

„Inżynieria wiedzy w konstruowaniu modeli sieci bayesowskich w medycynie”

Tematyka mojej pracy naukowej w ciągu ostatnich 13 lat (po uzyskaniu doktoratu) koncentrowała się na budowaniu modeli probabilistycznych opartych na sieciach bayesowskich w dziedzinie medycyny. W pracy mojej kładłam szczególny nacisk na praktyczne zastosowania, w których zainteresowana byłam inżynierią wiedzy. Problem ten jest trudny, ponieważ mamy do czynienia z dużą różnicą pomiędzy rachunkiem prawdopodobieństwa a intuitywnym przetwarzaniem niepewności przez człowieka, a więc zarówno eksperta jak i użytkownika systemów wspomagania decyzji w warunkach niepewności.

Sieci bayesowskiej w ciągu ostatnich dwóch dekad stały się popularnym narzędziem do modelowania trudnych problemów w warunkach niepewności. Sieć bayesowska składa się z dwóch części: (1) jakościowej, która jest acyklicznym grafem skierowanym modelującym zmienne losowe i zależności probabilistyczne między nimi oraz z (2) części ilościowej reprezentowanej przez warunkowe rozkłady prawdopodobieństwa. Inżynieria wiedzy w budowaniu modeli sieci bayesowskich sprowadza się między innymi do pozyskiwania wiedzy w danej dziedzinie, transformacji tej wiedzy do acyklicznego grafu skierowanego oraz wyznaczania warunkowych rozkładów prawdopodobieństwa. Model sieci bayesowskiej może być skonstruowany w oparciu o wiedzę eksperta bądź też nauczony z danych. Proces tworzenia takiego modelu wymaga niejednokrotnie łączenia danych, które pochodzą z różnych źródeł. Źródłem danych może być opinia eksperta, dane kliniczne, dane z badań przesiewowych, dane histopatologiczne, dane ankietowe, itp. Ważną rolę odgrywa w tym procesie inżynier wiedzy, który powinien opanować techniki pozyskiwania wiedzy od eksperta, rozumieć modelowaną dziedzinę, posiadać umiejętności przetworzenia danych medycznych, również w sytuacji gdy są one niekompletne.

W ramach swoich badań przeprowadziłam szereg eksperymentów, które badały wpływ dokładności parametrów numerycznych modeli sieci bayesowskich na ich jakość diagnostyczną. Jeden z przeprowadzonych eksperymentów sprowadzał się do zaokrąglania dokładności parametrów numerycznych modeli sieci bayesowskich i badanie ich jakości diagnostycznej [Oniśko et al. 2013]. W innych badaniach modelowany był różnego rodzaju szum: od szumu symetrycznego do szumu, który modeluje zachowanie eksperta określającego parametry numeryczne sieci bayesowskiej, tj. ekspert zbyt pewny siebie lub ekspert niepewny swoich estymacji [Oniśko et al. 2003; Drużdżel et al. 2008; Oniśko et al. 2011]. Wyniki tych eksperymentów pokazały, że szum i ogólnie jakość parametrów nie mają istotnego wpływu na jakość diagnostyczną modeli. Znaczące pogorszenie jakości diagnostycznej modeli sieci bayesowskich spowodowane było zaokrągleniem małych wartości prawdopodobieństw do zera. Wyniki te są istotną wskazówką dla inżynierów wiedzy, którzy konstruują modele sieci bayesowskich. Szczególnie istotne jest to w sytuacji, gdy inżynierowie wiedzy korzystają z wiedzy ekspertów lub gdy parametry numeryczne modeli są estymowane z niekompletnych zbiorów danych.

Przeprowadziłam również badania w zakresie struktury graficznej sieci bayesowskich oraz wpływu tej struktury na jakość diagnostyczną modeli probabilistycznych [Oniśko et al. 2014]. W ramach tych badań w pierwszej kolejności zmieniana była struktura sieci bayesowskiej poprzez systematyczne usuwanie krawędzi grafu w określonym porządku. W drugiej kolejności odwracany był kierunek krawędzi grafu, również w określonym porządku. Za każdym razem sprawdzana była jakość diagnostyczna modelu. Wyniki naszych eksperymentów pokazały, że zmiany w strukturze sieci bayesowskiej, szczególnie w przypadku odwracania kierunku krawędzi, nie mają dużego wpływu na jakość diagnostyczną modelu.

