Autoreferat

1 Informacje podstawowe

1.1 Imię i nazwisko

Marek Śmieja

1.2 Posiadane dyplomy, stopnie naukowe

(1) *Uzyskany stopień:* Doktor nauk matematycznych, dyscyplina: informatyka
Data uzyskania tytułu: styczeń 2015
Miejsce: Wydział Matematyki i Informatyki Uniwersytetu Jagiellońskiego
Tytuł rozprawy: Entropia w kodowaniu i klastrowaniu danych
Promotor: prof. dr hab. Jacek Tabor

(2) *Uzyskany tytuł:* Licencjat informatyki
Data uzyskania tytułu: lipiec 2011
Miejsce: Instytut Informatyki Uniwersytetu Jagiellońskiego

(3) *Uzyskany tytuł:* Magister matematyki
Data uzyskania tytułu: wrzesień 2009
Miejsce: Instytut Matematyki Uniwersytetu Jagiellońskiego
Tytuł pracy: Zastosowanie metody Monte Carlo w wycenie opcji finansowych
Promotor: prof. dr hab. Armen Edigarian

1.3 Zatrudnienie w jednostkach naukowych

- 2015-10-01 – obecnie: Adiunkt w Instytucie Informatyki i Matematyki Komputerowej, Uniwersytetu Jagiellońskiego
- 2012-11-01 – 2015-09-30: Asystent w Instytucie Informatyki i Matematyki Komputerowej, Uniwersytetu Jagiellońskiego

2 Aktywność naukowa

2.1 Staże i wizyty naukowe

- 2019-03-21 – 2019-06-20: staż podoktorski w grupie Pattern and Image Analysis, Instituto de Telecomunicações, Instituto Superior Técnico, Universidade de Lisboa, Portugalia, współpraca z prof. Mario A. T. Figueiredo,

- 2018-11-01 – 2019-01-31: staż w Instytucie Podstaw Informatyki,
  Polska Akademia Nauk, Warszawa,
  współpraca z dr hab. Pawłem Morawieckim,

- 2018-09-10 – 2018-09-13: wizyta w Know-Center Gmbh,
  Graz University of Technology, Austria,
  współpraca z dr Bernhard C. Geiger'em,

- 2018-06-25 – 2018-06-29: wizyta w Grupie Uczenia Maszynowego,
  Politechnika Wrocławska,
  współpraca z prof. dr hab. Michałem Woźniakiem

2.2 Udział w grantach badawczych

(1) 2019 – 2023: *Sztuczne sieci neuronowe inspirowane biologicznie*,
Team-Net (FNP), nr POIR.04.04.00-00-14DE/18-00,
funkcja: młody naukowiec.

(2) 2020 – 2023: *Generowanie rzeczywistych obrazów za pomocą modeli opartych na architekturze autoenkodera*,
Opus (NCN), nr 2019/33/B/ST6/00894,
funkcja: wykonawca.

(3) 2019 – 2022: *Głębokie przetwarzanie danych strukturalnych.*
Opus (NCN), nr 2018/31/B/ST6/00993,
funkcja: kierownik.

(4) 2018 – 2021: *Polifarmakologiczna platforma skriningowa in silico*,
Lider (NCBiR), nr 0137/L-9/2017,
funkcja: wykonawca.

(5) 2018 – 2021: *Efektywne metody uczenia nienadzorowanego z zastosowaniami w głębokim nauczaniu*,
Opus (NCN), nr 2017/25/B/ST6/01271,
funkcja: wykonawca.

(6) 2017 – 2020: *Dodatkowa informacja w grupowaniu danych i zagadnieniach pokrewnych*
Sonata (NCN), nr 2016/21/D/ST6/00980
funkcja: kierownik.

(7) 2016 – 2020: *Budowanie algorytmów grupowania danych w oparciu o uogólnione rozkłady normalne oraz rozkłady nie gaussowskie*,
Sonata (NCN), nr 2015/19/D/ST6/01472
funkcja: wykonawca.

(8) 2016 – 2019: *Teoria analizy niekompletnych danych*,
Opus (NCN), nr 2015/19/B/ST6/01819
funkcja: wykonawca.

(9) 2015 – 2017: *Paradygmat minimalizacji pamięci w klastrowaniu*
Opus (NCN), nr 2014/13/B/ST6/01792,

funkcja: wykonawca.

(10) 2015 – 2016: *Rozwój metod uczenia maszynowego z zastosowaniem do przewidywania aktywności związków chemicznych*,
Preludium (NCN), nr 2014/13/N/ST6/01832,
funkcja: kierownik.

