

AUTOREFERAT

Analiza i projektowanie miar konfirmacji o pożądanym właściwościach
z punktu widzenia oceny reguł wyindukowanych z danych

dr inż. Izabela Szczęch

Instytut Informatyki
Wydział Informatyki
Politechnika Poznańska

Poznań, 29 sierpnia 2017r.

Pracę dedykuję moim dwóm najwspanialszym i najwartościowszym „osiągnięciom”
- Helence i Adasiowi
oraz ich „współautorowi” - mojemu kochanemu mężowi Bartkowi.

I. IMIĘ I NAZWISKO

Izabela Szczęch

II. POSIADANE STOPNIE I TYTUŁY NAUKOWE

1. **Stopień doktora nauk technicznych** - Politechnika Poznańska, Wydział Informatyki i Zarządzania; dyscyplina - informatyka:
 - data nadania: 18 marca 2008r.;
 - tytuł rozprawy doktorskiej: "Multicriteria attractiveness evaluation of decision and association rules";
 - promotor: prof. dr hab. inż. Roman Słowiński;
 - studium doktoranckie w latach 2004r.-2008r.
2. **Tytuł zawodowy magistra inżyniera** - Politechnika Poznańska, Wydział Informatyki i Zarządzania; kierunek - informatyka; specjalizacja – projektowanie i eksploatacja systemów informatycznych:
 - data nadania: 10 września 2004r.;
 - tytuł pracy magisterskiej: "Knowledge Extraction from the Database of the Polish Registry of Congenital Malformations";
 - promotor: prof. dr hab. inż. Roman Słowiński;
 - studia na X semestrze na Université Paris Dauphine we Francji w ramach międzynarodowej wymiany studenckiej Socrates/Erasmus i uzyskanie Certificate of Appreciation z Université Paris Dauphine (podwójny dyplom).
3. **Tytuł zawodowy inżyniera** - Politechnika Poznańska, Wydział Elektryczny; kierunek - informatyka:
 - data nadania: 12 lipca 2002r.;
 - tytuł pracy inżynierskiej: "Projekt i implementacja modułu ewidencji zmian chorobowych z wykorzystaniem technologii Oracle Forms Developer 6i oraz Oracle Application Server";
 - promotor: dr inż. Dariusz Wawrzyniak.

III. DOTYCHCZASOWE ZATRUDNIENIE W JEDNOSTKACH NAUKOWYCH

1. **Analityk-projektant aplikacji**
 - Instytut Informatyki, Wydział Informatyki i Zarządzania, Politechnika Poznańska;
 - okres: 1 listopada 2004r. - 30 września 2008r.
 - w tym urlop macierzyński w okresie: 8 sierpnia 2008r. – 30 września 2008r.
2. **Adiunkt**
 - Wydział Informatyki i Komunikacji Wizualnej, Wyższa Szkoła Nauk Humanistycznych i Dziennikarstwa w Poznaniu (obecnie Collegium Da Vinci);
 - okres: 1 października 2009r. - 30 września 2014r.

3. Adiunkt

- Instytut Informatyki, Wydział Informatyki, Politechnika Poznańska;
- okres: od 1 października 2008r.
 - w tym urlopy macierzyńskie w okresach: 1 października 2008r. – 11 grudnia 2008r., oraz 19 czerwca 2015r. – 16 czerwca 2016r.

IV. WSKAZANIE OSIĄGNIĘCIA NAUKOWEGO WYNIKAJĄCEGO Z USTAWY O STOPNIACH NAUKOWYCH I TYTULE NAUKOWYM ORAZ O STOPNIACH I TYTULE W ZAKRESIE SZTUKI

A. TYTUŁ OSIĄGNIĘCIA NAUKOWEGO

Analiza i projektowanie miar konfirmacji o pożądanych właściwościach z punktu widzenia oceny reguł wyindukowanych z danych

B. LISTA PRAC WCHODZĄCYCH W SKŁAD OSIĄGNIĘCIA NAUKOWEGO

- [P1] Salvatore Greco, Roman Słowiński, Izabela Szczęch
Properties of rule interestingness measures and alternative approaches to normalization of measures.
Information Sciences vol. 216 (2012), 1–16.
 Pkt MNiSW(2012)=40 ; IF(2012)=3,643
- [P2] Salvatore Greco, Roman Słowiński, Izabela Szczęch
Finding Meaningful Bayesian Confirmation Measures.
Fundamenta Informaticae vol. 127 (2013), 161–176.
 Pkt MNiSW(2017)=20 ; IF(2013)=0,479
- [P3] Robert Susmaga, Izabela Szczęch
Can Interestingness Measures Be Usefully Visualized?
International Journal of Applied Mathematics and Computer Science vol. 25, no. 2 (2015), 323–336.
 Pkt MNiSW(2017)=25 ; IF(2015)=1,037
- [P4] Robert Susmaga, Izabela Szczęch
Visualization support for the analysis of properties of interestingness measures.
Bulletin of the Polish Academy of Sciences Technical Sciences vol. 63, issue 1 (2015), 315–327.
 Pkt MNiSW(2017)=25 ; IF(2015)=1,087
- [P5] Salvatore Greco, Roman Słowiński, Izabela Szczęch
Measures of rule interestingness in various perspectives of confirmation.
Information Sciences vol. 346–347C (2016), 216–235.
 Pkt MNiSW(2017)=45 ; IF(2016)=4,832
- [P6] Robert Susmaga, Izabela Szczęch
Selected group-theoretic aspects of confirmation measure symmetries.
Information Sciences vol. 346–347C (2016), 424–441.
 Pkt MNiSW(2017)=45 ; IF(2016)=4,832

- [P7] Krystyna Napierała, Jerzy Stefanowski, Izabela Szczęch
Increasing the Interpretability of Rules Induced from Imbalanced Data by Using Bayesian Confirmation Measures.
 In: Appice, A., Ceci, M., Loglisci, C., Masciari, E., Ras, Z.W. (Eds.) *New Frontiers in Mining Complex Patterns – 5th International Workshop, NFMCP 2016, Held in Conjunction with ECML-PKDD 2016, Riva del Garda, Italy, September 19, 2016, Revised Selected Papers*
 LNAI 10321, 84-98, Springer (2017).
 Pkt MNiSW=0 ; konf. jeszcze nie indeks. w WoS (indeksowane 2nd, 4th)
- [P8] Dariusz Brzezinski, Zbigniew Grudziński, Izabela Szczęch
Bayesian Confirmation Measures in Rule-based Classification.
 In: Appice, A., Ceci, M., Loglisci, C., Masciari, E., Ras, Z.W. (Eds.) *New Frontiers in Mining Complex Patterns – 5th International Workshop, NFMCP 2016, Held in Conjunction with ECML-PKDD 2016, Riva del Garda, Italy, September 19, 2016, Revised Selected Papers*
 LNAI 10321, 39-53, Springer (2017).
 Pkt MNiSW=0 ; konf. jeszcze nie indeks. w WoS (indeksowane 2nd, 4th)

C. OMÓWIENIE CELU NAUKOWEGO WW. PRAC ORAZ OSIĄGNIĘTYCH WYNIKÓW WRAZ Z OMÓWIENIEM ICH EWENTUALNEGO WYKORZYSTANIA

Cykl prac składających się na osiągnięcie naukowe dotyczy odkrywania wiedzy z danych. Istotą tego obszaru informatyki jest poszukiwanie nowych, potencjalnie użytecznych i interpretowalnych wzorców, rozumianych jako pewna regularność, podobieństwo lub zależność funkcyjna występująca pomiędzy elementami danych [Fay91]. Odkrywane wzorce mogą być reprezentowane np. w formie asocjacji, drzew i reguł decyzyjnych, skupień, modeli regresji wielowymiarowej [Pia91]. Przedstawiany cykl prac koncentruje się na wzorcach w formie reguł decyzyjnych, które stanowią wyrażenia w postaci:

jeżeli spełnione określone warunki *to* określona decyzja,

gdzie część warunkowa (przesłanka) zawiera koniunkcję warunków elementarnych tworzonych w oparciu o wartości atrybutów opisujących obiekty, a część decyzyjna (konkluzja, hipoteza) określa przynależność obiektu spełniającego te warunki do danej klasy decyzyjnej. Zapis symboliczny reguły często przyjmuje formę: $E \rightarrow H$, gdzie E i H oznaczają odpowiednio część warunkową i decyzyjną.

Większa niż przy innych formach reprezentacji wiedzy czytelność dla człowieka oraz modułarna budowa wpłynęły na dużą popularność reguł. W efekcie, reguły decyzyjne są obecnie powszechnie wykorzystywane do reprezentacji wiedzy w procesach wspomaganie decyzji w różnorodnym kontekście zastosowań [Han00].

Poszukiwanie reguł zwykle odbywa się drogą indukcyjnego uogólnienia opisu przykładów uczących [Ste01]. Indukcję reguł w odkrywaniu wiedzy można rozpatrywać w dwóch kontekstach: predykcyjnym oraz deskrypcyjnym. Istotą indukcji ukierunkowanej na predykcję jest generowanie z danych zbioru reguł, które następnie będą wykorzystywane do klasyfikowania nowych obiektów, czyli przewidywania ich przydziału do pewnych klas obiektów na podstawie wiedzy reprezentowanej w regułach, a wynikającej z analizy przykładów klasyfikacji obiektów w przeszłości. Celem indukcji ukierunkowanej na

deskrypcję jest natomiast wyszukanie reguł najlepiej charakteryzujących procesy opisane przez dane, a zatem reguł ułatwiających interpretację zależności pomiędzy wartościami atrybutów a definicją klasy decyzyjnej w tych procesach.

Wspólnym problemem dotyczącym zarówno indukcji reguł predykcyjnych jak i deskrypcyjnych jest częste generowanie zbyt dużej liczby reguł, która może powodować chociażby przeuczenie klasyfikatora zbudowanego z reguł, czy też uniemożliwić ekspertowi analizę i korzystanie z reguł. W tym kontekście uwidacznia się potrzeba oceniania użyteczności reguł, która pozwala na zawężanie zbioru reguł i koncentrowanie uwagi tylko na regułach najbardziej użytecznych.

Ocena reguł zwykle przeprowadzana jest z wykorzystaniem pewnych ilościowych wskaźników zwanych *miarami atrakcyjności*. W literaturze dotyczącej odkrywania wiedzy i eksploracji danych można znaleźć bardzo wiele propozycji miar tego typu [Gen06, McG05]. Każda z miar atrakcyjności kładzie nacisk na uwzględnienie innych cech pojedynczej reguły, czy całego ich zbioru. Ogólnie miary te można podzielić na subiektywne i obiektywne [Ste01b]. Miary subiektywne biorą pod uwagę wiedzę dziedzinową i przekonania eksperta, podczas gdy miary obiektywne są niezależne od dziedziny zastosowania czy użytkownika i obliczane są wyłącznie na podstawie analizowanego zbioru danych. Przedmiotem badań w przedstawianym cyklu prac są miary obiektywne [Fre99].

Mnogość obiektywnych miar atrakcyjności proponowanych w literaturze powoduje, że sam proces wyboru właściwej miary staje się nietrywialny. To z kolei rodzi często pokusę upraszczania wyboru przez ograniczanie się do kryterium popularności miary. Niestety powszechność stosowania miary może jednak wynikać jedynie z prostoty jej definicji, a nie z istotnych właściwości miary, określających jej zachowanie w pewnych sytuacjach.

W tym kontekście bardzo ważnym i aktualnym tematem badawczym jest analizowanie właściwości miar atrakcyjności reguł i projektowanie nowych miar tak, by spełniały one właściwości zgodne z założoną racjonalnością. Tego typu właściwości będziemy nazywać pożądanymi. Stosowanie miar o pożądanymi właściwościami pozwala na znajdowanie użytecznych reguł, bo miary te odzwierciedlają przez swoje właściwości racjonalne oczekiwania co do ich zachowania. Dzięki temu miary te promują reguły polepszające interpretowalność reprezentowanych procesów.

Przedstawiany cykl publikacji koncentruje się na grupie miar atrakcyjności, zwanych miarami konfirmacji. Ich wspólną cechą jest spełnianie właściwości konfirmacji [Pop59, Mah05, Fit01], którą formalnie można ująć w następujący sposób:

Dla reguły $E \rightarrow H$ miara atrakcyjności $c(H, E)$ spełnia właściwość konfirmacji, gdy

- $c(H, E) > 0$ wtedy i tylko wtedy gdy $P(H|E) > P(H)$,
- $c(H, E) = 0$ wtedy i tylko wtedy gdy $P(H|E) = P(H)$,
- $c(H, E) < 0$ wtedy i tylko wtedy gdy $P(H|E) < P(H)$,

gdzie $P(H|E)$ oznacza prawdopodobieństwo zajścia konkluzji pod warunkiem zajścia przesłanki, a $P(H)$ oznacza prawdopodobieństwo zajścia konkluzji.