Sieci bayesowskie są często stosowane w modelowaniu problemów diagnostycznych. W ramach badań, które prowadziłam w ciągu ostatnich lat, zbudowałam kilka diagnostycznych modeli sieci bayesowskich [Lacave et al. 2006; Oniśko 2008]. Na przykład model BPH do diagnozowania łagodnego przerostu prostaty [Makal et al. 2004] lub też model AutismNET wspierający wczesne diagnozowanie autyzmu [Szczygieł et al. 2014]. Oba modele zostały zbudowane w oparciu o literaturę medyczną oraz wiedzę eksperta. Inny model diagnostyczny został zbudowany w oparciu o dane immunohistochemiczne reprezentujące biomarkery i pozwala na diagnozowanie różnicowe pomiędzy dwoma typami nowotworów [Jones et al. 2013]. Dodatkowo model wyznacza wartość diagnostyczną biomarkerów wskazując jednocześnie, które z nich najlepiej różnicują uwzględnione diagnozy.

Agnieszka Oniśko
Wydział Informatyki Politechnika Białostocka
Magee-Womens Hospital, University of Pittsburgh Medical Center, Pittsburgh, USA

Sieci bayesowskie mogą również być stosowane w modelowaniu prognozy medycznej. Jeden z moich modeli, Pittsburgh Cervical Cancer Screening Model, jest narzędziem wyznaczającym ryzyko zachorowania na raka szyjki macicy. Model jest dynamiczną siecią bayesowską, która została zbudowana w oparciu o wiedzę eksperta oraz wyniki badań przesiewowych zgromadzonych na Uniwersytecie Pittsburgh'skim na przestrzeni 11 lat (ponad milion rekordów) [Austin et al. 2010; Oniśko et al. 2015; Austin et al. 2016]. Kolejny model sieci bayesowskiej, który zbudowałam dotyczy prognozowania nowotworu endometrium w oparciu o wyniki badań przesiewowych [Jing et al.]. Celem tego modelu było zidentyfikowanie kobiet, które mają zwiększone ryzyko zachorowania na nowotwór endometrium i jednocześnie wskazanie kobiet, dla których to ryzyko jest minimalne i które nie muszą być poddawane dodatkowym inwazyjnym badaniom diagnostycznym.

Ważnym elementem modelowania za pomocą sieci bayesowskich jest odkrywanie probabilistycznych zależności z danych. W oparciu o dane ankietowe zbudowaliśmy sieć bayesowską pozwalającą na analizę czynników ryzyka i przyczyn występowania próchnicy u trzylatków [Łąguna 2014]. Jeden z wyników tej analizy wskazał, że czyszczenie jamy ustnej jest jednym z najważniejszych czynników zapobiegania próchnicy u dzieci, nawet jeśli dziecko spożywa dużą ilość słodyczy.

Wyniki moich prac wskazują na potencjał stosowania sieci bayesowskich w praktyce. W każdej z dziedzin, w której pracowałam, stworzone modele były w stanie rozwiązać zadany problem. W kilku wypadkach, stworzone modele zostały praktycznie zastosowane.

Mój dorobek naukowy po doktoracie składa się z 46 publikacji. Web of Science indeksuje 35 moich publikacji, 266 cytowań, oraz h-index równy 8. Szczegółowe zestawienie wskaźników dorobku naukowego znajduje się poniżej.