(11) 2013 – 2015: *Klastrowanie w przestrzeniach metrycznych*,
Iuventus Plus (MNiSW), nr IP2012 055972,
funkcja: kierownik.

(12) 2011 - 2014: *Uogólnienie entropii i wymiaru entropijnego oraz ich zastosowania*,
Opus (NCN), nr 2011/01/B/ST6/01887,
funkcja: wykonawca.

3 Wskazanie osiągnięcia habilitacyjnego

Tytuł: **Metody uczenia maszynowego dla danych niekompletnych**

3.1 Wprowadzenie

Metody uczenia maszynowego osiągają imponujące wyniki w typowych zadaniach klasyfikacji czy regresji na popularnych zbiorach danych[1], co zwiększa zainteresowanie ich zastosowaniami komercyjnymi. W przypadku rozpoznawania obrazów możliwości metod uczenia maszynowego niejednokrotnie przewyższają percepcję ludzką [9, 6]. Jednak skuteczność tych metod zależy w dużej mierze od ilości i jakości dostarczonych danych.

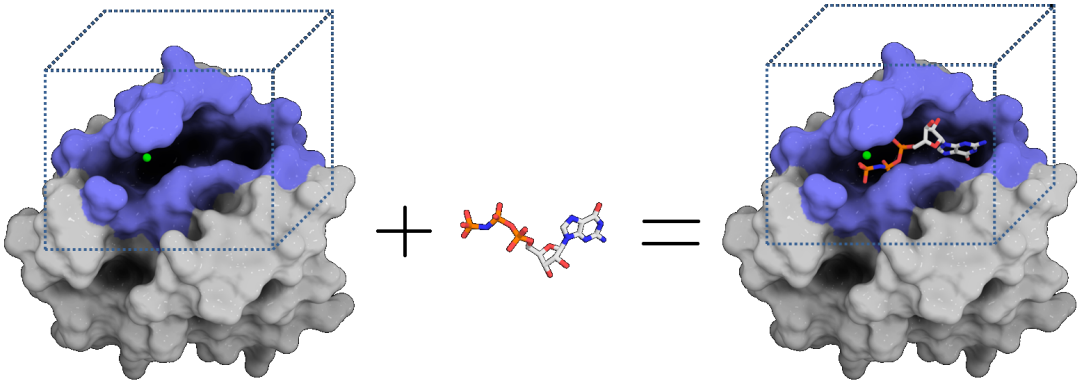
W praktyce dostęp do większej ilości danych odbywa się często za cenę spadku ich jakości. Dane bez etykiet są zwykle łatwe do pozyskania, natomiast każda próba odkrycia etykiety bądź zależności pomiędzy przykładami wiąże się z zaangażowaniem eksperta, co prowadzi do wzrostu kosztów całego procesu uczenia maszynowego. Przykładowo bazy danych zawierają miliony syntetyzowalnych związków chemicznych, podczas gdy aktywność biologiczna wobec określonego białka jest potwierdzona zwykle dla zaledwie kilkuset związków[2] [7, 18], ponieważ wymaga to kosztownych badań laboratoryjnych, Rysunek 1. Takie zjawisko powoduje, że zmienna opisująca etykietę (klasę) bywa w rzeczywistości niekompletna.

Inną sytuacją są brakujące wartości atrybutów, co również powoduje znaczące obniżenie jakości danych. Braki mogą być spowodowane czynnikami losowymi, na przykład awarią czujnika rejestrującego działanie maszyny lub wynikają z pewnego procesu generowania danych. W szczególności brak określonego atrybutu może mieć związek z klasą, którą chcemy odkryć. Dla przykładu zły stan zdrowia pacjenta uniemożliwia wykonania kompletu badań, co skutkuje pojawieniem się braków w opisie badań chorego pacjenta [3].

Obie sytuacje prowadzą do niekompletnego opisu danych – po stronie etykiet bądź atrybutów. O ile człowiek bardzo dobrze radzi sobie z przetwarzaniem takich danych, o tyle bezpośrednie

[1] ImageNet: `https://paperswithcode.com/sota/image-classification-on-imagenet`

[2] PubChem: `https://pubchem.ncbi.nlm.nih.gov/`, ChEMBL: `https://www.ebi.ac.uk/chembl/`



Rysunek 1: Laboratoryjne testowanie aktywności biochemicznej związków (dopasowanie liganda do białka) może być wspomagane metodami uczenia maszynowego. Ich sukces zależy od ilości i jakości związków o potwierdzonej aktywności dostępnych do trenowania modeli. Źródło rysunku: `https://dockthor.lncc.br/v2/`.

zastosowanie standardowych metod uczenia maszynowego nie jest możliwe w tym przypadku. Stanowi to silną motywację do podjęcia badań nad konstrukcją nowych bądź adaptacją istniejących modeli do sytuacji danych niekompletnych.