Poruszając się na gruncie oceny reguł indukowanych z danych, można przyjąć, że odpowiednie częstości określone na podstawie zbioru danych mogą być użyte do estymowania powyższych prawdopodobieństw. Jak łatwo zauważyć, warunki definiujące właściwość konfirmacji określają jedynie, kiedy miara ma przyjmować wartości dodatnie,

kiedy zero, a kiedy wartości ujemne. W literaturze jest zatem wiele propozycji miar confirmacji [Mah05, Fit01], które różnią się jednak między sobą licznymi innymi właściwościami.

Użyteczność miar confirmacji, a zarazem ich przewaga pod kątem interpretowalności procesów opisanych przez dane nad miarami nie spełniającymi właściwości confirmacji, może być zilustrowana następującym przykładem. Dla rzutu kostką do gry, w którym liczba wyrzuconych oczek jest ze zbioru {1, 2, 3, 4, 5, 6}, rozważmy przykładową przesłankę reguły: $E =$ "liczba wyrzuconych oczek jest podzielna przez dwa" oraz dwie konkluzje: $H_1 =$ "liczba wyrzuconych oczek jest szóstką", $H_2 =$ "liczba wyrzuconych oczek nie jest szóstką".

Ocenę reguł $E \rightarrow H_1$ oraz $E \rightarrow H_2$ przeprowadzamy najpierw z wykorzystaniem *ufności* (*confidence*) [Agr93] czyli bardzo popularnej miary oceny reguł, która nie posiada właściwości confirmacji, a jest definiowana jako stosunek liczby obiektów spełniających przesłankę i konkluzję do liczby obiektów spełniających przesłankę: $confidence(E \rightarrow H) = P(E \wedge H) / P(E) = P(H|E)$. Dla pierwszej reguły $confidence(E \rightarrow H_1) = 1/3$, a dla drugiej $confidence(E \rightarrow H_2) = 2/3$. Jeśli użytkownik kieruje się miarą *ufności*, reguła druga jest zatem jednoznacznie preferowana nad regułą pierwszą. Co więcej, wysoka wartość tej miary dla $E \rightarrow H_2$ najprawdopodobniej zmotywuje użytkownika do użycia reguły i podjęcia sugerowanych przez nią działań.

Ocena powyższych reguł z wykorzystaniem dowolnej miary confirmacji prowadzi jednak do zupełnie innych wniosków. W szczególności, reguła $E \rightarrow H_1$ charakteryzuje się dodatnią wartością miar confirmacji, gdyż $P(H_1|E) = 1/3 > P(H_1) = 1/6$. Z kolei regułę $E \rightarrow H_2$ charakteryzuje ujemna wartość miar confirmacji, gdyż $P(H_2|E) = 2/3 < P(H_2) = 5/6$. Oznacza to, że reguła $E \rightarrow H_2$ jest w rzeczywistości myląca, bo zajęcie przesłanki zmniejsza prawdopodobieństwo zajęcia konkluzji. Kierowanie się zatem wysoką wartością *ufności* byłoby w przypadku reguły drugiej niekorzystne dla użytkownika.

Przedstawiony przykład ilustruje zarówno praktyczne znaczenie miar confirmacji, które za pomocą swej skali od razu wskazują reguły nieprzydatne (dysconfirmujące, $c(H,E) < 0$), jak również ułomność popularnej miary *ufności*, która niestety jest dość powszechnie domyślnie używana w praktyce. Wobec znaczenia właściwości confirmacji, przedstawiany cykl prac stawia miary confirmacji w centrum zainteresowania.

Zadanie wyboru miary confirmacji spełniającej pożądane właściwości, choć znacznie zawężone z uwagi na powyżej przedstawioną podstawową właściwość confirmacji, wymaga nietrywialnego przyjrzenia się wielu alternatywnym miarom. Sama właściwość confirmacji nie jest bowiem jedynym czynnikiem gwarantującym właściwy wybór. Należy zatem sformułować na gruncie określonej racjonalności zbiór pożądanych właściwości miar confirmacji i przeanalizować dostępne miary pod tym kątem, i ewentualnie zaprojektować nowe miary.

Prezentowany cykl prac stanowi obszerną propozycję obejmującą analizę i projektowanie miar confirmacji o pożądanych właściwościach z punktu widzenia oceny reguł wyindukowanych z danych. Przechodzi on od wnikliwych badań nad pożadanymi właściwościami miar confirmacji reguł do odkrywania zależności pomiędzy różnymi właściwościami, modyfikowania istniejących właściwości i formułowania nowych. Określenie zbioru pożądanych właściwości stanowi punkt wyjścia do analizy miar confirmacji

proponowanych w literaturze, pozwala także na wyznaczenie metod ich normalizacji oraz projektowanie nowych. Analiza miar pod kątem posiadanych przez nie właściwości może być wspierana przez zaproponowaną metodę wizualizacji miar, umożliwiającą łatwe pozyskanie informacji o zachowaniu się miar w całych ich dziedzinach. Usprawnia to znacząco proces doboru miar do poszczególnych zastosowań, jak również wspiera proces projektowania nowych miar. Praktyczną przydatność miar konfirmacji o pożądanych właściwościach do oceny reguł wyindukowanych z danych potwierdzono eksperymentalnie w kontekście danych niezbalansowanych i zbalansowanych, a także problemów dwu i wieloklasowych.

Omówienie celów i wyników naukowych prac wchodzących w skład prezentowanego cyklu zostanie podzielone na trzy części, grupujące najważniejsze opracowane analizy, metody i zastosowania:

- [H1] Analiza i projektowanie miar konfirmacji i ich właściwości [P1, P2, P5, P6].
- [H2] Metoda wizualizacji miar konfirmacji jako podejście wspierające analizę i projektowanie miar konfirmacji [P3, P4].
- [H3] Zastosowania miar konfirmacji o pożądanych właściwościach do oceny reguł wyindukowanych z danych [P7, P8].

H1. Analiza i projektowanie miar konfirmacji i ich właściwości

U podstaw (nieformalnego) określenia właściwości konfirmacji leży oczekiwanie, że miara będzie przyjmowała wartości dodatnie, gdy przesłanka reguły będzie pozytywnie wpływała na wnioskowanie o konkluzji, zero, gdy przesłanka nie będzie w ogóle wpływała na konkluzję, i wreszcie wartości ujemne, gdy przesłanka reguły będzie negatywnie wpływała na wnioskowanie o konkluzji. Formalna definicja często używana w literaturze wykorzystuje zapis operujący na prawdopodobieństwach $P(H|E)$ oraz $P(H)$, jak przytoczono we wstępie. Badając wnikliwie miary konfirmacji, trzeba mieć jednak na względzie fakt, że alternatywnych zapisów formułujących samą właściwość konfirmacji jest więcej [Fit01, Mah05]. W pracy [P1], a potem bardziej szczegółowo w [P5], zostały opisane cztery sposoby sformalizowania właściwości konfirmacji zwane perspektywami:

- perspektywa *Bayesowska* (Bayesian) wymagająca, by:
 - $c(H, E) > 0 \Leftrightarrow P(H|E) > P(H)$,
 - $c(H, E) = 0 \Leftrightarrow P(H|E) = P(H)$,
 - $c(H, E) < 0 \Leftrightarrow P(H|E) < P(H)$;
- perspektywa *silna Bayesowska* (strong Bayesian) wymagająca, by:
 - $c(H, E) > 0 \Leftrightarrow P(H|E) > P(H|\neg E)$,
 - $c(H, E) = 0 \Leftrightarrow P(H|E) = P(H|\neg E)$,
 - $c(H, E) < 0 \Leftrightarrow P(H|E) < P(H|\neg E)$;
- perspektywa *potencjalnościowa* (likelihoodist) wymagająca, by:
 - $c(H, E) > 0 \Leftrightarrow P(E|H) > P(E)$,
 - $c(H, E) = 0 \Leftrightarrow P(E|H) = P(E)$,
 - $c(H, E) < 0 \Leftrightarrow P(E|H) < P(E)$;

- perspektywa *silna potencjalnościowa* (strong likelihoodist) wymagająca, by:
 - $c(H, E) > 0 \Leftrightarrow P(E|H) > P(E|\neg H)$,
 - $c(H, E) = 0 \Leftrightarrow P(E|H) = P(E|\neg H)$
 - $c(H, E) < 0 \Leftrightarrow P(E|H) < P(E|\neg H)$.

Perspektywy są zatem możliwymi różnymi sposobami wyrażenia tego samego pojęcia, czyli właściwości konfirmacji. Poprzez użycie w definicji innych prawdopodobieństw, każda z nich w naturalny sposób kładzie nacisk na różne aspekty, pozwalając przez to spojrzeć na właściwość konfirmacji pod innym kątem (z innej perspektywy). Przykłady ilustrujące różnice między perspektywami zostały szczegółowo omówione w [P5], gdzie poruszono również kwestię pewnej odpowiedniości perspektyw Bayesowskich i potencjalnościowych, z uwagi na fakt, że reguła $E \rightarrow H$ w podejściu Bayesowskim odpowiada regule $H \rightarrow E$ w podejściu potencjalnościowym.

Przedstawione cztery perspektywy są zapisami wykorzystującymi inne prawdopodobieństwa, a zatem przyjmującymi wartości nieokreślone w różnych sytuacjach (np. gdy $P(E)=0$ wówczas $P(H|E)$ jest nieokreślone w perspektywie *Bayesowskiej* i *silnej Bayesowskiej*). Jednak mimo tych różnic, wszystkie cztery perspektywy są logicznie równoważne, przez co rozumie się, że „przełączają” się one między wartościami dodatnimi, zerem i wartościami ujemnymi w tych samych momentach, jeśli nie prowadzą do wartości nieokreślonych ([P1], [Gla13]). Jak pokazano szczegółowo w [P5], warunki ujęte w każdej z powyższych perspektyw można też sprowadzić do jednego sformułowania określanego jako *ogólna definicja właściwości konfirmacji*. Warto podkreślić, że choć wszystkie perspektywy (pomijając przypadki wartości nieokreślonych) w ten sam sposób wymagają, by miara przyjmowała wartości dodatnie, zero i wartości ujemne, to jednak są różnymi sformułowaniami i inspirują do innego spojrzenia na właściwość konfirmacji.

W szczególności perspektywa *Bayesowska* stawała się często punktem wyjścia do definiowania właściwości miar konfirmacji. Tak było m.in. w przypadku właściwości $Ex1$ zaproponowanej w [Cru07], czy właściwości L (logicality) omawianej w [Cru07, Gla13]. Obie te właściwości opisują oczekiwania wobec zachowania się miar w ekstremalnych sytuacjach, czyli gdy w danych nie ma kontrprzykładów albo pozytywnych przykładów dla reguły. Analizy prowadzone w pracach [P1, P5] wskazały jednak, że właściwości $Ex1$ oraz L mogą prowadzić do uporządkowania reguł niezgodnego z racjonalną intuicją. Przykłady paradoksalnych sytuacji (wynikających ze spełnienia właściwości $Ex1$ czy L), w których bardzo duży przyrost konfirmacji był pomijany z uwagi na pojedynczy kontrprzykład dla reguły, zostały zilustrowane i skomentowane w pracach [P1, P5]. Stały się one jednocześnie motywacją do poszukiwania modyfikacji tych właściwości. W efekcie, w pracy [P1] zaproponowano uogólnienie właściwości $Ex1$ do właściwości weak $Ex1$ i analogicznie właściwości L do weak L . Uzupełnieniem rozważań w [P1] są wyniki omówione w [P5] definiujące właściwość $Ex1'$, która jako komplementarna względem oryginalnej propozycji $Ex1$ [Cru07], współtworzy właściwość weak $Ex1$. Analogiczne badania dotyczą też właściwości L [P5]. Dodatkowo, w [P5] przedstawiono propozycję potencjalnościowego odpowiednika właściwości weak $Ex1$, oznaczanego weak $L-Ex1$ (likelihoodist counterpart of weak $Ex1$), a w pracy [P1] potencjalnościowych odpowiedników właściwości $Ex1$ oraz L , oznaczanych odpowiednio jako $L-Ex1$ oraz $L-L$. W publikacjach [P1, P5] dopełnieniem badań

związanych z właściwościami jest analiza miar konfirmacji pod względem spełniania zaproponowanych właściwości.

