjnego może zmienić rekomendację. Decydent powinien być świadomy

takich sytuacji. Dodatkowym usprawnieniem zwiększającym użyteczność metod wspomaganie decyzji z punktu widzenia użytkownika może być umożliwienie obserwacji bezpośrednich konsekwencji wprowadzania informacji do modelu. W ramach pracy badawczej zaproponowano szereg rozwiązań odnoszących się do powyższych spostrzeżeń. Kolejne rozdziały przybliżają rozważane zagadnienia i proponowane metody.

# Wielokryterialne sortowanie na podstawie pośredniej informacji preferencyjnej

W wielokryterialnym wspomaganiu decyzji sortowaniem nazywa się zadanie przydziału wariantów decyzyjnych do znanych klas uporządkowanych względem preferencji. Wykorzystując addytywną funkcję wartości jako model reprezentacji wiedzy, taka operacja może przebiegać dwojako. Popularniejszym podejściem jest porównywanie globalnej wartości wariantu z rozdzielającymi kolejne klasy progami, które stają się parametrami modelu lub są zadane z góry. Alternatywnie globalne wartości klasyfikowanych wariantów decyzyjnych można porównywać z wartościami przykładowych wariantów dostarczonych przez użytkownika, reprezentujących granice klas.

Dla tak zdefiniowanego problemu zaproponowano metodę modelowania preferencji, która integruje dwa znane typy informacji preferencyjnej: przykładowe przydziały i oczekiwane licznosci klas oraz nowo wprowadzone i sformalizowane porównania par wariantów w odniesieniu do przydziałów (np. jeden wariant musi być w klasie nie gorszej niż inny). Wszystkie trzy typy mogą mieć formę precyzyjną lub nieprecyzyjną. Równocześnie metoda zapewnia trzy rodzaje wyników odpornej regresji porządkowej, które bezpośrednio korespondują z tymi typami: przydziały do klas, minimalne i maksy-

malne możliwe licznosci klas oraz relacje preferencji pomiedzy parami wariantow w odniesieniu do przydzialow (np. jeden wariant jest zawsze w klasie wyzszej niz inny). Przydzialy i relacje mozna rozwazac w sensie mozliwym i koniecznym, tj. gdy sa wspierane przez co najmniej jeden lub wszystkie modele spójne z preferencjami decydenta. Uzyskanie wynikow wymaga rozwiązania problemow programowania calkowitoliczbowego.

Dostarczana informacja preferencyjna wprowadza ograniczenia na dozwolone parametry modelu preferencji. To przeklada sie na redukcje przestrzeni zawierajacej instancje modelu spójne z preferencjami decydenta i w konsekwencji na zbieznosc wynikow. Przedzialy mozliwych przydzialow zawezajaja sie, liczba par wariantow, co do ktorych konieczna relacja preferencji w odniesieniu do przydzialow jest ustalona, rośnie, a minimalne i maksymalne licznosci klas zbiegaja ku sobie.

Modelowanie preferencji jest istota wielokryterialnego wspomaganie decyzji. Zwiększenie mozliwosci wyrazania swoich przekonań przez decydentow bezposrednio wplywa na uzytecznosc metod. Ponadto jezeli proces uzyskiwania informacji preferencyjnej ma forme dialogu z uzytkownikiem to korespondencja pomiedzy wejściem a wyjściem metody umozliwia decydentowi obserwacje konsekwencji dostarczanej przez niego informacji preferencyjnej. To powinno ulatwiac zrozumienie samego procesu wypracowywania rekomendacji.

Jednak czasami interakcja z decydentem nie jest mozliwa, a model budowany jest na podstawie istniejacego zbioru informacji preferencyjnej. W pracy przedstawiono sposob postepowania w takim przypadku na przykladzie klasyfikacji protokolow syntezy nanocząstek srebra do pieciu klas zwiazanych z jakością i spełnieniem norm zrównowazonego rozwoju (od *wyjatkowo marginalnej* do *kompleksowej*). Model

został skonstruowany na podstawie zbioru przydziałów 48 wariantów referencyjnych ocenionych na ośmiu kryteriach, takich jak przedział wielkości cząstek, czas reakcji czy typ osprzętu. W ramach analizy zaprezentowano metodę radzenia sobie z niespójnościami w wejściowej informacji preferencyjnej. Wykryto maksymalne podzbiory spójnych przydziałów (konstrukty preferencyjne), które dalej były analizowane osobno. Posłużyły do zrozumienia wpływu poszczególnych kryteriów na rekomendacje oraz do konstrukcji odrębnych modeli i jednego reprezentatywnego. Następnie wszystkie modele wykorzystano do klasyfikacji nowych protokołów. W ramach badań pokazana została również komplementarność pomiędzy wynikami odpornej i stochastycznej regresji porządkowej. Pierwsze uzyskuje się rozwiązując problemy programowania matematycznego, a drugie wykorzystując symulacje Monte Carlo.