Celem badań było opracowanie narzędzi uczenia maszynowego, które pozwalają efektywnie pracować z danymi niekompletnymi. Pod uwagę została wzięta zarówno sytuacja, w której opis klasy jest niekompletny jak również przypadek pojawienia się braków w zbiorze atrybutów.

W skład osiągnięcia wchodzi 10 prac opublikowanych w czasopismach z listy JCR (7 prac) bądź w materiałach konferencyjnych posiadających kategorię A* (1 praca) lub A (2 prace) według rankingu CORE[3]. Średnia liczba punktów MNiSW przypadających na jedną publikację wynosi[4] 172, a średni impact factor (IF) prac to[5] 5.590. W dziewięciu pracach byłem pierwszym autorem a w jednej byłem drugim autorem.

Poniższa lista prac, stanowiących osiągnięcie naukowe, została uporządkowana tematycznie i będzie później omawiana w takiej kolejności. Dla każdej z prac podaję liczbę punktów i wskaźnik IF aktualne na rok 2020 oraz liczbę cytowań (bez autocytowań[6]) na podstawie bazy Web of Science (WoS) oraz Google Scholar (GS) aktualne na dzień 14.11.2020. Wpisuję również swój wkład w powstanie każdej z prac. Oświadczenia współautorów o ich wkładzie znajdują się w załączniku nr 5.

[A1] Marek Śmieja, Oleksandr Myronov, Jacek Tabor.
Semi-supervised discriminative clustering with graph regularization.
Konwledge-Based Systems, 151, p. 24–36, 2018.
Punkty MNiSW: 200, IF: 5.921.

[3] `http://portal.core.edu.au/conf-ranks/`

[4] Z uwagi na zmianę systemu punktacji czasopism, podaję liczbę punktów, a także wskaźnik impact factor aktualne na rok 2020.

[5] Z obliczenia średniego wskaźnika IF wyłączone zostały publikacje konferencyjne jako że nie posiadają wskaźnika IF.

[6] Autocytowanie rozumiane jest jako cytowanie przez któregokolwiek ze współautorów pracy.

[A2] Marek Śmieja, Łukasz Struski, Mario A. T. Figueiredo.
A Classification-Based Approach to Semi-Supervised Clustering with Pairwise Constraints
Neural Networks. 127, p. 193–203, 2020
Punkty MNiSW: 200, IF: 5.535.

[A3] Marek Śmieja, Magdalena Wiercioch.
Constrained clustering with a complex cluster structure.
Advances in Data Analysis and Classification, 11/3, pp. 493–518, 2017.
Punkty MNiSW: 100, IF: 1.603.

[A4] Marek Śmieja, Bernhard C. Geiger.
Semi-supervised cross-entropy clustering with information bottleneck constraint.
Information Sciences, 421, pp. 245–271, 2017.
Punkty MNiSW: 200, IF: 5.910.

[A5] Marek Śmieja, Łukasz Struski, Jacek Tabor.
Semi-supervised model-based clustering with controlled clusters leakage.
Expert Systems with Applications, 85, pp. 146-157, 2017.
Punkty MNiSW: 140, IF: 5.452.

[A6] Marek Śmieja, Maciej Wołczyk, Jacek Tabor, Bernhard C. Geiger.
SeGMA: Semi-Supervised Gaussian Mixture Auto-Encoder.
IEEE Transactions on Neural Networks and Learning Systems,
DOI:10.1109/TNNLS.2020.3016221, p. 12, 2020.
Punkty MNiSW: 200, IF: 8.793.

[A7] Marek Śmieja, Łukasz Struski, Jacek Tabor, Mateusz Marzec.
Generalized RBF kernel for incomplete data.
Knowledge-Based Systems, 173, p. 150-162, 2019.
Punkty MNiSW: 200, IF: 5.921.

[A8] Marek Śmieja, Łukasz Struski, Jacek Tabor, Bartosz Zieliński, Przemysław Spurek.
Processing of missing data by neural networks.
Advances in Neural Information Processing Systems 31 (NeurIPS), p. 2719–2729, 2018.
Punkty MNiSW: 200, Core rank: A*.

[A9] Tomasz Danel, Marek Śmieja, Łukasz Struski, Przemysław Spurek, Łukasz Maziarka.
Processing of Incomplete Images by (Graph) Convolutional Neural Networks.