Niespełnianie przez wiele miar konfirmacji właściwości Ex1 stało się motywacją do przedstawienia w [Cru07] sposobu transformacji miar, tak by powstała w jej wyniku miara posiadała właściwość Ex1. Normalizacja zaproponowana w [Cru07] przekształca wiele miar znanych w literaturze, do jednej wspólnej postaci, zwanej miarą Z, która spełnia właściwość Ex1. Jak pokazano jednak w pracy [P1] istnieją również inne sposoby transformacji miar inspirowane różnymi sposobami interpretowania maksymalnego/minimalnego potwierdzenia (konfirmowania) konkluzji przez przesłankę reguły, czyli czerpiące z właściwości konfirmacji. W [P1] zdefiniowano cztery nowe podejścia do normalizacji miar, omówiono różnice między nimi i zastosowano je do popularnych miar konfirmacji. W szczególności, zaproponowano potencjalnościowy odpowiednik transformacji z [Cru07], który doprowadził do zdefiniowania nowej miary A, będącej potencjalnościowym odpowiednikiem miary Z i spełniającej właściwość L-Ex1. Swoista komplementarność miary Z oraz A stała się także motywacją do zdefiniowania w [P1] nowych miar konfirmacji (c_1 , c_2 , c_3 oraz c_4) bazujących na Z oraz A. Agregacja miar składowych przeprowadzona została tak, by nowe miary posiadały pożądane właściwości. Porównanie w [P1] wyników zastosowania zaproponowanych normalizacji pozwoliło zidentyfikować miary, które różnymi metodami przekształcają się do tej samej miary, oraz zidentyfikować podejścia, które różne miary transformują do tej samej postaci. Dla miar znormalizowanych zaproponowanymi metodami przeanalizowano ich właściwości (m.in. Ex1, L, weak Ex1, weak L) oraz określono ich porządkową nierównoważność [P1].

Prace [P2, P6] skupiają się na kolejnych właściwościach często rozważanych w kontekście miar konfirmacji - na właściwościach symetrii, stanowiących cały zestaw właściwości, określających oczekiwania wobec zachowania się miar w sytuacjach, gdy zanegowana jest przesłanka lub konkluzja reguły, albo gdy są one zamienione miejscami. Badania nad właściwościami symetrii podejmowane były już wcześniej w literaturze (m.in. [Fit01, Eel02, Cru07]), jednakże konkluzje z nich płynące różniły się i nie dawały spójnej odpowiedzi na pytanie, które symetrie są pożądane i powinny być spełniane przez użyteczne miary. W tym kontekście, w pracy [P2] podjęto zadanie zestawienia i krytycznego porównania istniejących podejść do wartościowania własności symetrii (czyli określania, które z nich są pożądane, a które nie powinny być przez miary spełniane). W wyniku wnikliwej analizy, spójnej z wcześniej prowadzoną w [P1] dyskusją na temat właściwości Ex1, weak Ex1 itd., zaproponowano w [P2] nowe podejście do wartościowania symetrii. Uargumentowano, że jedynie symetrie ES, HS, oraz EHS powinny być uznawane za pożądane. Wobec nowej propozycji dotyczącej symetrii, przeprowadzono w [P2] analizę miar konfirmacji pod względem tych i innych właściwości (m.in. monotoniczności M [Gre04], weak Ex1) i wskazano miary, które powinny być uznane za szczególnie atrakcyjne.

Praca [P6] jest kontynuacją rozważań nad właściwościami symetrii prowadzoną jednak tym razem na gruncie inspiracji pochodzących z teorii grup [Hum96]. Jak już wspomniano powyżej, propozycje w literaturze [Eel02, Cru07, P2] różnią się ze względu na zestaw rozważanych symetrii. W szczególności można zaobserwować w niektórych swoistą niekompletność, rozumianą jako sytuację, w której złożenie dwóch symetrii z

proponowanego zbioru daje w efekcie symetrię nie zawartą w nim. W pracy [P6] pokazano, że pożądana kompletność byłaby zagwarantowana, gdyby symetrie i ich złożenie tworzyły grupę algebraiczną (symetrie jako elementy grupy, a złożenie symetrii jako operacja w grupie). Określono też niezbędne rozszerzenia zbiorów symetrii z poszczególnych propozycji z literatury prowadzące do uzyskania grupy algebraicznej, a w konsekwencji wyeliminowania niekompletności.

Po rozwiązaniu problemu niekompletności, skupiono się w [P6] na próbie spojrzenia przez pryzmat teorio-grupowy na dyskusje dotyczące wartościowania jako pożądanego i niepożądanego poszczególnych symetrii, a w szczególności na poszukiwaniu możliwości zobiektywizowania wartościowania symetrii. W naturalny sposób zrodziło się więc pytanie o zasady wartościowania złożenia symetrii tworzących grupę. Punktem wyjścia do tych rozważań stało się określenie, czy element neutralny grupy powinien być wartościowany jako pożądanym, czy niepożądanym. W pracy [P6] uzasadniono, że element neutralny należy traktować jako pożądanym, z tego zaś dalej wynikają następane zasady wartościowania złożenia symetrii. Analiza dotychczasowych propozycji literaturowych pokazała jednak, że zasady te są przez niektórych autorów naruszane, co czyni te podejścia niespójnymi w sensie zgodności z wprowadzonymi zasadami. W [P6] wykazano, że istnieje tylko jeden spójny podział symetrii na pożądanego i niepożądanego, który ma interpretację w kontekście oceny reguł wyindukowanych z danych. Jest on zbieżny w wartościowaniu proponowanym w pracy [P2].

Warto podkreślić, że przedstawione w [P6] oryginalne spojrzenie na dyskutowane w literaturze właściwości symetrii poprzez pryzmat grup algebraicznych, pozwoliło na otrzymanie formalnych wniosków, niezależnych od uzasadnień proponowanych przez różnych autorów dla ich zbiorów symetrii. Spojrzenie takie wydaje się być zatem ważnym i obiektywnym elementem w dyskusji w tym obszarze.

W pracy [P5] podjęto próbę spojrzenia na miary konfirmacji poprzez wspomniane wcześniej perspektywy właściwości konfirmacji i powiązania definicji konkretnych miar z perspektywami. Pozwoliło to na połączenie, nierozważanych dotąd wspólnie, aspektów jakościowych (reprezentowanych przez perspektywy) i ilościowych (reprezentowanych przez definicje miar) oceny miar.

W szczególności, badania w ramach pracy [P5] doprowadziły do zaproponowania nowego zestawu właściwości miar opartych na oczekiwaniu monotoniczności miary względem prawdopodobieństw definiujących perspektywy konfirmacji. Przykładowo, w kontekście perspektywy *Bayesowskiej* (wymagającej, by: $c(H, E) > 0 \Leftrightarrow P(H|E) > P(H)$, itd.) miarę będziemy uznawali za monotoniczną względem tej perspektywy, jeśli będzie ona niemalejąca względem $P(H|E)$ oraz nierosnąca względem $P(H)$. W pracy [P5] zdefiniowano monotoniczność względem każdej z przytoczonych wcześniej perspektyw oraz podano przykłady popularnych miar konfirmacji spełniających te właściwości. Rozważania zostały uzupełnione ciekawym twierdzeniem, mówiącym o relacjach pomiędzy zaproponowanymi monotonicznościami, a dokładniej o tym, że miara będąca silnie monotoniczna względem jednej perspektywy, nie jest silnie monotoniczna względem innych perspektyw.

Związek miary z konkretną perspektywą wyrażony za pomocą zaproponowanej monotoniczności w naturalny sposób wpisuje się w podstawowe oczekiwania wobec zachowania się miary, stanowi zatem ważne ogólne dopełnienie właściwości konfirmacji. W tym kontekście rodzi się potrzeba sprawdzenia istnienia zależności między monotonicznością w konkretnej perspektywie a innymi właściwościami (np. Ex1, weak Ex1, symetrie). Określenie ich pozwoliłoby wnioskować o spełnianiu bądź niespełnianiu pewnych właściwości już na podstawie wiedzy o monotoniczności miary w danej perspektywie. Praca [P5] w systematyczny sposób bada relacje między monotonicznościami w różnych perspektywach a właściwościami: monotoniczności M [Gre04], weak Ex1 [P1, P5], weak L_Ex1 [P1, P5], weak L [P1, P5], maximality/minimality [Gla13], oraz całym zestawem właściwości symetrii [Eel02, Cru07, P2]. Przykładowo, wykazano, że miary monotoniczne względem *silnie Bayesowskiej* albo *silnie potencjalnościowej* perspektywy zawsze spełniają właściwości weak L. Szczególnie obszerne i nietrywialne analizy zostały przeprowadzone dla właściwości symetrii. Seria wielu twierdzeń i dowodów określa m.in., czy przy monotoniczności względem konkretnych perspektyw, istnieją miary, które spełniają dane symetrie.

Podsumowaniem pracy [P5] jest scharakteryzowanie popularnych miar konfirmacji pod kątem ich monotoniczności w perspektywie oraz posiadanych przez nie innych właściwości. Wnioski z tych analiz wskazały miary szczególnie atrakcyjne ze względu na swoje właściwości, a przez to użyteczne w praktycznych zastosowaniach. Wnikliwe rozważania dotyczące zależności między właściwościami stanowią też wartościową wiedzę, która może wspierać poszukiwanie nowych miar i modyfikowanie już istniejących.

H2. Metoda wizualizacji miar konfirmacji jako podejście wspierające analizę i projektowanie miar konfirmacji

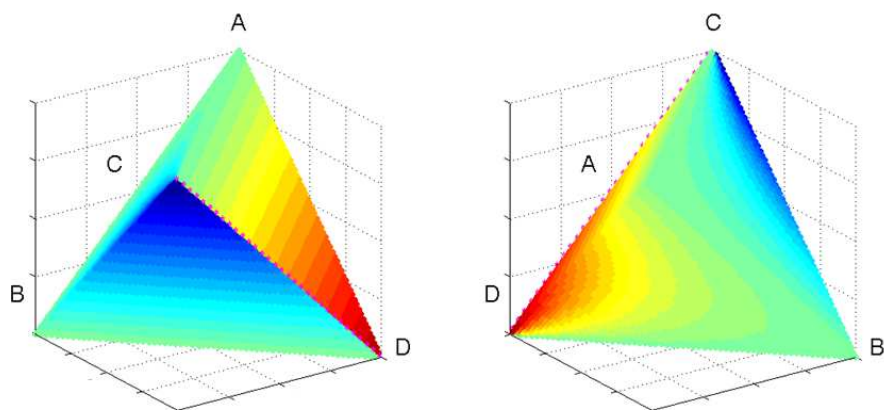
Formalna analiza miar konfirmacji pod kątem spełnianych przez nie właściwości jest pracą niejednokrotnie bardzo czasochłonną i trudną, bo wymagającą potencjalnie zaawansowanego aparatu matematycznego. Usprawnienie tego procesu to nie tylko korzyści związane z czasem, ale też możliwość rozwijania metod projektowania nowych miar (w tym metod automatycznych lub półautomatycznych).

W odpowiedzi na powyższe potrzeby, w pracy [P3] została zaprezentowana (a następnie rozwinięta w [P4]) propozycja oryginalnej metody wizualizacji miar konfirmacji, pozwalająca na łatwe i szybkie badanie analizowanych miar w całych ich dziedzinach. Takie całościowe spojrzenie na wartości i dziedzinę miary jest bardzo istotne, bo pozwala na wnioskowanie o zachowaniu się miary w dowolnej sytuacji (reprezentowanej przez konkretne wartości z jej dziedziny) i ma tym samym przewagę nad wizualizacją miary w wycinkowym kontekście opartym jedynie na wartościach przyjmowanych przez miarę dla konkretnego zbioru danych i reguł. Tylko dostęp do informacji o zachowaniu się we wszystkich obszarach dziedziny pozwala na formułowanie ogólnych wniosków o miarach.

Wprowadzona w pracy [P3] metoda wizualizacji bazuje na tym, że miary konfirmacji posiadają czterowymiarową dziedzinę, reprezentowaną w formie macierzy kontyngencji o rozmiarze 2×2 i dodatniej sumie. Przy założeniu, że suma ta jest stała i po odpowiednim jej

przeskalowaniu możliwe jest reprezentowanie pierwotnie czterowymiarowej dziedziny miar w postaci czworościanu w barycentrycznym układzie współrzędnych, który z kolei można wizualizować w przestrzeni trójwymiarowej [Flo06, War03, War04]. Nałożenie na każdy z punktów czworościanu pewnego koloru skojarzonego z konkretną wartością poprzez mapę kolorów, pozwala następnie wizualizować wartość miary przyjmowaną dla konkretnych argumentów dziedziny (czyli wpisów w macierzy kontyngencji). Metoda wykorzystywana w [P3, P4] wizualizuje zatem miary confirmacji w postaci kolorowych czworościanów.

Przykład wizualizujący miarę M [Mor88] pokazano na Rys. 1. Aby umożliwić oglądanie wszystkich ścian zewnętrznych czworościanu, rysunek zawiera dwa widoki z różnych stron. Nazwy wierzchołków czworościanu (A, B, C, D) nawiązują bezpośrednio do nazw elementów macierzy kontyngencji definiującej miarę, zastosowana zaś mapa kolorów zmienia się od ciemno niebieskiego (reprezentującego minimalne wartości miary) poprzez błękitny (reprezentujący wartości wokół zera) aż do ciemno czerwonego (reprezentującego maksymalne wartości). Rys. 1 pozwala zaobserwować m.in. położenie wartości ekstremalnych, czy pewną ciekawą niesymetrię w zachowaniu miary M uwypukloną poprzez przebieg izokwant (linii o tym samym kolorze, łączących punkty o tej samej wartości miary), które mają charakter równoległych linii na ścianach BCD oraz ACD, przechodzący w bardziej koncentryczny na ścianach ABD oraz ABC.