# Interaktywny proces uzyskiwania informacji preferencyjnej w formie przykładów decyzji

Uzyskiwanie informacji preferencyjnej może mieć formę interaktywnego procesu, który polega na iteracyjnym dostarczaniu jej przez decydenta oraz włączaniu na bieżąco do konstruowanego modelu. Popularną i stosunkowo prostą formą takiej informacji są przykłady decyzji: przydziały do klas w problemie sortowania oraz porównania parami w rankingu i wyborze. Aby uzyskać rekomendację odporną, tj. zgodną ze wszystkimi instancjami modelu spójnymi z preferencjami, decydent musi dostarczyć wiele przykładów decyzji. Ich wybór oraz kolejność wprowadzania wpływają na długość procesu analizy, czyli liczbę interakcji pomiędzy użytkownikiem a metodą.

Dla problemów wielokryterialnego rankingu, wyboru i sortowania zaproponowano szereg metod aktywnego uczenia, które dla każdej interakcji wybierają pytanie do decydenta. Pytanie dotyczy dwóch wariantów (jednego w przypadku sortowania), a odpowiedzią jest przykład decyzji, czyli wskazanie lepszego z pary (lub przydział do klasy w przypadku sortowania). Celem proponowanych metod jest wybór takich pytań, aby możliwie zminimalizować liczbę interakcji z decyden-



tem. Rozmiar rozwinięcia wszystkich możliwych scenariuszy pytanie-odpowiedź (przestrzeń przeszukiwania) jest zależny wykładniczo od rozmiaru instancji problemu decyzyjnego. W związku z tym, zaproponowane w rozprawie algorytmy mają charakter heurystyczny.

Na każdym etapie interakcji algorytm ocenia wszystkie możliwe do zadania pytania, którymi są pary wariantów, co do których relacja preferencji jest nieustalona, lub niesklasyfikowane precyzyjnie warianty w przypadku sortowania. Ocena jest dwustopniowa. W pierwszym kroku określone zostają udziały możliwych odpowiedzi w zmniejszeniu niepewności rekomendacji. Następnie są one agregowane do jednej wartości dla ocenianego pytania.

Proponowane algorytmy różnią się między sobą typem wyników odpornej lub stochastycznej regresji porządkowej wykorzystanym do oceny oczekiwanego zysku z odpowiedzi oraz metodami agregacji. Celem heurystyk jest wybór takiego pytania, które bez względu na udzieloną odpowiedź będzie miało możliwie największy wkład w minimalizację niepewności. Część heurystyk symuluje kolejne stany interakcji w celu oceny zysku z odpowiedzi, a inne wykorzystują jedynie bieżący stan.

Jakość algorytmów została sprawdzona w testach z użyciem sztucznych i prawdziwych zbiorów danych oraz w eksperymencie z rzeczywistymi decydentami. Zbadano skuteczność heurystyk przy zmieniającej się liczbie wariantów decyzyjnych i kryteriów oraz klas w przypadku sortowania. Sprawdzono wpływ sposobu generowania sztucznych zbiorów danych na ich właściwości. Ponadto zbadano również zachowanie algorytmów dedykowanych problemowi wielokryterialnego sortowania przy rozwiązywaniu remisów, ograniczaniu rozważanych wariantów w każdej iteracji do losowo wybranego podzbioru oraz stosowaniu alternatywnych kryteriów stopu.

Testy wskazały najlepsze algorytmy oraz wykazały, że prezentowane metody mogą być skuteczniejsze od istniejących rozwiązań. Największy zysk ze stosowania zaproponowanych heurystyk wyboru pytań można osiągnąć dla problemów z większymi liczbami wariantów oraz mniejszymi liczbami kryteriów. Eksperyment z rzeczywistymi decydentami, przeprowadzony dla problemu wielokryterialnego rankingu, potwierdził wyniki uzyskiwane w testach na sztucznych danych.

Spośród zaproponowanych algorytmów dla rankingu, jednym z najskuteczniejszych okazał się wybór pytania, które minimalizuje oczekiwaną liczbę par wariantów, dla których relacja preferencji nie będzie ustalona w następnym etapie interakcji z decydentem. Podobna metoda osiąga dobre rezultaty w przypadku wielokryterialnego wyboru. Zamiast liczby par wykorzystuje ona liczbę potencjalnie optymalnych wariantów. Z kolei dla sortowania najlepsze wyniki uzyskała heurystyka, która do oceny wskazuje wariant o największej entropii możliwych przydziałów do klas.