International Conference on Neural Information Processing (ICONIP), Lecture Notes on Computer Science 12533, DOI:10.100978-3-030-63833-7_43, p. 12, 2020.
Punkty MNiSW: 140, Core rank: A.

[A10] Marek Śmieja, Maciej Kołomycki, Łukasz Struski, Mateusz Juda, Mario A. T. Figueiredo. *Iterative Imputation of Missing Data using Auto-encoder Dynamics.* International Conference on Neural Information Processing (ICONIP), Lecture Notes on Computer Science 12534, DOI:10.1007978-3-030-63836-8_22, p. 12, 2020.

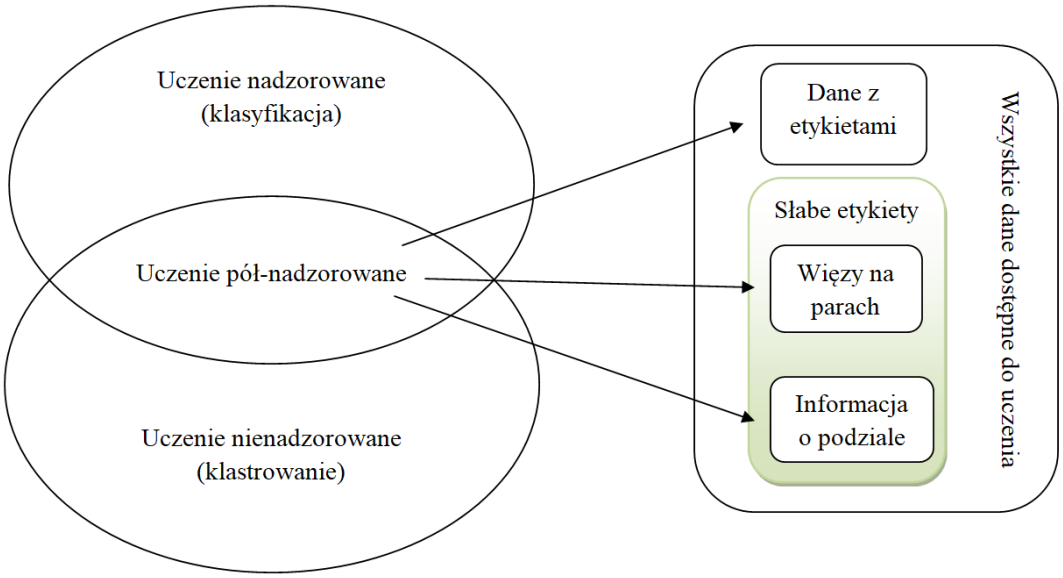
Powyższe prace dotyczą zarówno aspektów teoretycznych jak i mają zastosowanie praktyczne. Z jednej strony poświęciłem szczególną uwagę teoretycznej analizie modeli, możliwości uzyskania chociaż częściowych rozwiązań analitycznych bądź też uzyskaniu takiej postaci funkcji kosztu, która może być łatwo optymalizowana technikami gradientowymi. Z drugiej strony dążyłem do osiągnięcia poprawy wyników eksperymentalnych w stosunku do powszechnie stosowanych rozwiązań.

Najważniejszymi pracami w kontekście wskazanego osiągnięcia są:

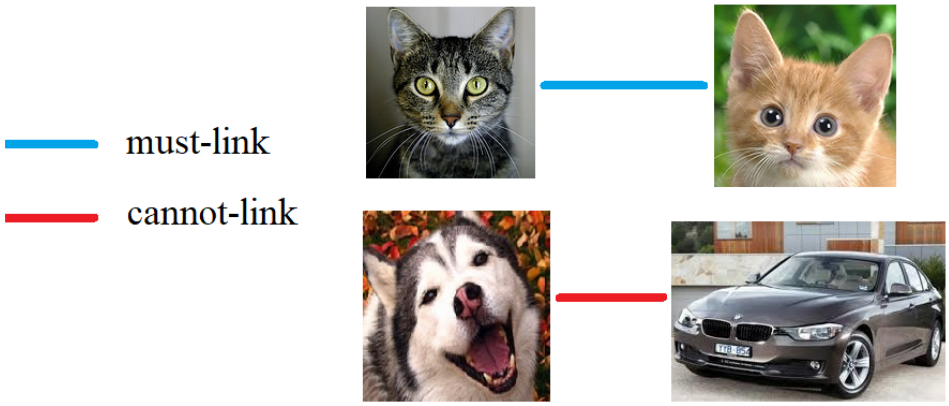
- Prace [A1, A2] wprowadzające dyskryminatywne modele grupowania z więzami, w których osiągnąłem znaczną poprawę wyników w stosunku do istniejących rozwiązań.
- Praca [A4], która definiuje model grupowania ze wstępnym podziałem z uwagi na przeprowadzoną dokładną analizę teoretyczną modelu.
- Praca [A6] prezentująca model generatywny z częściowym etykietowaniem, w której uzyskałem bardzo ciekawe własności eksperymentalne (ciągły transfer stylu).
- Prace [A8, A7] pokazujące jak uczyć metody kernelowe oraz sieci neuronowe na brakujących danych. Oba modele są łatwe w użyciu, dają bardzo dobre wyniki eksperymentalne, a ich budowa ma jasną motywację teoretyczną.