Rys. 1. Przykładowa wizualizacja miary confirmacji M .

Warto podkreślić jednocześnie uniwersalność zaproponowanej techniki, bo choć została ona zademonstrowana w pracach [P3, P4] na konkretnych miarach – miarach confirmacji – to z powodzeniem można jej używać dla dowolnych współczynników definiowanych w oparciu o elementy macierzy 2×2 o stałej, dodatniej sumie.

Zaproponowaną metodę wizualizacji wykorzystano w pracy [P3] do wspierania podstawowych analiz związanych z miarami confirmacji, m.in., do szybkiego uwidaczniania gdzie (w dziedzinie, czyli konkretnych punktach czworościanu) znajdują się wartości ekstremalne czy zerowe miar, w jakich obszarach miary charakteryzują się wzrostem/spadkiem wartości, oraz do określania profili gradientu (przebiegu izokwant). Takie badania mogą wspomagać chociażby wnioskowanie o tym, czy miary są porządkowo równoważne czy nie, co jest istotne do określenia nie tylko dla miar uznanych już w literaturze, ale również dla nowo proponowanych. Dodatkowo, jak zilustrowano w pracy [P3], zaprezentowaną metodę można też z powodzeniem wykorzystywać do badania grup miar,

poprzez wizualizację różnic między miarami, czy wariancji w całych grupach miar. To z kolei pozwala sprawnie wyciągać wnioski np. o tym w jakich obszarach dziedzin miary różnią się najbardziej/najmniej, albo też czy miary z jakiejś grupy podobnie się zachowują czy nie. Takie wyniki są istotne z praktycznego punktu widzenia, bo pozwalają np. w wielokryterialnej ocenie reguł uniknąć stosowania kilku miar o podobnym zachowaniu, co mogłoby mieć znamiona redundancji i niekorzystnie wpływać na efektywność całego procesu oceny reguł.

Praca [P4] pokazuje jak zaproponowana metoda wizualizacji może być zastosowana do wspierania analizowania miar konfirmacji pod kątem posiadanych przez nie właściwości. Dla takich właściwości jak: monotoniczności M, Ex1, weak Ex1, logiczności L, weak L, maximality/minimality, czy zestawu właściwości symetrii zostały opracowane i opisane konkretne techniki wskazujące jak wizualizacja może być wykorzystana do identyfikowania faktu spełniania/niespełniania ich przez miary. Niejednokrotnie, już sama wizualna analiza zewnętrznych warstw czworościanu (nawet bez umożliwianego przez narzędzia implementujące proponowaną metodę wizualizacji badania wnętrza czworościanu) ujawniała kontrprzykłady dla konkretnych właściwości, czyli zachowania się miary naruszające warunki definiowane przez właściwości. To pozwala wnioskować o nieposiadaniu przez miarę konkretnych właściwości, przez co ogólnie proces analizowania jej cech może być znacząco usprawniony.

Warto również podkreślić, szczególną użyteczność proponowanej metody wizualizacji i technik weryfikowania spełniania przez miary konkretnych właściwości w kontekście miar konfirmacji definiowanych jako agregacje innych miar konfirmacji (np. miar c_1, c_2, c_3 , czy c_4 formułowanych w oparciu o miary A oraz Z w pracy [P1]). W tych przypadkach, analizy teoretyczne wymagałyby rozważania bardzo wielu przypadków, a wręcz pewnej hierarchii przypadków, byłyby zatem czasochłonne i żmudne.

Opracowana w pracach [P3, P4] metoda wizualizacji doczekała się różnych implementacji, które od implementacji w MATLAB'ie ewoluowały w końcu do aplikacji przeglądarkowej [Brz17], która w tej chwili pozwala już na wspieranie analizowania nie tylko miar konfirmacji, ale dowolnych innych współczynników definiowanych w oparciu o macierze kontyngencji, np. popularnych miar oceny klasyfikatorów (F-score, G-mean [He13, Jap11]). Użytkownik ma możliwość zdefiniowania swoich własnych miar, przez co wspierany jest proces ich projektowania. Narzędzie wspiera też analizę miar parametrycznych (np. $c_1, c_2, F_\beta, IBA_\alpha(G\text{-mean})$ [P1, Jap11, Gar09]), pozwalając m.in. badać wpływ parametrów na zachowanie się miar i spełnianie konkretnych właściwości, co stanowi ciekawą i bardzo użyteczną alternatywę lub uzupełnienie rozmaitych rozważań teoretycznych.

H3.Zastosowania miar konfirmacji o pożądanym właściwościach do oceny reguł wyindukowanych z danych

Praktyczna użyteczność miar konfirmacji o pożądanym właściwościach w kontekście oceny reguł wyindukowanych z danych została potwierdzona eksperymentami obliczeniowymi opisanymi w publikacjach [P7, P8]. W szczególności, w pracy [P7] zaproponowano wykorzystanie miar konfirmacji S [Chr99] oraz N [Noz81] do zawężania zbioru reguł

generowanych przez klasyfikator BRACID [Nap12], specjalizujący się w indukcji reguł dla danych niezbalansowanych. Korzystne wyniki w [P7] zainspirowały nas do podjęcia szerszych badań eksperymentalnych dotyczących używania miar confirmacji w klasyfikacji binarnej i wieloklasowej, oraz dla danych zbalansowanych i niezbalansowanych, które zostały podsumowane w publikacji [P8].

Indukcja reguł ukierunkowana na predykcję jest zadaniem często rozważanym, dlatego w literaturze jest wiele propozycji algorytmów generowania reguł klasyfikacyjnych [Fur12]. Jednak generowanie reguł z danych, w których występuje duża dysproporcja pomiędzy licznosciami obiektów w poszczególnych klasach (czyli indukcja reguł dla danych niezbalansowanych), jest problemem szczególnie wymagającym. Trafne rozpoznawanie nowych obiektów z klasy mniejszościowej jest często bardzo ważne dla eksperta, z uwagi na specyfikę tej klasy (np. są to dane dotyczące chorych pacjentów, czy nietypowych zachowań klientów bankowych). Zaproponowany w literaturze algorytm BRACID (Bottom-up induction of Rules And Cases for Imbalanced Data) [Nap12] został zaprojektowany tak, by działać szczególnie dobrze właśnie w kontekście danych niezbalansowanych. Eksperymentalne weryfikacje użyteczności BRACID'a pokazały, że jest on znacząco lepszy niż wiele popularnych klasyfikatorów dostosowanych do indukcji reguł z danych niezbalansowanych [Nap12].

Chcąc rozciągnąć skuteczność BRACID'a z perspektywy predykcyjnej na deskrypcyjną, problemem okazała się bardzo duża liczba reguł generowana przez to podejście. Stała się ona barierą dla ekspertów (m.in., lekarzy), którzy zostali przytłoczeni rozmiarem zbioru reguł. W kontekście konkretnych zastosowań, sama skuteczność predykcyjna okazała się niewystarczająca, bo poza nią użytkownicy chcieli też mieć wgląd w zbiór reguł, analizować go i interpretować poszczególne reguły.

W naturalny sposób zrodziła się zatem potrzeba zwiększenia interpretowalności zbioru reguł generowanych przez BRACID'a, tak by był on mniejszy, ale składał się z reguł ukierunkowanych na perspektywę deskrypcyjną, czyli reprezentujących w interesujący i przystępny sposób regularności charakteryzujące dane. Celem pracy [P7] stało się poszukiwanie strategii redukcji zbioru reguł wyindukowanych z BRACID'a, które zwiększyłyby interpretowalność reguł, nie obniżając jednak drastycznie jego skuteczności predykcyjnej. Warto podkreślić, że choć trafność klasyfikacyjna pełniła jedynie rolę drugorzędowego kryterium, celowo nie pominięto jej zupełnie, by zachować pewną równowagę między przydatnością predykcyjną i deskrypcyjną.

Zaproponowane w pracy [P7] strategie oparte zostały na wykorzystaniu miar confirmacji o pożądanych właściwościach do zawężania zbioru reguł. Spośród wielu miar confirmacji do eksperymentów wybrano dwie miary: S [Chr99] oraz N [Noz81], które jako nieliczne charakteryzują się właściwościami monotoniczności M, maximality/minimality, weak Ex1, weak L, czy korzystnymi symetriami HS, ES oraz EHS. To właśnie posiadanie tych właściwości gwarantuje, że miary będą zachowywały się zgodnie z przyjętą racjonalnością i ich zachowanie będzie łatwo interpretowalne dla eksperta. Przykładowo, miary będą przez swoje wartości odpowiednio odzwierciedlały sytuacje wzrostu liczby przykładów pozytywnych albo wzrostu liczby kontrprzykładów, będą przyjmowały wartości ekstremalne tylko w określonych właściwościach warunkach, czy wreszcie będą symetrycznie traktowały reguły,

w których zaprzeczona jest konkluzja/przesłanka. Dodatkową motywacją do skorzystania z miar S oraz N był fakt, że same ich definicje są stosunkowo łatwo interpretowalne.

W ramach strategii wyboru podzbioru reguł wygenerowanych przez BRACID'a, poza miarami confirmacji wzięto również pod uwagę *wparcie* reguł (czyli liczbę obiektów spełniających przesłankę i konkluzję reguły) i inspiracje pochodzące z odkrywania skupień, a pozwalające ukierunkować proces wyboru na reguły opisujące podgrupy [Gam02]. W ten sposób strategię zaproponowaną w pracy [P7] umożliwiają wybieranie reguł dobrze reprezentowanych w danych (o silnym *wsparciu*), o dobrych cechach deskrypcyjnych (o wysokich wartościach miar confirmacji) oraz pokrywających różne obszary zbioru danych.

W serii eksperymentów na 11 zbiorach danych z repozytorium UCI [Lic13] opracowane strategię wykorzystujące miary S oraz N porównane zostały ze strategią bazującą tylko na *wsparciu* reguł oraz z podejściem bez ograniczania zbioru reguł. Analiza wyników tych badań pokazuje, że zaproponowane strategię rzeczywiście pozwalają na korzystne zawężenie zbioru reguł przy nieznacznym spadku trafności klasyfikacyjnej oraz na wybranie reguł, które średnio charakteryzują się wyższymi wartościami miar confirmacji niż w pozostałych podejściach. Eksperymentalnie potwierdzona została zatem w pracy [P7] użyteczność miar confirmacji do oceny reguł wyindukowanych z danych.

Wyniki badań z wykorzystaniem miar S oraz N zainspiowały nas do rozszerzenia prac również na kontekst innych miar confirmacji w klasyfikacji binarnej i wieloklasowej, dla danych zbalansowanych i niezbalansowanych. W efekcie, w publikacji [P8] skoncentrowano się na problemie zawężania zbioru reguł, stanowiącego klasyfikator regułowy, przy jednoczesnym dbaniu, by wybrane reguły były korzystne w ujęciu deskrypcyjnym. Zaproponowano wykorzystanie miar confirmacji (jako miar definiowanych do oceny pojedynczych reguł z perspektywy deskrypcyjnej) do sortowania i ograniczania generowanego zbioru reguł. Wpływ użycia miar confirmacji w klasyfikacji regułowej badano poprzez nowo opracowany algorytm CM-CAR (Confirmation Measure Class Association Rules), uogólniający metodę CBA [Liu98]. Potrzeba zaproponowania nowego algorytmu wynikała z faktu, że popularne klasyfikatory nastawione są zwykle bardziej na optymalizację trafności klasyfikacji, czy pokrycia zbioru uczącego, niż na cechy deskrypcyjne generowanych reguł.

Algorytm CM-CAR w oparciu o miary confirmacji generuje określoną przez użytkownika liczbę reguł. Zaczyna on od znalezienia wszystkich zbiorów częstych (minimalne *wsparcie* jest progiem określanym na wejściu przez użytkownika), z których tworzy reguły decyzyjne, czyli takie, które mają w konkluzji jedynie pewien wyróżniony atrybut (decyzję). Następnie wykorzystuje dwie listy miar atrakcyjności, zdefiniowane przez użytkownika, by odpowiednio sortować i filtrować na ich podstawie reguły. Według wartości miar z pierwszej listy reguły są uporządkowane kolejno. Jeśli zatem w danych pojawi się przykład pokrywający wiele reguł, używana do klasyfikacji jest ta znajdująca się wyżej na liście. Druga lista miar jest natomiast wykorzystywana do tego, by zawęzić zbiór wygenerowanych reguł do k najlepszych pod względem wartości miar z listy. Warto podkreślić, że to właśnie dzięki użyciu dwóch oddzielnych zbiorów miar, algorytm może osobno badać ich wpływ na aspekty predykcyjne i deskrypcyjne końcowego zbioru reguł, stanowiącego model klasyfikacyjny.