Ostateczny dobór heurystyki powinien jednak zależeć od konkretnego problemu. W przypadku pojedynczej analizy bardziej opłacalna może okazać się optymalizacja najgorszego przypadku, a przy powtarzalnym eksperymencie (np. z większą liczbą decydentów) oczekiwany przypadek.

## *Analiza post factum*

Kiedy rekomendacja została już wypracowana, można zbadać jej stabilność przy pomocy zaproponowanej metody analizy *post factum*. Pozwala ona określić minimalną wymaganą zmianę (polepszenie) w ocenach wariantu na wybranych kryteriach, tak aby dopuścić obecnie nieosiągnięty cel lub znaleźć maksymalne pogorszenie, aby zagwarantować zachowanie bieżącego stanu. Celem (stanem) może być zajęcie pozycji w rankingu (w szczególności bycie najlepszym), zachodzenie relacji preferencji pomiędzy parą wariantów lub przydział do klasy. Wszystkie cele mogą być rozważane w sensie możliwym i koniecznym. Ten pierwszy oznacza dopuszczalność wyniku przez co najmniej jeden model spójny z preferencjami decydenta, a drugi dotyczy sytuacji, w której wszystkie takie modele wspierają dany wynik. Ponadto analiza *post factum* może operować w odniesieniu do pojedynczej instancji modelu preferencji lub do zbioru instancji będących wynikiem odpornej regresji porządkowej oraz pozwala na badanie osiągalności celów przez jeden wariant podczas gdy zmieniają się oceny innego.

Przykładem typowego zadania analizy *post factum* jest sprawdzenie w jaki sposób oceny z wybranego podzbioru kryteriów drugiego w rankingu wariantu musiałyby się zmienić, aby zapewnić mu pierwszą pozycję. Podobnie z drugiej strony można rozpatrywać na ile obecnie najlepszy wariant może się pogorszyć, aby jego pozycja nie uległa zmianie.

W pracy badawczej zaproponowano metodę operującą jednocześnie na całym wybranym podzbiornie kryteriów z wykorzystaniem wyszukiwania binarnego. Zaprezentowano również uogólnienie tego podejścia, które wymaga rozwiązania problemu wielokryterialnej optymalizacji i generuje zbiór niezdominowanych rozwiązań – wektorów zmian koniecznych do osiągnięcia analizowanego celu. Drugie podejście jest bardziej kosztowne obliczeniowo, ale dostarcza szerszego wglądu w problem decyzyjny, ilustrując kompromisy występujące pomiędzy kryteriami. Umożliwia również analizę pogarszania ocen na jednych kryteriach z równoczesnym polepszaniem na innych.

Możliwości analizy *post factum* zostały zaprezentowane na przykładzie studium przypadku. Problem dotyczył pięciu możliwych sposobów rewitalizacji nieużywanej linii kolejowej w północnych Włoszech. Warianty postępowania (m.in. utworzenie pasa zieleni, renowacja starej stacji kolejowej oraz przywrócenie transportu) ocenione były na dziewięciu kryteriach, w tym pod kątem wpływu na środowisko, kosztów inwestycji oraz liczby potencjalnych użytkowników. W analizowany problem zaangażowanych było dwóch rzeczywistych decydentów. W ramach badań sprawdzono wpływ zmian ocen poszczególnych wariantów na rekomendacje. Analizowano m.in. jakie minimalne polepszenia na kryteriach są wymagane, aby przywrócenie transportu było preferowane ponad renowację stacji kolejowej dla wszystkich spójnych modeli preferencji, jakie warunki muszą zaistnieć, aby potencjalnie optymalne warianty osiągnęły najwyższą pozycję w rankingu dla obu decydentów oraz jak dodatkowe wydatki na powiększenie terenów zielonych mogą sprawić, że najbardziej preferowanym scenariuszem stałaby się renowacja stacji kolejowej.

Analiza *post factum* może znaleźć zastosowanie m.in.

w projektowaniu, planowaniu i formułowaniu wymagań. Uzupełnia analizę odporności i pozwala uprościć dyskusję pomiędzy decydentami, dając bardziej szczegółowy wgląd jak jeden wariant przewyższa inny i w odniesieniu do których kryteriów. Ponadto narzędzia analizy *post factum* mogą wskazać jak istniejące warianty powinny zostać zmienione, aby osiągały wymagane cele, lub pomóc w projektowaniu nowych lepszych wariantów decyzyjnych.

## Podsumowanie

Zaproponowano nowy typ pośredniej i nieprecyzyjnej informacji preferencyjnej dla wielokryterialnego sortowania opartego na addytywnej funkcji wartości. Są to porównania parami operujące na przydziałach do klas. Razem z przykładami przydziałów oraz oczekiwanymi licznosciami klas zostały one zorganizowane w jedno narzędzie wspomagania decyzji. Integruje ono również trzy rodzaje wyników analizy odporności, które korespondują z typami informacji preferencyjnej. Ta odpowiedniość pomiędzy wejściem i wyjściem metody umożliwi decydentom obserwację wpływu dostarczanych przez nich informacji na rekomendacje.