3.2 Zakres badań

W przypadku niekompletności etykiet, mamy do czynienia z *uczeniem pół-nadzorowanym* (ang. semi-supervised learning) [4], gdzie część przykładów uczących posiada informację dotyczącą przynależności do określonych klas, a dla pozostałej części ta informacja nie jest znana, Rysunek 2. Zamiast etykiet można również wykorzystać tak zwane *słabe etykiety* (ang. weak labels), które są zwykle tańsze w pozyskaniu, gdyż nie wymagają identyfikacji i nazywania klas. Jednym z przykładów słabych etykiet są *więzy na parach* (ang. pairwise constraints) [2], które określają czy para punktów pochodzi z tej samej klasy (ang. must-link) czy też reprezentuje różne klasy (ang. cannot-link), Rysunek 3. Typowy proces pozyskiwania więzów polega na odpytywaniu eksperta czy dana para przykładów powinna zostać zgrupowana razem. W pewnych zastosowaniach proces określania więzów można częściowo zautomatyzować: w [A1] rozważyłem możliwość nadawania więzów na podstawie wewnętrznego podobieństwa; w segmentacji obrazów często zakłada się, że sąsiednie piksele obrazu z dużym prawdopodobieństwem reprezentują ten sam obiekt [5]. Choć więzy nie pozwalają na jednoznaczne zidentyfikowanie klasy, są one powszechnie wykorzystywane w problemach *pół-nadzorowanego klastrowania*



Rysunek 2: Związek pomiędzy różnymi typami uczenia oraz danymi dostępnymi do uczenia w sytuacji pół-nadzorowanej.

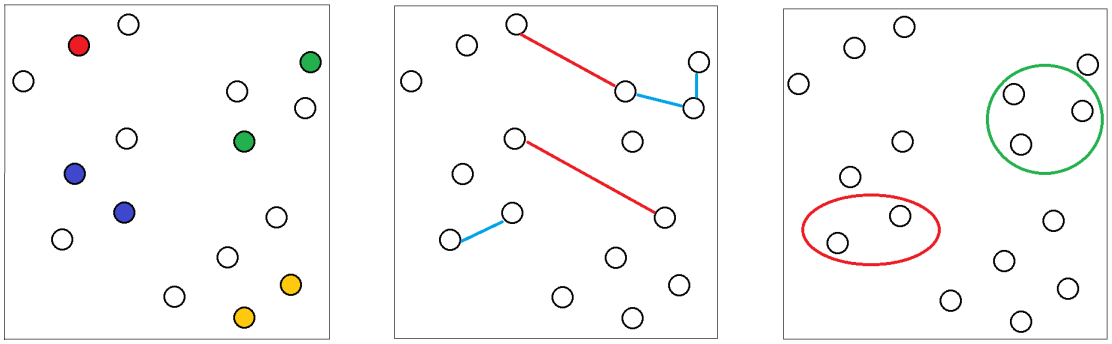


Rysunek 3: Przykłady więzów typu must-link i cannot-link nałożonych na zbiorze obrazkowym.

(ang. semi-supervised clustering), gdzie celem jest wydobycie grup a nie nazwanie klas. Alternatywą dla więzów jest *informacja o podziale* (ang. partition-level side information) [11], gdzie manualne grupowanie dokonywane jest na niewielkim podzbiorze danych. Mając do dyspozycji podzbiór przykładów, często łatwiej podzielić je na grupy niż definiować więzy pomiędzy wszystkimi parami przykładów. W odróżnieniu od etykietowania części zbioru (występującego w klasyfikacji), informacja o podziale nie nazywa klas ani nie musi pokrywać wszystkich klas. Rysunek 4 obrazuje podstawowe różnice pomiędzy etykietami i słabymi etykietami.