Eksperymenty przeprowadzone na 20 różnorodnych (binarnych/wieloklasowych, zbalansowanych/niezbilansowanych) zbiorach danych z repozytorium UCI [Lic13] podzielono na dwie części, z których każda badała 12 miar konfirmacji. W pierwszej kolejności eksperymety skupiały się na ocenie przydatności miar konfirmacji do zawężania zbioru reguł, były zatem prowadzone przy jednej ustalonej liście miar do sortowania (złożonej z miar *ufności*, *wsparcia* i kryterium długości reguły). Lista miar odpowiadających za zawężanie zbioru reguł podlegała natomiast zmianie, tak by po kolei uwzględnić wszystkie 12 analizowanych miar konfirmacji. Warto podkreślić, że ustalenie listy sortującej powoduje, że na różnice w modelu klasyfikacyjnym wpływać mogą jedynie miary z drugiej listy, co pozwala na porównywanie miar konfirmacji pod kątem przydatności do zawężania zbioru reguł. Wersję referencyjną (porównawczą) algorytmu stanowiło podejście korzystające przy zawężaniu zbioru reguł z miary *ufności*, czyli w praktyce sprowadzające się do metody CBA proponowanej w [Liu98].

W drugiej części eksperymentów nacisk położony był na weryfikację użyteczności miar konfirmacji w kontekście predykcyjnym, a zatem obie listy (do sortowania i do filtrowania) przygotowane były tak, by jedną z analizowanych miar konfirmacji używać jako czynnik decydujący zarówno o aspektach predykcyjnych jak i deskrypcyjnych. Analogicznie jak w pierwszej części eksperymentów, miara *ufności* stanowiła swoisty punkt referencyjny.

Modele klasyfikacyjne utworzone w ramach wszystkich eksperymentów oceniane były za pomocą kilku współczynników oceny klasyfikatorów. Wyniki badań podsumowane w pracy [P8] pokazują, że miary c_1 [P1], F [Kem52] oraz Z [Cru07] ogólnie dobrze sprawdzają się jako miary do ograniczania zbioru reguł i sortowania, a dodatkowo w kontekście problemów binarnych i danych niezbilansowanych wyróżniały się także miary N [Noz81], S [Chr99] oraz c_3 [P1].

Praca [P8] potwierdza eksperymentalnie ogólną użyteczność miar konfirmacji w ocenie reguł wyindukowanych z danych o różnej charakterystyce. Wskazuje, które miary zasługują na szczególną uwagę, jeśli celem jest otrzymanie zawężonego zbioru reguł o dobrych cechach deskrypcyjnych, który sprawdza się także w perspektywie predykcyjnej.

Wyniki podsumowane w publikacjach [P7, P8] stanowią potwierdzenie praktycznej użyteczności rozważań i analiz formalnych w pracach [P1, P2, P5, P6] oraz analiz wspieranych wizualizacjami w [P3, P4].

Podsumowanie

Badania przedstawione w ramach cyklu publikacji [P1]-[P8], wykorzystując metody informatyczne, wnoszą istotny wkład w rozwój metodyki odkrywania wiedzy. Dokonano w nich wnikliwej analizy wielu właściwości miar oceny reguł i zaproponowano nowe właściwości korygujące istniejące podejścia i uzupełniające je. Określono zależności między właściwościami, wspomagające badanie właściwości miar i projektowanie nowych. Przeanalizowano też popularne miary konfirmacji pod względem ich właściwości i zdefiniowano nowe miary atrakcyjne z uwagi na ich właściwości. Opracowano metody wizualizacji miar wspierające badanie ich cech i właściwości oraz projektowanie nowych

miar. Praktyczną użyteczność miar konfirmacji do oceny reguł wyindukowanych z danych potwierdzono poprzez szereg eksperymentów na danych o różnej charakterystyce.

Podstawowe elementy oryginalnego wkładu naukowego zawarte w ośmiu publikacjach wchodzących w skład przedstawionego cyklu są następujące:

- zaprezentowano usystematyzowane omówienie różnych perspektyw konfirmacji wraz z uzasadnieniem formalnym równoważności między nimi i sprowadzeniem ich do ogólnej definicji właściwości konfirmacji [P5];
- zaproponowano uogólnienie właściwości Ex1 oraz L do właściwości weak Ex1 oraz weak L, a także sformułowano potencjalnościowe odpowiedniki właściwości Ex1 oraz L [P1];
- opracowano cztery alternatywne metody normalizacji miar konfirmacji i przedstawiono ich zastosowanie w popularnych miarach konfirmacji [P1];
- przeprowadzono analizę znormalizowanych miar konfirmacji pod kątem posiadanych przez nie właściwości oraz zależności między nimi [P1];
- zaproponowano miarę konfirmacji A będącą potencjalnościowym odpowiednikiem miary Z z [Cru07], a następnie cztery miary konfirmacji (c_1, c_2, c_3, c_4) bazujące na komplementarności miar Z i A [P1];
- dokonano zestawienia i krytycznego porównania dotychczasowych podejść do wartościowania właściwości symetrii proponowanych w literaturze i zaproponowano nowy podział właściwości symetrii na pożądane i niepożądane [P2];
- przeanalizowano miary konfirmacji w kontekście nowej propozycji wartościowania symetrii, a także pod kątem innych pożądanych właściwości (m.in., monotoniczności M, weak Ex1, weak L) [P2];
- zaproponowano nowe spojrzenie na dyskutowane w literaturze zestawy właściwości symetrii poprzez pryzmat grup algebraicznych, co pozwoliło na zidentyfikowanie w nich i rozwiązanie problemów niekompletności i niespójności [P6];
- wykazano, że istnieje tylko jeden podział symetrii na pożądane i niepożądane, który ma interpretację w kontekście oceny reguł wyindukowanych z danych [P6];
- zaprezentowano zestaw nowych właściwości monotoniczności miar względem perspektyw konfirmacji i udowodniono zależności pomiędzy nimi [P5];
- przedstawiono i udowodniono szereg twierdzeń dotyczących zależności pomiędzy monotonicznością miary względem perspektywy a innymi właściwościami (m.in. monotonicznością M, weak Ex1, weak L, maximality/minimality) [P5];
- wywiedziono szereg zależności pomiędzy monotonicznością miary względem perspektywy a rozważanymi w literaturze właściwościami symetrii [P5];
- dokonano analizy miar konfirmacji pod kątem ich monotoniczności względem różnych perspektyw oraz posiadanych przez nie właściwości [P5];
- zaproponowano i zweryfikowano metodę wizualizacji miar konfirmacji w postaci czworościanów w barycentrycznym układzie współrzędnych, umożliwiając tym samym pozyskiwanie informacji o zachowaniu się miary we wszystkich obszarach jej dziedziny [P3, P4];
- przedstawiono sposoby wykorzystania zaproponowanej wizualizacji do badania cech charakterystycznych (m.in., ekstrema, wartości zerowe, gradienty) pojedynczych miar oraz grup miar (m.in., różnice między miarami, wariancja w grupach miar) [P3];

- opracowano techniki wykorzystania wizualizacji do wnioskowania o właściwościach miar confirmacji [P4];
- dokonano szczegółowej analizy cech i właściwości miar confirmacji w oparciu o zaproponowaną metodę wizualizacji [P3, P4];
- zaproponowano koncepcję użycia miar confirmacji dla zwiększania interpretowalności reguł wyindukowanych z niezbalansowanych danych [P7];
- opracowano i zaimplementowano strategie zawężania zbioru reguł wykorzystujące miary confirmacji, *wsparcie* reguł oraz inspiracje pochodzące z odkrywania grup [P7];
- potwierdzono praktyczną przydatność miar confirmacji o pożądanych właściwościach do zwiększania interpretowalności reguł i zawężania ich zbioru [P7];
- przeprowadzono weryfikację użyteczności miar confirmacji w ocenie reguł wyindukowanych z danych zbalansowanych/niezbalansowanych oraz dla problemów dwu i wieloklasowych, z wykorzystaniem opracowanego na potrzeby tego zadania algorytmu CM-CAR [P8];
- potwierdzono użyteczność miar confirmacji do generowania zawężonego zbioru reguł łączącego wymagania aspektów deskrypcyjnych i predykcyjnych [P8].

Literatura

- [Agr93] Agrawal R., Imielinski T., Swami A., Mining associations between sets of items in massive databases, in: *Proc. of the 1993 ACM-SIGMOD International Conference on Management of Data*, 207-216, (1993).
- [Brz17] Brzezinski D., Stefanowski J., Susmaga R., Szczęch I., *Tetrahedron: Barycentric Measure Visualizer*, praca przyjęta do publikacji na konferencji ECML PKDD 2017 w Skopie (Macedonia).
- [Chr99] Christensen D., *Measuring confirmation*. *Journal of Philosophy* 96, 437-461 (1999).
- [Cru07] Crupi V., Tentori K., Gonzalez M., On Bayesian measures of evidential support: Theoretical and empirical issues, *Philosophy of Science*, 74, 229–252 (2007).
- [Eel02] Eells E., Fitelson B., Symmetries and asymmetries in evidential support, *Philosophical Studies* 107 (2) 129–142 (2002).
- [Fay99] Fayyad U.M., Piatetsky-Shapiro G., Smyth P., Uthurusamy R., *Advances in knowledge discovery and data mining*, AAAI/MIT Press, (1999).
- [Fit01] Fitelson B., *Studies in Bayesian Confirmation Theory*. Ph.D. thesis, University of Wisconsin, Madison, (2001).
- [Flo06] Floater M.S., Hormann K., Kos G., A general construction of barycentric coordinates over convex polygons, *Advances in Computational Mathematics* 24(1–4): 311–331 (2006).
- [Fre99] Freitas A., On rule interestingness measures. *Knowledge-Based Systems*, 12, 309–315 (1999).
- [Fur12] Furnkranz J., Gamberger D., Lavrac N., *Foundations of Rule Learning*, Springer Verlag (2012).
- [Gam02] Gamberger D., Lavrac N., Expert-Guided Subgroup Discovery: Methodology and Application. *Journal of Artificial Intelligence Research*, 17(1), 501–527 (2002).

- [Gar09] Garcia V., Mollineda R. A., Sanchez J. S., Index of Balanced Accuracy: A Performance Measure for Skewed Class Distributions, in: *Proc. 4th Iberian Conf. Pattern Recognition Image Analysis*, 441–448, (2009).
- [Gen06] Geng L., Hamilton H., Interestingness measures for data mining: A survey. *ACM Computing Surveys*, 38(3) (2006).
- [Gla13] Glass D.H., Confirmation measures of association rule interestingness, *Knowledge-Based Systems* 44, 65–77, (2013).
- [Gre04] Greco S., Pawlak Z., Słowiński R., Can Bayesian confirmation measures be useful for rough set decision rules? *Engineering Applications of Artificial Intelligence* 17 (4) 345–361, (2004).
- [Han00] Han J., Kamber M., *Data mining: Concepts and Techniques*, San Francisco, Morgan Kaufmann, (2000).
- [He13] He H., Ma Y. (Eds.), *Imbalanced Learning: Foundations, Algorithms and Applications*, IEEE - Wiley, (2013).
- [Hum96] Humphreys J.F., *A Course in Group Theory*, Oxford University Press, Cambridge, UK, (1996).
- [Jap11] Japkowicz N., Shah M., *Evaluating Learning Algorithms: A Classification Perspective*, Cambridge University Press, ISBN 9780521196000, (2011).
- [Kem52] Kemeny J., Oppenheim P., Degrees of factual support. *Philosophy of Science* 19, 307-324, (1952).
- [Lic13] Lichman M., UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science, (2013).
- [Liu98] Liu B., Hsu W., Ma Y., Integrating classification and association rule mining. In: *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining*. pp. 80-86 (1998).
- [Mah05] Maher P., *Confirmation Theory*. The Encyclopedia of Philosophy, second ed., Macmillan Reference, USA, (2005).
- [McG05] McGarry, K., A survey of interestingness measures for knowledge discovery. *The Knowledge Engineering Review*, 20(1), 39–61 (2005).
- [Mor88] Mortimer H., *The Logic of Induction*. Paramus, Prentice Hall (1988).
- [Nap12] Napierała K., Stefanowski J., BRACID: a comprehensive approach to learning rules from imbalanced data, *Journal of Intelligent Information Systems*, 39(2), 335–373 (2012).
- [Noz81] Nozick R., *Philosophical Explanations*. Clarendon Press, Oxford, UK (1981).
- [Pia91] Piatetsky-Shapiro G., Matheus C., *Knowledge discovery in databases*, AAAI/MIT Press, (1991).
- [Pop59] Popper K.R., *The Logic of Scientific Discovery*, Hutchinson, London, (1959).
- [Ste01] Stefanowski J., *Algorytmy indukcji reguł decyzyjnych w odkrywaniu wiedzy*, Politechnika Poznańska, Instytut Informatyki, (2001).