Bibliografia

- [Oniśko et al. 2013] Agnieszka Oniśko, Marek J. Druzdzel. Impact of Precision of Bayesian Networks Parameters on Accuracy of Medical Diagnostic Systems, *Artificial Intelligence in Medicine*, 57(3):197-206, 2013.
- [Oniśko et al. 2003] Agnieszka Oniśko and Marek J. Druzdzel. Effect of Imprecision in Probabilities on Bayesian Network Models: An Empirical Study. AIME Workshop on Model-based and Qualitative Reasoning in Biomedicine, Protaras, Cypr, 19-22 października 2003.
- [Druzdzel et al. 2008] Marek J. Druzdzel, Agnieszka Oniśko. The Impact of Overconfidence Bias on Practical Accuracy of Bayesian Network Models: An Empirical Study. In *Working Notes of the 2008 Bayesian Modelling Applications Workshop, Special Theme: How Biased Are Our Numbers? Part of the Annual Conference on Uncertainty in Artificial Intelligence (UAI-2008)*, Helsinki, Finlandia, 9 lipca 2008.
- [Oniśko et al. 2011] Agnieszka Oniśko and Marek J. Druzdzel. Impact of Quality of Bayesian Networks Parameters on Accuracy of Medical Diagnostic Systems: An Empirical Study. In *Working Notes of the Workshop Probabilistic Models in Biomedicine. European Conference on Artificial Intelligence in Medicine (AIME-11)*, Bled, Slovenia, 2011.
- [Oniśko et al. 2014] Agnieszka Oniśko, Marek J. Druzdzel: Impact of Bayesian network model structure on the accuracy of medical diagnostic systems, *Proceedings of 13th International Conference on Artificial Intelligence and Soft Computing (ICAISC'2014)*, eds. Leszek Rutkowski, Marcin Korytkowski, Rafal Scherer, Ryszard Tadeusiewicz, Lofti A. Zadeh, Jacek M. Zurada, Heidelberg, Springer, Lecture Notes in Computer Science, Vol. 8467: 167-178, Zakopane, 1-5 czerwca 2014.
- [Lacave et al. 2006] Carmen Lacave, Agnieszka Oniśko, Francisco J. Díez. Use of Elvira's explanation facility for debugging probabilistic expert systems. *Knowledge-Based Systems*, Elsevier, vol. 19(8): 730-738, 2006.
- [Oniśko 2008] Agnieszka Oniśko. Medical Diagnosis, In Patrick Naim, Olivier Pourret, and Bruce Marcot(eds), *Bayesian Networks: A Practical Guide to Applications*, Wiley & Sons, strony: 15-32, marzec 2008.
- [Makal et al. 2004] Jarosław Makal, Andrzej Nazarkiewicz, Agnieszka Oniśko, Piotr Orzechowski. Expert system for diagnosis of Benign Prostatic Hyperplasia. *Journal of Automation and Measurements*, No. 7/8: 193-196, 2004.
- [Jones et al. 2013] Mirka W. Jones, Agnieszka Oniśko, David J. Dabbs, Esther Elishaev, Rohit Bhargava. Immunohistochemistry and HPV in situ hybridization in distinction between endocervical and endometrial adenocarcinoma: A comparative tissue microarray study of 76 tumors, *International Journal of Gynecological Cancer*, 23(2):380-4, 2013.
- [Szczygieł et al. 2014] Justyna Szczygieł, Agnieszka Oniśko, Jolanta Świdorska, Elżbieta Krysiwicz, Jerzy Sienkiewicz. Probabilistic graphical model supporting early diagnosis of autism spectrum disorder, *Advances in Computer Science Research*, No. 11, s. 151-164, 2014.
- [Austin et al. 2010] R. Marshall Austin, Agnieszka Oniśko, Marek J. Druzdzel. The Pittsburgh Cervical Cancer Screening Model. A Risk Assessment Tool. *Archives of Pathology and Laboratory Medicine*, 134:744-750, 2010.
- [Oniśko et al. 2015] Agnieszka Oniśko, R. Marshall Austin: Dynamic Bayesian Network for Cervical Cancer Screening, In Peter J.F. Lucas and Arjen Hommersom(eds), *Foundations of Biomedical Knowledge Representations. Methods and Applications*, Springer, *Lectures Notes in Artificial Intelligence* 9521: 207-218, 2015.
- [Austin et al. 2016] R. Marshall Austin, Agnieszka Oniśko. Increased Cervical Cancer Risk Associated with Extended Screening Intervals after Negative Human Papilloma Virus (HPV) Test Results: Bayesian Risk Estimates using the Pittsburgh Cervical Cancer Screening Model. *Journal of American Society of Cytopathology*, 5 (1):9-14, 2016.
- [Jing et al. 2016] Jing Yu, Agnieszka Onisko, R. Marshall Austin. Bethesda System Reporting of Benign-Appearing Endometrial Cells in Women 40 and Older: Analysis of Predictive Value from a Large Academic Women's Hospital Database, United States and Canadian Academy of Pathology's 105th Annual Meeting, 12-18 marca 2016, Seattle, WA, USA.
- [Łaguna 2014] Wojciech Łaguna. Modele probabilistyczne w wyznaczaniu czynników ryzyka próchnicy u dzieci trzyletnich. Praca magisterska, Wydział Informatyki, Politechnika Białostocka, lipiec 2014.

Wskaźniki dorobku naukowego

Publikacje po doktoracie (2003-2016)

Rodzaj publikacji	Liczba publikacji
Artykuły w czasopismach z JCR	8
Streszczenia w czasopismach z JCR	14
Artykuły w czasopismach spoza JCR	6
Rozdziały w książkach	6
Publikacje w materiałach konferencyjnych	12
Suma:	46

1. Cytowania Web of Science

Indeksy cytowań	Wartość
Liczba publikacji	35
Liczba cytowań	266
Liczba cytowań bez autocytowań	251
Liczba artykułów cytujących	234
Liczba artykułów cytujących bez autocytowań	225
Średnia liczba cytowań na publikację	7.60
h-index	8

2. Cytowania Google Scholar

Indeksy cytowań	Wartość
Liczba cytowań	738
h-index	13
i10-index	17

3. Cytowania w Research Gate

Indeksy cytowań	Wartość
Liczba cytowań	509
h-index	12
RG score	20.2