W pierwszej części swoich badań skupiłem się na użyciu informacji o klasach w typowych problemach uczenia nienadzorowanego:

- **Grupowanie z więzami.** W pracach [A2, A1] zostały opracowane dyskryminatywne algorytmy grupowania naśladujące modele klasyfikacyjne. Skonstruowana funkcja kosztu, bazująca na funkcji softmax, może być optymalizowana gradientowo, a co za tym idzie użyta w mode-



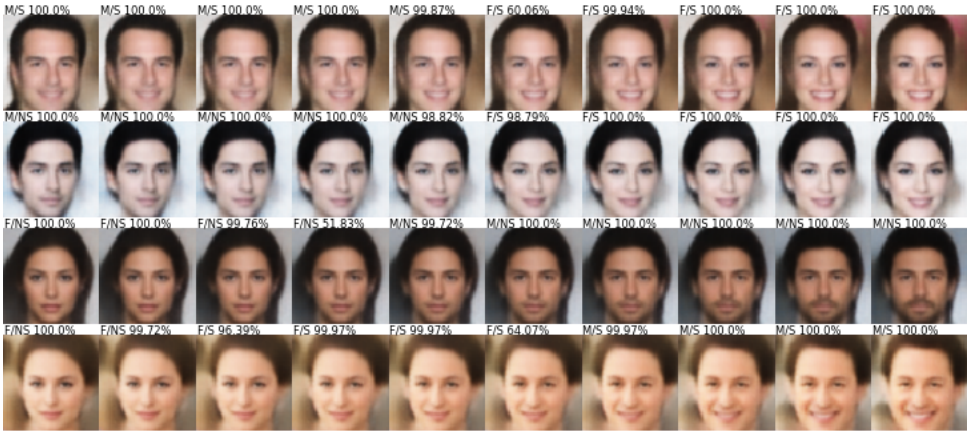
(a) Etykiety (b) Więzy na parach (c) Informacja o podziale

Rysunek 4: Porównanie różnych typów informacji o klasach w problemach pół-nadzorowanych: etykiety (4a) służą do identyfikacji wszystkich klas i są powszechnie stosowane w klasyfikacji; więzy na parach (4b) określają czy para punktów należy do tej samej grupy; informacja o podziale (4c) zawiera przypisania podzbioru elementów do grup. W odróżnieniu od etykiet, słabe etykiety nie muszą pokrywać wszystkich klas i nie identyfikują (nazywają) klas.

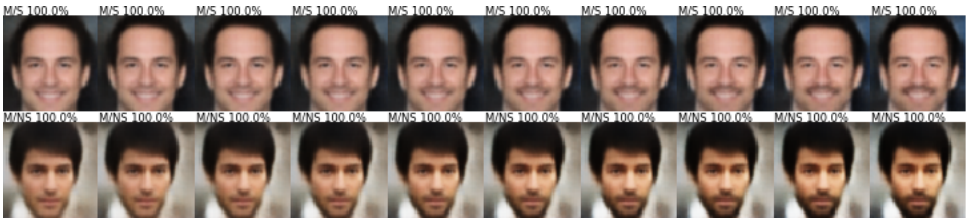
lach głębokiego uczenia. Zaproponowane modele potrafią bardzo efektywnie wykorzystać informację zawartą w więzach, przez co osiągają wyniki lepsze niż modele standardowe [15, 17, 16].
Praca [A3] proponuje model doboru skomplikowania modelu klastrowania na podstawie przekazanych więzów. W szczególności klaster może zostać opisany mieszanką rozkładów gaussowskich zamiast pojedynczym rozkładem Gaussa, co stanowi przewagę nad konkurencyjnymi modelami z podobnej klasy [17, 12], które stosują rozkłady jednomodalne do opisu klastra.

- **Grupowanie z informacją o podziale.** Alternatywą dla więzów jest przekazanie informacji o klasach we wstępnym podziale danych, co zmniejsza ryzyko popełnienia błędów w tak podanej informacji. Prace [A4, A5] wprowadzają model mieszanki gaussowskiej, który jest jest w stanie dobrać liczbę klastrów odpowiednią do opisu danych oraz zachowania informacji przekazanej przez eksperta. Na szczególna uwagę zasługuje dokładana analiza teoretyczna zaproponowanych modeli.
- **Model generatywny z częściowym etykietowaniem.** Praca [A6] wprowadza model generatywny uczony z wykorzystaniem niewielkiej liczby etykiet, który pozwala na generowanie danych z wybranej klasy. Dzięki zastosowaniu rozkładu mieszanki gaussowskiej w przestrzeni ukrytej, możliwa jest interpolacja pomiędzy przykładami różnych klas, co wyróżnia nasz model spośród konkurencyjnych rozwiązań [13, 8]. Istotną konsekwencją jest możliwość zmiany atrybutów opisujących klasę w sposób ciągły, na przykład modyfikacja zdjęcia twarzy polegająca na stopniowym dodawaniu zarostu lub nadaniu uśmiechu o różnym natężeniu, Rysunek 5.