- [Ste01b] Stefanowski J., Vanderpooten D., Induction of decision rules in classification and discovery-oriented perspectives. *International Journal of Intelligent Systems*. 16 (1), 13–28, (2001).
- [War04] Ware C., *Information Visualization: Perception for Design, 2nd Edition*, Morgan Kaufmann, Waltham, MA (2004).
- [War03] Warren J., On the uniqueness of barycentric coordinates, in R. Goldman and R. Krasauskas (Eds.), *Topics in Algebraic Geometry and Geometric Modeling*, Contemporary Mathematics, Vol. 334, American Mathematical Society, Providence, RI, USA, 93–99, (2003).

V. OSIĄGNIĘCIA NAUKOWO-BADAWCZE OGÓŁEM

Wyniki moich badań zostały opisane w publikacjach naukowych, których pełne zestawienie przedstawiono poniżej. Lista uwzględnia podział na prace opublikowane przed i po uzyskaniu stopnia doktora. Zestawienie jest chronologiczne, z jednym wyjątkiem: praca [A6] jest publikacją doktoratu w czasopiśmie *Transactions on Rough Sets X* (LNCS series), została zatem ujęta w dorobku „przed doktoratem” mimo, że została wydana rok po uzyskaniu stopnia doktora. Dla każdej publikacji podano wartości współczynnika wpływu (Impact Factor; IF), a także liczbę punktów z listy Ministerstwa Nauki i Szkolnictwa Wyższego (Pkt MNiSW).

- [A1] Izabela Brzezińska, Roman Słowiński
Monotonicity of a Bayesian confirmation measure in rule support and confidence
 In T. Burczyński, W. Cholewa, W. Moczulski (Eds.) *Recent Developments in Artificial Intelligence Methods*,
 AI-METH Series, Gliwice, Poland (2005), 39-42.
 Pkt MNiSW=10; konf.indeks. w WoS
- [A2] Roman Słowiński, Izabela Brzezińska, Salvatore Greco
Application of Bayesian Confirmation Measures for Mining Rules from Support-Confidence Pareto-optimal set
 In: Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., Żurada, J. (eds.), *ICAISC 2006*,
 LNAI, vol. 4029, pp. 1018-1026. Springer, Heidelberg (2006).
 Pkt MNiSW=10; konf.indeks. w WoS
- [A3] Izabela Brzezińska, Salvatore Greco, Roman Słowiński
Mining Pareto-optimal rules with respect to support and confirmation or support and anti-support
Engineering Applications of Artificial Intelligence, Volume 20 no. 5 (2007) pp.587-600.
 Pkt MNiSW(2007)=20 ; IF(2007)=0,762
- [A4] Roman Słowiński, Izabela Szczęch, Mirosław Urbanowicz, Salvatore Greco
Mining Association Rules with respect to Support and Anti-support - Experimental Results
 In: Kryszkiewicz, M., Peters, J.F., Rybinski, H., Skowron, A. (eds.) *RSEISP2007*
 LNAI, vol. 4585, pp. 534-542. Springer, Heidelberg (2007).
 Pkt MNiSW=10; konf.indeks. w WoS

- [A5] Savatore.Greco, Roman Słowiński, Izabela Szczęch
Analysis of monotonicity properties of some rule interestingness measures
 In: Materiały II Krajowej Konferencji nt. *Technologie Przetwarzania Danych*. Poznań, 24-26.09.2007,
 Wydawnictwo Politechniki Poznańskiej, Poznań, 2007, str. 151-161.
 Pkt MNiSW=0 ; konf. nie indeks. w WoS
- [A6] Izabela Szczęch
Multicriteria attractiveness evaluation of decision and association rules
Transactions on Rough Sets X (LNCS series), vol. 5656/2009, pp. 197-274, Springer, Heidelberg (2009).
 Pkt MNiSW(2009)=10 ; IF(2009)= brak
- przed uzyskaniem stopnia doktora ↑
-
- po uzyskaniu stopnia doktora ↓
- [A7] Salvatore Greco, Roman Słowiński, Izabela Szczęch,
Assessing the quality of rules with a new monotonic interestingness measure Z
 In: Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.), *Artificial Intelligence and Soft Computing (ICAISC 2008)*,
 LNAI, vol. 5097, pp. 556-565. Springer, Heidelberg (2008).
 Pkt MNiSW=10; konf. indeks w WoS
- [A8] Roman Słowiński, Salvatore Greco, Izabela Szczęch
Analysis of monotonicity properties of new normalized rule interestingness measures
 In: Brezillon, P., Coppin, G., Lenca, Ph. (eds), *Human Centered Processes*. Proc. of the 3rd International Conference HCP-2008, Delft, 8-12 June 2008, pp.231-242.
 Pkt MNiSW=0; konf. nie indeks. w WoS
- [A9] Salvatore Greco, Roman Słowiński, Izabela Szczęch
Analysis of monotonicity properties of some rule interestingness measures
Control and Cybernetics vol. 38 (2009) No.1 pp. 9-25.
 Pkt MNiSW(2009)=15 ; IF(2009)=0,378
- [A10] Salvatore Greco, Roman Słowiński, Izabela Szczęch
Alternative normalization schemas for Bayesian confirmation measures.
 In: Hüllermeier, E., Kruse, R., Hoffmann, F. (eds.), *Computational Intelligence for Knowledge-Based Systems Design (IPMU 2010)*, LNAI, vol. 6178, pp. 230-239. Springer-Verlag, Heidelberg (2010).
 Pkt MNiSW= 10 ; konf. indeks. w WoS
- [A11] Izabela Szczęch, Salvatore Greco, Roman Słowiński
New property for rule interestingness measures
 In: Ganzha, M., Maciaszek, L., Paprzycki, M. (eds), *Proceedings of the Federated Conference on Computer Science and Information Systems FedCSIS 2011*, Szczecin, 18-21 September 2011, pp.103-108 (ISBN 978-83-60810-39-2).
 Pkt MNiSW=0; konf. nie indeks. w WoS (indeksowane 2012,2013,2014,2015,2016)

- [P1] Salvatore Greco, Roman Słowiński, Izabela Szczęch
Properties of rule interestingness measures and alternative approaches to normalization of measures
Information Sciences vol. 216 (2012), 1–16.
Pkt MNiSW(2012)=40 ; IF(2012)=3,643
- [A12] Salvatore Greco, Roman Słowiński, Izabela Szczęch
Analysis of symmetry properties for Bayesian confirmation measures.
In: Li, T., Nguyen, H.S., Wang, G., Grzymała-Busse, J., Janicki, R., Hassanién, A.E., Yu, H. (eds.) *RSKT 2012 LNAI*, vol. 7414, pp. 207-214. Springer-Verlag Berlin Heidelberg (2012)
Pkt MNiSW=0; konf. nie indeks. w WoS (indeksowane 1st, 2nd, 3rd, 4th, 6th, 8th, 9th, 10th)
- [A13] Robert Susmaga, Izabela Szczęch
Statistical significance of Bayesian confirmation measures
Research Report RA-10/12, Institute of Computing Science, Poznań University of Technology, Poznań, (2012).
Pkt MNiSW=0; nie indeks. w WoS
- [A14] Robert Susmaga, Izabela Szczęch
The Property of χ^2_{01} -Concordance for Bayesian Confirmation Measures
In: Torra, V., Narukawa, Y., Navarro-Arribas, G., Megias, D. (eds.) *Modeling Decisions for Artificial Intelligence. Proceedings of the 10th International Conference MDAI 2013*, Barcelona, 20-22 November 2013, LNCS, vol. 8234, pp. 226-236. Springer (2013).
Pkt MNiSW=0; konf. nie indeks. w WoS (indeksowane 7th, 8th, 11th, 13th)
- [A15] Robert Susmaga, Izabela Szczęch
Visualization of interestingness measures
In: Proceedings of the 6th Language & Technology Conference: *Human Language Technologies as a Challenge for Computer Science and Linguistics*, Poznań, 07-09 December 2013, Fundacja UAM, Poznań, pp. 95–99 (2013).
Pkt MNiSW=15; konf. indeks. w WoS
- [P2] Salvatore Greco, Roman Słowiński, Izabela Szczęch
Finding Meaningful Bayesian Confirmation Measures
Fundamenta Informaticae vol. 127 (2013), 161–176.
Pkt MNiSW(2017)=20 ; IF(2013)=0,479
- [A16] Robert Susmaga, Izabela Szczęch
Visual-Based Detection of Properties of Confirmation Measures
In: Andreasen, T., Christiansen, H., Cubero Talavera J.C., Raś, Z.W. (eds.) *Foundations of Intelligent Systems. Proceedings of the 21st International Symposium on Methodologies for Intelligent Systems*, ISMIS 2014, Roskilde, Denmark, June 25-27, 2014. LNCS, vol. 8502, pp. 133-143. Springer (2014).
Pkt MniSW=0; konf. nie indeks. w WoS (indeksowane 6th, 12th, 13th, 18th, 19th)
- [A17] Salvatore Greco, Roman Słowiński, Izabela Szczęch
Four perspectives of confirmation
Research Report RA-01/14, Institute of Computing Science, Poznań University of Technology, Poznań, (2014).
Pkt MNiSW=0; nie indeks. w WoS

- [P3] Robert Susmaga, Izabela Szczęch
Can Interestingness Measures Be Usefully Visualized?
International Journal of Applied Mathematics and Computer Science vol. 25, no. 2 (2015), 323–336.
Pkt MNiSW(2017)=25 ; IF(2015)=1,037
- [P4] Robert Susmaga, Izabela Szczęch
Visualization support for the analysis of properties of interestingness measures
Bulletin of the Polish Academy of Sciences Technical Sciences vol. 63, issue 1 (2015), 315–327.
Pkt MNiSW(2017)=25 ; IF(2015)=1,087
- [P5] Salvatore Greco, Roman Słowiński, Izabela Szczęch
Measures of rule interestingness in various perspectives of confirmation
Information Sciences vol. 346–347C (2016), 216–235.
Pkt MNiSW(2017)=45 ; IF(2016)=4,832
- [P6] Robert Susmaga, Izabela Szczęch
Selected group-theoretic aspects of confirmation measure symmetries
Information Sciences vol. 346–347C (2016), 424–441.
Pkt MNiSW(2017)=45 ; IF(2016)=4,832
- [P7] Krystyna Napierała, Jerzy Stefanowski, Izabela Szczęch
Increasing the Interpretability of Rules Induced from Imbalanced Data by Using Bayesian Confirmation Measures
In: Appice, A., Ceci, M., Loglisci, C., Masciari, E., Ras, Z.W. (Eds.) *New Frontiers in Mining Complex Patterns – 5th International Workshop, NFMCP 2016, Held in Conjunction with ECML-PKDD 2016, Riva del Garda, Italy, September 19, 2016, Revised Selected Papers*. LNAI 10321, 84–98, Springer (2017) .
Pkt MNiSW=0 ; konf. jeszcze nie indeks. w WoS (indeksowane 2nd, 4th)
- [P8] Dariusz Brzezinski, Zbigniew Grudziński, Izabela Szczęch
Bayesian Confirmation Measures in Rule-based Classification
In: Appice, A., Ceci, M., Loglisci, C., Masciari, E., Ras, Z.W. (Eds.) *New Frontiers in Mining Complex Patterns – 5th International Workshop, NFMCP 2016, Held in Conjunction with ECML-PKDD 2016, Riva del Garda, Italy, September 19, 2016, Revised Selected Papers*. LNAI 10321, 39–53, Springer (2017).
Pkt MNiSW=0 ; konf. jeszcze nie indeks. w WoS (indeksowane 2nd, 4th)
- [A18] Dariusz Brzeziński, Jerzy Stefanowski, Robert Susmaga, Izabela Szczęch
Tetrahedron: Barycentric Measure Visualizer
praca przyjęta do publikacji na konferencji ECML PKDD 2017 w Skopie (Macedonia).
- [A19] Dariusz Brzeziński, Jerzy Stefanowski, Robert Susmaga, Izabela Szczęch
Visual-based analysis of classification measures with applications to imbalanced data
ArXiv e-prints (2017), arXiv:1704.07122
publikacja zgłoszona do czasopisma *Information Sciences* 21.07.2017r.

VI. WSKAŹNIKI BIBLIOMETRYCZNE

A. PODSUMOWANIE ZE WZGLĘDU NA IMPACT FACTOR

- Sumaryczny *2-letni Impact Factor* czasopism według listy Journal Citation Reports zgodny z rokiem opublikowania wynosi **17,05**;

B. PODSUMOWANIE ZE WZGLĘDU NA PUNKTY MINISTERSTWA NAUKI I SZKOLNICTWA WYŻSZEGO

- Sumaryczna liczba punktów MNiSW za publikacje wynosi **310** (w tym 240 po uzyskaniu stopnia doktora);
- Liczba publikacji wg odpowiadających im punktów MNiSW:
45p. - 2 publikacje, 40p. - 1 publikacja, 25p. - 2 publikacje,
20p. - 2 publikacje, 15p. - 2 publikacje, 10p. - 6 publikacji.