Zaproponowano również kilkanaście heurystycznych metod dla wyboru pytań w interaktywnym procesywnym uzyskiwaniu informacji preferencyjnej od decydenta. Metody te wykorzystują wyniki stochastycznej i odpornej regresji porządkowej, aby ocenić nieznaną odpowiedź decydenta. Celem ich wykorzystania jest uniknięcie włączania do modelu bezwartościowej informacji oraz zmniejszenie liczby interakcji z decydentem. Jakość zaproponowanych heurystyk została oceniona z wykorzystaniem sztucznych i prawdziwych zbiorów danych oraz w eksperymencie z rzeczywistymi decydentami.

Kolejnym elementem rozprawy jest przedstawienie narzędzia analizy *post factum*, która ocenia rekomendacje z perspektywy ich stabilności i pozwala na sprawdzenie zależności pomiędzy kryteriami. Bada ono, jak powinny polepszyć się

oceny na kryteriach poszczególnych wariantów decyzyjnych, aby obecnie nieosiągalny cel decyzyjny stał się możliwy, lub jak pogorszyć takie oceny, aby utrzymać bieżący stan (pozycję w rankingu, relację preferencji lub przydział do klasy).

Część zaproponowanych metod wraz z istniejącymi rozwiązaniami wielokryterialnego wspomaganie decyzji zostało zastosowane w praktyce. Wykorzystano je do oceny protokołów syntezy nanocząstek srebra oraz w problemie wyboru sposobu rewitalizacji nieużywanej linii kolejowej. Poza wsparciem decydentów, aplikacje miały na celu demonstrację użyteczności proponowanych metod oraz komputerowego wspomaganie decyzji w ogóle.

Problemy decyzyjne są obecne w wielu obszarach, m.in. w ekonomii, zarządzaniu i nauce. Dziedzina wielokryterialnego wspomaganie decyzji dostarcza narzędzi, które umożliwiają ich analizę oraz wypracowywanie rekomendacji. Rozwój metod opisanych w niniejszej rozprawie może pomóc w podejmowaniu skuteczniejszych decyzji.

## Najważniejsze publikacje

- [P1] Kadziński, M., Ciomek, K., and Słowiński, R. (2015). Modeling assignment-based pairwise comparisons within integrated framework for value-driven multiple criteria sorting. *European Journal of Operational Research*, 241(3):830–841, DOI: 10.1016/j.ejor.2014.09.050.
- [P2] Kadziński, M., Ciomek, K., Rychły, P., and Słowiński, R. (2016). Post factum analysis for robust multiple criteria ranking and sorting. *Journal of Global Optimization*, 65(3):531–562, DOI: 10.1007/s10898-015-0359-3.
- [P3] Ciomek, K., Kadziński, M., and Tervonen, T. (2017a). Heuristics for prioritizing pair-wise elicitation questions with additive multi-attribute value models. *Omega*, 71:27 – 45, DOI: 10.1016/j.omega.2016.08.012.
- [P4] Ciomek, K., Kadziński, M., and Tervonen, T. (2017b). Heuristics for selecting pair-wise elicitation questions in multiple criteria choice problems. *European Journal of Operational Research*, 262(2):693 – 707, DOI: 10.1016/j.ejor.2017.04.021.
- [P5] Kadziński, M., Cinelli, M., Ciomek, K., Coles, S. R., Nadagouda, M. N., Varma, R. S., and Kirwan, K. (2018). Co-constructive development of a green chemistry-based model for the assessment of nanoparticles synthesis.



*European Journal of Operational Research*, 264(2):472 – 490, DOI: 10.1016/j.ejor.2016.10.019.

- [P6] Ciomek, K., Ferretti, V., and Kadziński, M. (2018). Predictive analytics and disused railways requalification: Insights from a post factum analysis perspective. *Decision Support Systems*, 105(Supplement C):34 – 51, DOI: 10.1016/j.dss.2017.10.010.



© 2019 Krzysztof Ciomek

Politechnika Poznańska  
Wydział Informatyki  
Instytut Informatyki  
Skład przy użyciu systemu L<sup>A</sup>T<sub>E</sub>X

Bib<sub>T</sub>E<sub>X</sub>:

```
@phdthesis{ Ciomek2019,  
  author = "Krzysztof Ciomek",  
  title = "{Advances in Preference Disaggregation and Post  
Factum Analysis with Multiple Criteria Additive Value Models}",  
  school = "Pozna{\n} University of Technology",  
  address = "Pozna{\n}, Poland",  
  year = "2019",  
}
```