Niekompletność atrybutów oznacza brak informacji o wartościach niektórych współrzędnych dla pewnych przykładów, co zwykle określa się mianem *danych brakujących* (ang. missing data) [10]. Generowanie wypełnień dla danych brakujących jest jednym z istotnych problemów rozważanych w przetwarzaniu obrazów (ang. image inpainting) [14], Rysunek 6, jak również występuje w zastosowaniach medycznych (niekompletne badania pacjenta) [3], czy przemysłowych (awaria czujnika rejestrującego zachowanie maszyny) [A7]. Pojawienie się braków



(a) Ciągła zmiana cech



(b) Zmiana intensywności cech

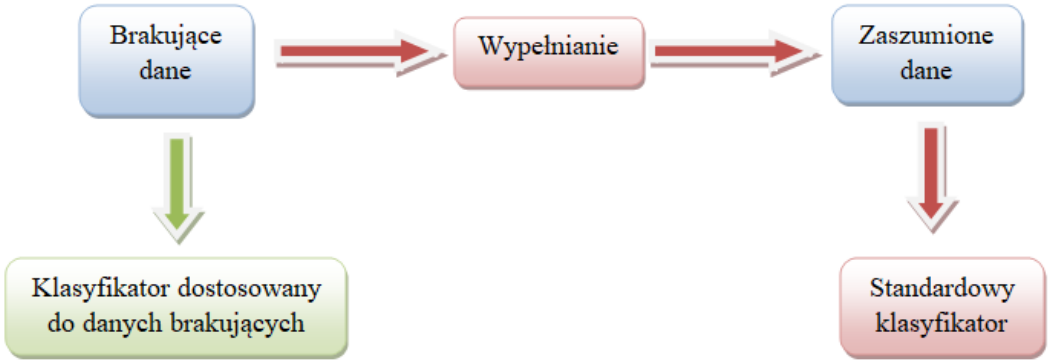
Rysunek 5: Przykład ciągłego transferu stylu pomiędzy klasami dla zaproponowanego modelu SeGMA [A6]. Pierwszy rysunek pokazuje ciągłą zmianę cech, a drugi wzmacnianie (osłabianie) cech.

uniemożliwia bezpośrednie zastosowanie standardowych metod uczenia bez wcześniejszego ich wypełnienia. Stąd typowy proces uczenia na danych brakujących sprowadza się do uzupełnienia braków i zastosowaniu standardowych metod uczenia na takich danych, Rysunek 7. Niestety niepoprawne wypełnienie prowadzi do błędnych danych, co z kolei negatywnie wpływa na trening modeli predykcyjnych. Pojawiają się zatem dwa komplementarne problemy: jak trenować modele bezpośrednio na niepełnych danych oraz jak wypełniać brakujące wartości?

W drugiej części badań zająłem się następującymi problemami uczenia maszynowego związanymi z brakującymi danymi:

- **Uczenie maszynowe z modelowaniem niepewności.** Podczas gdy wypełnienie braku niesie ze sobą ryzyko pomyłki, to zamodelowanie brakującej wartości rozkładem prawdopodobieństwa jest bezpieczniejsze. W pracach [A7, A8] dokonałem adaptacji popularnych modeli, takich jak kernel SVM oraz sieci neuronowych, w przypadku gdy brakujące dane zostały zamodelowane za pomocą rozkładów gaussowskich. Istota modeli polega na uśrednianiu funkcji jądrowej (dla SVM), bądź funkcji aktywacji (dla sieci neuronowych) w sposób analityczny, biorąc pod uwagę wszystkie możliwe wypełnienia. Uzyskanie wzorów analitycznych pozwala na szybsze uczenie (w porównaniu z generowaniem wielu wypełnień) i dokładniejszą estymację (w porównaniu z imputacją).
- **Uczenie maszynowe z ignorowaniem braków.** Praca [A9] przedstawia sposób reprezentowania obrazów z brakującymi wartościami za pomocą grafów i ich przetwarzania wykorzy-

Rysunek 6: Przykładowe wypełnia dla problemu inpaintingu obrazów [19]. O ile w przypadku obrazów wiele rozwiązań, nawet zróżnicowanych, jest akceptowalnych, o tyle w zastosowaniach medycznych czy przemysłowych dążymy zwykle od odtworzenia prawdziwych danych.