C. LICZBA CYTOWAŃ (wg danych z 12.07.2017r.)

- Web of Science: **118** (w tym 74 cytowania obce);
- Scopus: **144** (w tym 76 cytowań obcych);
- Google Scholar: **188**.
-

D. INDEKS HIRSCHA (wg danych z 12.07.2017r.)

- Web of Science: **5**;
- Scopus: **6**;
- Google Scholar: **8**.

VII. PORÓWNANIE TEMATYKI PRAC OPUBLIKOWANYCH PRZED I PO UZYSKANIU STOPNIA DOKTORA

Tematyka podejmowana w pracach opublikowanych *przed uzyskaniem stopnia doktora* koncentrowała się wokół zagadnień eksploracji danych i uczenia maszynowego, a w szczególności oceny reguł decyzyjnych i asocjacyjnych ze względu na wiele kryteriów jednocześnie. W centrum zainteresowania stawiana była wielokryterialna ocena reguł z jednakową konkluzją, weryfikująca reguły pod względem różnych miar atrakcyjności jednocześnie (nie ograniczonych do miar confirmacji).

W ramach tych prac m.in.:

- przeanalizowano właściwości wybranych miar oraz wykazano zachodzenie związków zawierania się między zbiorami reguł niezdominowanych względem wybranych miar [A1, A2, A3, A5, A6];
- zaproponowano nowe podejście do wielokryterialnej oceny reguł z jednakową konkluzją oparte na mierze *wsparcia* i *anty-wsparcia*. Jak wykazano w rozprawie, podejście to gwarantuje, że zbiór reguł Pareto-optimalnych w przestrzeni *wsparcie-anty-wsparcie* zawiera wszystkie reguły optymalne ze względu na dowolną miarę atrakcyjności z pożądaną właściwością monotoniczności M (zaproponowaną w [Gre04]) [A3, A4, A6];

- przedstawiono metodę określania w przestrzeni *wsparcie–anty-wsparcie* obszaru reguł z dodatnią wartością konfirmacji, co pozwoliło na analizę reguł interesujących z punktu widzenia minimalnego *wsparcia* reguły, maksymalnego dopuszczalnego *anty-wsparcia*, dodatniej konfirmacji i właściwości monotoniczności M [A3, A4, A6];
- zaprezentowano i zweryfikowano system wielokryterialnej oceny reguł, oparty na modyfikacji schematu apriori [Agr93], umożliwiając generowanie reguł z uwzględnieniem różnych miar atrakcyjności [A6].

W cyklu *habilitacyjnym* omówiono w sposób kompleksowy zagadnienia dotyczące analizy i projektowania miar konfirmacji o pożądanym właściwościach z punktu widzenia oceny reguł wyindukowanych z danych. Sformułowano szereg twierdzeń dotyczących miar i ich właściwości, opracowano wizualne metody wspierania badania miar oraz przeprowadzono wiele eksperymentów potwierdzających użyteczność miar konfirmacji w kontekście oceny reguł wyindukowanych z danych. Najważniejsze elementy oryginalnego wkładu naukowego, które nie znajdują odpowiedników we wcześniejszych pracach, dotyczą:

- przeprowadzenia dogłębnej analizy właściwości miar konfirmacji i opracowania nowych właściwości oraz istotnych modyfikacji już istniejących [P1, P2, P5, P6];
- zaproponowania właściwości monotoniczności miar względem perspektyw, łączących aspekty jakościowe wyrażane przez perspektywy z ilościowymi wyrażanymi przez definicje miar [P5];
- formalnego określenia relacji między właściwościami, umożliwiającego wnioskowanie o spełnianiu przez miarę pewnych właściwości na podstawie posiadania przez nią innych [P5];
- rozszerzenia analizy miar konfirmacji o nowe właściwości [P1, P5];
- przedstawienia oryginalnego spojrzenia na dyskutowane w literaturze właściwości symetrii poprzez pryzmat grup algebraicznych, które prowadzi do formalnych wniosków, niezależnych od dotychczasowych, niejednoznacznych rozważań różnych autorów na temat zbiorów symetrii i wartościowania symetrii w obrębie tych zbiorów [P6];
- zaproponowania nowych miar konfirmacji o pożądanym właściwościach [P1, P5];
- opracowania metody wizualizacji miar konfirmacji z wykorzystaniem układu barycentrycznego, która usprawnia badanie cech i właściwości miar, a przez to znacząco ułatwia proces doboru miar do konkretnych zastosowań oraz projektowanie nowych miar o pożądanym zachowaniu [P3, P4];
- eksperymentalnej weryfikacji użyteczności miar konfirmacji do oceny reguł wyindukowanych z danych, a w szczególności określenia, które miary szczególnie dobrze sprawdzają się przy zadaniach ograniczania zbioru reguł tak, by bez pominięcia aspektów deskrypcyjnych dobrze sprawdzały się też w perspektywie predykcyjnej [P7, P8];

VIII. OMÓWIENIE NAJWAŻNIEJSZYCH CELÓW NAUKOWYCH PRAC STANOWIĄCYCH POZOSTAŁY DOROBEK NAUKOWY

Prace, których nie ujęto w cyklu stanowiącym osiągnięcie habilitacyjne, podnoszą tematy teoretyczne oraz praktyczne dotyczące miar i ich właściwości, często blisko związane z problemami poruszonymi w artykułach [P1]-[P8]. W szczególności dotyczą:

- analizy właściwości miar [A7, A8, A9, A11, A17];
- propozycji metod normalizacji miar [A8, A10];
- badania właściwości symetrii [A12];
- propozycji statystycznych właściwości miar [A13, A14];
- propozycji metod wizualizacji miar [A15, A16];
- propozycji wspieranej poprzez wizualizacje analizy miar oceny klasyfikatorów [A18, A19].

Tematyka prac [A13, A14] jest zbieżna z tematyką omawianą w prezentowanym cyklu publikacji, bo dotyczy właściwości miar konfirmacji, spogląda na nie jednak przez pryzmat statystycznych zależności [Has03]. W szczególności przedmiotem tych publikacji jest badanie zgodności miar konfirmacji ze statystycznie istotną zależnością pomiędzy występowaniem przesłanki i konkluzji reguły. Istniejące miary nie były bowiem zwykle definiowane z zamiarem posiadania takiej zgodności. Tego typu rozważania są istotne, gdy analizowane dane są obciążone błędami obserwacyjnymi. Może wówczas dochodzić do bardzo niepożądanych sytuacji, w których analizowana miara zaniża lub zawyża oceny reguł. Przykładowo, w sytuacjach gdy zależność w danych jest słaba, miara wskazująca silną konfirmację (lub dyskonfirmację) mogłaby skłaniać użytkownika do podejmowania działań/decyzji w oparciu o wysoką wartość miary, mimo, że nie byłyby one uzasadnione faktycznymi zależnościami w danych.

W tym kontekście prace [A13, A14] proponują współczynnik oceniający zależność pomiędzy występowaniem przesłanki a konkluzji w danych eksperymentalnych, i wprowadzają sposób kwantyfikowania poziomu zgodności (konkordancji) tego współczynnika z analizowaną miarą konfirmacji. Odstępstwa od idealnego poziomu zgodności mogą być dokładnie badane i wizualizowane dzięki przedstawianym w pracach wykresach rozrzutu oraz specjalizowanym histogramom. Pozwoliło to następnie nadać miarom interpretacje w kategoriach ryzyka: miara konfirmacji, która zaniża oceny reguł jest nazywana miarą wykazującą awersję do ryzyka, podczas gdy miara, która zawyża oceny reguł jest nazywana miarą wykazującą skłonność do ryzyka. Eksperymentalne badanie zgodności popularnych miar konfirmacji ze zdefiniowanym współczynnikiem zależności pomiędzy występowaniem przesłanki a konkluzji, uzupełnia charakterystykę tych miar oraz może pozytywnie wpływać na sposoby definiowania nowych miar w przyszłości.

Spójny tematycznie zestaw prac w ramach pozostałego dorobku naukowego stanowią też publikacje [A18, A19]. Proponują one zaadaptowanie metod wizualizacji miar konfirmacji proponowanych w [P3, P4] do miar oceny klasyfikatorów. Tematyka dotyczy zatem jednego z najważniejszych elementów uczenia nadzorowanego, czyli złożonego problemu wyboru miary do oceny klasyfikatorów [Dom12, Jap11]. Podobnie jak w kontekście

miar oceny reguł, również wybór współczynnika oceniającego klasyfikator często dokonywany jest automatycznie, bez wnikliwej analizy właściwości. Problem odpowiedniej oceny klasyfikatora przybiera na znaczeniu szczególnie dla problemów, w których występują znaczące dysproporcje pomiędzy licznościami klas [HeG09]. Liczba miar proponowanych w literaturze do oceny klasyfikatorów w kontekście danych niezbalansowanych jest duża, rodzi się zatem potrzeba określania ich właściwości, dopiero bowiem poznanie charakterystyki zachowania się miar pozwoli na świadome wybranie miary do konkretnego problemu klasyfikacyjnego.

W pracach [A18, A19] przedstawiono możliwości wizualizowania miar oceny klasyfikatorów z wykorzystaniem układu barycentrycznego [War03]. W szczególności, zaproponowano sposób tworzenia specyficznych przecięć czworościanu reprezentującego miarę, których analiza pozwala wnioskować o zmianach w zachowaniu się miary, gdy zmiana ulega proporcja między klasami w danych. Jest to istotne np. w zastosowaniach dotyczących strumieni danych, gdzie niezbalansowanie może ulegać zmianie i ważna jest wiedza o swoistej odporności miary na zmiany proporcji klas.

Dodatkowo, praca [A19] podejmuje problem porównania miar oceny klasyfikatorów, by ułatwić proces wyboru miary do konkretnego zastosowania. Proponowana jest lista dziesięciu nowych właściwości szczególnie istotnych w kontekście danych niezbalansowanych. Popularne miary oceny klasyfikatorów (zarówno nieparametryczne jak i parametryczne) są weryfikowane pod względem posiadanych właściwości z wykorzystaniem wizualizacji. Dodatkowo dla miar parametrycznych, analitycznie wskazano progowe wartości parametrów warunkujące posiadanie konkretnych właściwości przez miary. W efekcie uwypuklone są różnice pomiędzy miarami, które budują wiedzę konieczną do trafnego wyboru miary.

Literatura

- [Dom12] Domingos P.M., A few useful things to know about machine learning, *Communications of the ACM* 55 (10) 78-87, (2012).
- [Gre04] Greco S., Pawlak Z., Słowiński R., Can Bayesian confirmation measures be useful for rough set decision rules, *Engineering Applications of Artificial Intelligence* 17 (4) 345–361, (2004).
- [Has03] Hastie T., Tibshirani R., Friedman J., *Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag (2003).
- [HeG09] He H., Garcia E. A., Learning from imbalanced data, *IEEE Trans. Knowl. Data Eng.* 21 (9), 1263–1284, (2009).
- [Jap11] Japkowicz N., Shah M., *Evaluating Learning Algorithms: A Classification Perspective*, Cambridge University Press, (2011).
- [War03] Warren J., On the uniqueness of barycentric coordinates, in R. Goldman and R. Krasauskas (Eds.), *Topics in Algebraic Geometry and Geometric Modeling*, Contemporary Mathematics, Vol. 334, American Mathematical Society, Providence, RI, USA, 93–99, (2003).

IX. DOROBEK DYDAKTYCZNY I ORGANIZACYJNY

W ramach pracy dydaktycznej prowadziłam zajęcia z następujących przedmiotów realizowanych na kierunku Informatyka na Wydziale Informatyki Politechniki Poznańskiej:

- Wspomaganie decyzji (studia niestacjonarne I stopnia);
- Narzędzia informatyki (studia stacjonarne I stopnia);
- Informatyzacja przedsiębiorstw (studia stacjonarne i niestacjonarne I stopnia);
- Hurtownie danych (studia stacjonarne II stopnia, specjalność Technologie Wytwarzania Oprogramowania);
- Hurtownie danych i eksploracja danych (studia stacjonarne II stopnia, specjalność Inteligentne Systemy Wspomagania Decyzji);
- Systemy baz danych i hurtowni danych (studia niestacjonarne II stopnia, specjalność: Technologie Wytwarzania Oprogramowania);
- Podstawy informatyki (studia stacjonarne i niestacjonarne I stopnia dla studentów Uniwersytetu Medycznego prowadzone w ramach współpracy Uniwersytetu Medycznego w Poznaniu z Politechniką Poznańską).