Rysunek 7: Standardowy proces uczenia z brakujących polega na wypełnieniu braków i zastosowaniu standardowych modeli uczenia maszynowego na wypełnionych danych (czerwone strzałki). Alternatywą jest taka adaptacja metod, aby umożliwić ich bezpośrednie trenowanie na danych brakujących (zielona strzałka).

stując konwolucyjne sieci grafowe. Ignorując brakujące wartości unikamy ryzyka związanego z błędnym wypełnieniem.

- **Generowanie wypełnień.** W pracy [A10] rozważyłem problem, gdy brakujące dane nie pojawiają się w czasie treningu, lecz w czasie testu. Bazując na własnościach funkcji rekonstrukcji auto-enkodera [1] zaproponowałem metodę uzupełniania braków za pomocą najbardziej prawdopodobnego wypełnienia. Zbudowany algorytm może wykorzystać dowolny model auto-enkodera, który jest uczony w sposób standardowy bez konieczności jego specjalizacji do zadania wypełniania.

3.3 Literatura

[1] G. Alain and Y. Bengio. What regularized auto-encoders learn from the data-generating distribution. *Journal of Machine Learning Research*, 15:3563–3593, 2014.

[2] Sugato Basu, Ian Davidson, and Kiri Wagstaff. *Constrained clustering: Advances in algorithms, theory, and applications*. CRC Press, 2008.

[3] Lora E Burke, Jacqueline M Dunbar-Jacob, and Martha N Hill. Compliance with cardiovascular disease prevention strategies: a review of the research. *Annals of Behavioral Medicine*, 19(3):239–263, 1997.

[4] Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. *IEEE Transactions on Neural Networks*, 20(3):542–542, 2009.

[5] Dong S Cheng, Vittorio Murino, and Mário Figueiredo. Clustering under prior knowledge with application to image segmentation. In *Advances in Neural Information Processing Systems*, pages 401–408, 2007.

[6] Samuel Dodge and Lina Karam. A study and comparison of human and deep learning recognition performance under visual distortions. In *2017 26th international conference on computer communication and networks (ICCCN)*, pages 1–7. IEEE, 2017.

[7] Sunghwan Kim, Paul A Thiessen, Evan E Bolton, Jie Chen, Gang Fu, Asta Gindulyte, Lianyi Han, Jane He, Siqian He, Benjamin A Shoemaker, et al. Pubchem substance and compound databases. *Nucleic acids research*, 44(D1):D1202–D1213, 2016.

[8] Durk P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. Semi-supervised learning with deep generative models. In *Proc. Advances in Neural Information Processing Systems (NIPS)*, pages 3581–3589, 2014.

[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[10] Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*, volume 793. John Wiley & Sons, 2019.

[11] Hongfu Liu and Yun Fu. Clustering with partition level side information. In *2015 IEEE international conference on data mining*, pages 877–882. IEEE, 2015.

[12] Z. Lu and T. Leen. Semi-supervised learning with penalized probabilistic clustering. In *Advances in Neural Information Processing Systems (NIPS)*, pages 849–856, 2004.

[13] Lars Maaløe, Casper Kaae Sønderby, Søren Kaae Sønderby, and Ole Winther. Auxiliary deep generative models. In *Proc. Int. Conf. on Machine Learning (ICML)*, pages 1445–1453, 2016.

[14] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016.

[15] Yuanli Pei, Xiaoli Z Fern, Teresa Vania Tjahja, and Rómer Rosales. Comparing clustering with pairwise and relative constraints: A unified framework. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 11(2):1–26, 2016.

[16] P. Qian, Y. Jiang, S. Wang, K. Su, J. Wang, L. Hu, and R. Muzic. Affinity and penalty jointly constrained spectral clustering with all-compatibility, flexibility, and robustness. *IEEE Transactions on Neural Networks and Learning Systems*, 28(5):1123–1138, 2017.

[17] Noam Shental, Aharon Bar-Hillel, Tomer Hertz, and Daphna Weinshall. Computing gaussian mixture models with em using equivalence constraints. In *Advances in neural information processing systems*, pages 465–472, 2004.

[18] Anne Mai Wassermann and Jürgen Bajorath. Bindingdb and chembl: online compound databases for drug discovery. *Expert opinion on drug discovery*, 6(7):683–687, 2011.

[19] Lei Zhao, Qihang Mo, Sihuan Lin, Zhizhong Wang, Zhiwen Zuo, Haibo Chen, Wei Xing, and Dongming Lu. Uctgan: Diverse image inpainting based on unsupervised cross-space translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5741–5750, 2020.