Dla każdego z ww. przedmiotów opracowałam oryginalny zestaw materiałów i ćwiczeń, uzupełniony również raportami:

- Izabela Szczęch, Barbara Wołyńska, Irmina Masłowska
Elementy funkcjonalności modułu finansowo-księgowego systemu Microsoft Dynamics NAV2009
Raport badawczy RB-8/2011, Instytut Informatyki, Politechniki Poznańskiej (2011);
- Izabela Szczęch, Bartosz Szczęch
Podstawy rachunkowości - materiały dydaktyczne do przedmiotu Informatyzacja Przedsiębiorstw
Raport badawczy RB-13/10, Instytut Informatyki, Politechniki Poznańskiej (2010).

Ponadto nawiązałam i koordynuję współpracę z firmą IT.integro, w ramach której Politechnika Poznańska otrzymała licencję akademicką na korzystanie ze zintegrowanego systemu ERP – Microsoft Dynamics NAV, który wykorzystywany jest na przedmiocie Informatyzacja przedsiębiorstw. Dodatkowo, współpraca ta zaowocowała gościnnymi wykładami praktyków prowadzących wdrożenia MS Dynamics NAV oraz organizowaniem corocznych staży i praktyk studenckich w IT.integro.

Począwszy od roku akademickiego 2009/10 sprawuję opiekę nad pracami magisterskimi i inżynierskimi realizowanymi przez studentów kierunku Informatyki na Wydziale Informatyki Politechniki Poznańskiej. Listę prac realizowanych pod moim kierunkiem przedstawiono poniżej.

PRACE DYPLOMOWE MAGISTERSKIE

- Śmiechowski F., System do wielokryterialnej oceny reguł decyzyjnych typu „at-least” i „at-most”, obrona pracy w 2010r.
- Zieliński J., Ewaluacja i porównanie algorytmów generowania reguł asocjacyjnych, obrona pracy w 2010r.

- Mwaba P., Ewaluacja i porównywanie algorytmów odkrywania reguł decyzyjnych, obrona pracy w 2011r.
- Matecki A., System do wielokryterialnej oceny reguł decyzyjnych, obrona pracy w 2011r.
- Jędrzejczak J., Porównanie efektywności zapytań analitycznych w hurtowniach typu ROLAP, MOLAP i HOLAP na przykładzie implementacji SQL Server 2008, obrona pracy w 2011r.
- Skroban P., System tworzenia stron www oparty na modelu SaaS, obrona pracy w 2011r.
- Krzyżaniak D., System do wspomaganie wielokryterialnej analizy budżetu finansowego, obrona pracy w 2012r.
- Dobroszczyk P., System magazynowo-inwentaryzacyjny na urządzenia mobilne wykorzystujący kody kreskowe, obrona pracy w 2012r.
- Gierszewski D., Dynamiczne projektowanie interfejsów WWW do systemu ERP Microsoft Dynamics NAV, obrona pracy w 2013r.
- Zakrzewski B., Integracja aplikacji wspomagającej zarządzanie projektami Microsoft Project 2013 z systemem ERP Microsoft Dynamics NAV 2013, obrona pracy w 2014r.
- Tomas D., Integracja sklepu internetowego VevoCart z systemem ERP Microsoft Dynamics NAV 2013, obrona pracy w 2014r.
- Malinowski A., Rozliczanie podróży służbowych w Microsoft Dynamics NAV, obrona pracy w 2017r.

PRACE DYPLOMOWE INŻYNIERSKIE

- Malinowski A., Moduł bankowości elektronicznej w Microsoft Dynamics NAV, obrona pracy w 2014r.
- Górka M., Szeląg M., Moduł elektronicznej obsługi zamówień i faktur w systemie w Microsoft Dynamics NAV, obrona pracy w 2016r.

W ramach pracy dydaktycznej prowadziłam również zajęcia z następujących przedmiotów realizowanych na kierunku Informatyka na Wydziale Informatyki i Komunikacji Wizualnej w Wyższej Szkole Nauk Humanistycznych i Dziennikarstwa w Poznaniu (obecnie Collegium Da Vinci):

- Wprowadzenie do informatyki (studia stacjonarne i niestacjonarne I stopnia);
- Narzędzia informatyki (studia stacjonarne i niestacjonarne I stopnia).

Dla każdego z ww. przedmiotów opracowałam oryginalny zestaw materiałów i ćwiczeń.

X. PREZENTACJE NA KONFERENCJACH NAUKOWYCH I SEMINARIACH

- **Krytyczny przegląd miar atrakcyjności reguł wyindukowanych z danych**
Seminarium Zakładu Inteligentnych Systemów Wspomagania Decyzji Instytutu Informatyki, Politechnika Poznańska, 2005, Poznań
- **Monotoniczność Bayesowskiej miary confirmacji względem współczynników zaufania i wsparcia reguł**
Seminarium Zakładu Inteligentnych Systemów Wspomagania Decyzji Instytutu Informatyki, Politechnika Poznańska, 2005, Poznań
- **Monotonicity of a Bayesian confirmation measure in rule support and confidence**
AI-METH 2005 Methods of Artificial Intelligence, Gliwice, Poland, 17-11-2005
- **Poszukiwanie reguł niezdominowanych względem wsparcia i anty-wsparcia**
Seminarium Zakładu Inteligentnych Systemów Wspomagania Decyzji, Politechnika Poznańska, 2006, Poznań
- **Wielokryterialna ocena atrakcyjności reguł asocjacyjnych i decyzyjnych**
Seminarium Instytutu Informatyki, Politechnika Poznańska, 2007, Poznań
- **Mining Association Rules with respect to Support and Anti-support - Experimental Results**
RSEISP2007, The Rough Sets and Emerging Systems Paradigms 2007 International Conference, Warszawa, 28-30.06.2007.
- **Analysis of monotonicity properties of some rule interestingness measures**
II Krajowa Konferencja nt. *Technologie Przetwarzania Danych*. Poznań, 24-26.09.2007
- **Przegląd własności miar oceny reguł**
Seminarium Zakładu Inteligentnych Systemów Wspomagania Decyzji Instytutu Informatyki oraz Sekcji "Inteligentnych Systemów Wspomagania Decyzji oraz Obliczeń Elastycznych" Komitetu Informatyki PAN, Politechnika Poznańska, 2009, Poznań
- **Alternative normalization schemas for Bayesian confirmation measures**
IPMU 2010, The 13th International Conference on Information Processing and Management of Uncertainty, Dortmund, Germany, 28.06-02.07.2010.
- **New property for rule interestingness measures**
FedCSIS 2011, Federated Conference on Computer Science and Information Systems, Szczecin, 18.09-21.09.2011.
- **Analysis of symmetry properties for Bayesian confirmation measures**
RSKT 2012, Rough Sets and Knowledge Technology 2012, 7th International Conference, Chengdu, China 17-20.08.2012.
- **Nowe spojrzenie na Bayesowskie miary confirmacji i ich właściwości**
Seminarium Zakładu Inteligentnych Systemów Wspomagania Decyzji Instytutu Informatyki oraz Sekcji "Inteligentnych Systemów Wspomagania Decyzji oraz Obliczeń Granularnych" Komitetu Informatyki PAN, Politechnika Poznańska, 2013, Poznań

- **The Property of χ^2_{01} -Concordance for Bayesian Confirmation Measures**
MDAI 2013, Modeling Decisions for Artificial Intelligence – 10th International Conference, Barcelona, Spain 20-22.11.2013.
- **Visualization of interestingness measures**
LTC 2013, Human Language Technologies as a Challenge for Computer Science and Linguistics – 6th Language & Technology Conference, Poznań, 07-09.12.2013.
- **Visual-Based Detection of Properties of Confirmation Measures**
ISMIS 2014, 21st International Symposium on Methodologies for Intelligent Systems, Roskilde, Denmark 25-27.06.2014.
- **Wizualizacja miar atrakcyjności i skuteczności w odkrywaniu wiedzy i uczeniu maszynowym**
Seminarium Zakładu Inteligentnych Systemów Wspomagania Decyzji Instytutu Informatyki oraz Sekcji "Inteligentnych Systemów Wspomagania Decyzji oraz Obliczeń Granularnych" Komitetu Informatyki PAN, Politechnika Poznańska, 2014, Poznań
- **Teoriogrupowe aspekty symetrii miar konfirmacji**
Seminarium Zakładu Inteligentnych Systemów Wspomagania Decyzji Instytutu Informatyki oraz Sekcji "Inteligentnych Systemów Wspomagania Decyzji oraz Obliczeń Granularnych" Komitetu Informatyki PAN, Politechnika Poznańska, 2017, Poznań
- **Miary konfirmacji i ich właściwości w ocenie reguł**
Seminarium Instytutu Informatyki, Politechnika Poznańska, 2017, Poznań

XI. UDZIAŁ W FINANSOWANYCH PROJEKTACH BADAWCZYCH

- 2005-2007 Projekt badawczy Ministerstwa Nauki i Szkolnictwa Wyższego pt. „**Inteligentne systemy wspomagania decyzji oparte na wiedzy odkrytej z danych**” („Intelligent decision support systems based on knowledge induced from data”) grant nr 3 T11F 021 27, umowa nr 1456/T11/2004/27 (wykonawca)
- 2008-2011 Projekt badawczy Ministerstwa Nauki i Szkolnictwa Wyższego pt. „**Komputerowe metody wspomagania decyzji w oparciu o modele wiedzy wyindukowane z danych alfanumerycznych i tekstowych**” („Computer methods for decision support based on knowledge models induced from alphanumerical and text data”) grant nr N N519 314435, umowa nr 3144/B/T02/2008/35 (wykonawca)
- 2012-2016 Projekt badawczy w ramach badań statutowych Instytutu Informatyki PP „**Inteligentne wspomaganie decyzji – aspekty implementacyjne i sprzętowe**” („Intelligent decision support – implementation and hardware aspects”) (wykonawca).
- 2014-2017 Projekt badawczy (OPUS ST6 Narodowe Centrum Nauki) pt. „**Uczenie się klasyfikatorów z nieźrównoważonych oraz zmiennych danych**” („Learning classifiers from imbalanced and evolving data”) grant nr DEC-2013/11/B/ST6/00963 (wykonawca)

XII. RECENZOWANIE PUBLIKACJI I UDZIAŁ W KOMITETACH PROGRAMOWYCH KONFERENCJI

A. RECENZJE DLA CZASOPISM

- European Journal of Operational Research (5);
- Information Sciences (5);
- Foundations of Computing and Decision Sciences Journal (1);
- Computational Intelligence (1);
- Fundamenta Informaticae (1);
- Computing and Informatics (1);
- Journal of Intelligent Information Systems (1);
- Studia Oeconomica Posnaniensia (3);
- Technical Sciences (1).

B. RECENZJE ARTYKUŁÓW W RAMACH KONFERENCJI

- QIMIE 2009, 2011, 2013, 2015 - *Workshop on Quality issues, measures of interestingness and evaluation of data mining models* organizowany przy konferencji PAKDD;
- AAIA 2011, 2012, 2013, 2014, 2015, 2016, 2017 - *International Symposium Advances in Artificial Intelligence and Applications* organizowane w ramach konferencji FedCSIS;
- CORES 2013 - *The 8th International Conference on Computer Recognition Systems*;
- SIAM 2015, 2016 – *International Conference on Data Mining* (jako recenzent pomocniczy).

C. UDZIAŁ W KOMITETACH PROGRAMOWYCH KONFERENCJI

- QIMIE 2011, 2013, 2015 - *Workshop on Quality issues, measures of interestingness and evaluation of data mining models* organizowany przy konferencji PAKDD;
- AAIA 2012, 2013, 2014, 2015, 2016, 2017 - *International Symposium Advances in Artificial Intelligence and Applications* organizowane w ramach konferencji FedCSIS.

XIII. NAGRODY I WYRÓŻNIENIA

- Nagroda Rektora Politechniki Poznańskiej za osiągnięcia naukowe uzyskane w roku akademickim 2007/2008.
- Nagroda Rektora Politechniki Poznańskiej za wybitne osiągnięcia naukowe w roku 2007 zespołowa II stopnia za cykl 19 publikacji na temat inteligentnych systemów wspomaganie decyzji.
- Nagroda Rektora Politechniki Poznańskiej za osiągnięcia naukowe uzyskane w roku akademickim 2008/2009.
- „The 2011 Zdzisław Pawlak Special Recognition” przyznana w czasie międzynarodowej konferencji *Federated Conference on Computer Science and Information Systems* 2011r. za publikację "New property for rule interestingness measures".

- Nagroda Rektora Politechniki Poznańskiej za osiągnięcia naukowe uzyskane w roku akademickim 2012/2013.
- Nagroda Rektora Politechniki Poznańskiej za wybitne osiągnięcia naukowe w roku 2013 zespołowa I stopnia za cykl 18 publikacji na temat inteligentnych systemów wspomagania decyzji.
- Nagroda Rektora Politechniki Poznańskiej za wyróżniającą się działalność badawczą potwierdzoną publikacjami w roku 2016.
- Medal Komisji Edukacji Narodowej za szczególne zasługi dla oświaty i wychowania, nadany 6 kwietnia 2017r.

Izabela Szczęch

Izabela Szczęch