

**List rekomendacyjny Promotora/ki rozprawy doktorskiej zgłoszonej  
do konkursu *Polskiego Stowarzyszenia Sztucznej Inteligencji*  
na najlepszą rozprawę doktorską w dziedzinie Sztucznej Inteligencji**

Imię i nazwisko autora/ki rozprawy:

Jakub Piekoszewski .....

Tytuł rozprawy:

Uogólnione samoorganizujące się sieci neuronowe o drzewopodobnych strukturach w grupowaniu danych ze szczególnym uwzględnieniem danych medycznych opisujących ekspresję genów .....

Imię i nazwisko oraz afiliacja Promotora/ki:

Marian B. Gorzałczany, Wydział Elektrotechniki, Automatyki i Informatyki, Politechnika Świętokrzyska .....

Imię i nazwisko oraz afiliacja Promotora pomocniczego:

Filip Rudziński, afiliacja jak wyżej .....

Rekomendacja rozprawy do konkursu, uwzględniająca w szczególności:

- 1) znaczenie rozprawy i jej wkład do Sztucznej Inteligencji,
- 2) innowacyjny i oryginalny charakter otrzymanych wyników,
- 3) możliwości zastosowań uzyskanych wyników,
- 4) zrealizowane publikacje związane z rozprawą.

Zgłoszona do konkursu rozprawa poświęcona jest uogólnionym, samoorganizującym się sieciom neuronowym o ewoluujących, drzewopodobnych strukturach topologicznych oraz ich zastosowaniom do budowy systemów grupowania danych (analizy skupień) w złożonych, wielowymiarowych zbiorach danych, w tym grupowania danych medycznych opisujących ekspresję genów.

Wyznaczanie liczby skupisk danych oraz ich prototypów jest jednym z podstawowych celów procesu grupowania danych, który – co należy podkreślić – zazwyczaj odbywa się przy braku jakiegokolwiek wstępnej wiedzy o właściwościach zbioru danych (w szczególności o liczbie skupisk, ich wielkości, kształcie, wzajemnej lokalizacji, gęstości zawartych w nich danych, separowalności, poziomie zaszumienia, itp.). Zdecydowana większość istniejących metod grupowania danych wymaga określenia z góry liczby skupisk danych. W związku z tym, konieczne jest wielokrotne powtarzanie procesu grupowania dla różnych, zakładanych z góry, ilości skupisk, a następnie wybór określonego rozwiązania z wykorzystaniem różnorodnych mniej lub bardziej subiektywnych wskaźników jakości grupowania. Ponadto, metody wykorzystujące prototypy skupisk danych zwykle generują po jednym punktowym prototypie dla każdego skupiska. Tego typu prototypy są bardzo niedoskonałą reprezentacją skupisk o nieregularnych kształtach, w tym „cienkich” skupisk szkieletowych.

Mając na uwadze powyższe i szereg innych mankamentów tradycyjnych metod grupowania danych, w rozprawie zaproponowano odmienne podejście do tego problemu, mające swe korzenie w obszarze współczesnych systemów sztucznej inteligencji (tzw. systemów "inteligencji obliczeniowej", ang. *computational intelligence*). Główna metoda badawcza rozprawy polega na wykorzystaniu proponowanych uogólnionych samoorganizujących się sieci neuronowych o ewoluujących, drzewopodobnych strukturach jako narzędzia do automatycznego wykrywania liczby skupisk danych w złożonych, wielowymiarowych zbiorach danych oraz automatycznego generowania ich wielopunktowych prototypów będących podstawą grupowania. Proponowane uogólnione sieci neuronowe wyposażone są w mechanizmy automatycznego podziału ich struktur topologicznych na podstruktury, ponownego łączenia określonych podstruktur oraz automatycznej modyfikacji liczby

neuronów sieci w celu jak najlepszego dopasowania struktur sieci do rozkładu oraz charakteru (w tym kształtów, rozmiarów, gęstości i lokalizacji w przestrzeni danych) skupisk danych w rozważanym zbiorze.

Rozprawa demonstruje również – w obszerny i pogłębiony sposób – praktyczne wykorzystanie proponowanego, oryginalnego narzędzia teoretycznego w różnorodnych i złożonych problemach grupowania danych z wykorzystaniem najpierw szeregu standardowych syntetycznych zbiorów dwu- i trójwymiarowych danych, następnie trzech złożonych, rzeczywistych zbiorów wielowymiarowych danych dostępnych na serwerze Uniwersytetu Kalifornijskiego w Irvine, a ostatecznie czterech zbiorów danych medycznych opisujących ekspresję genów, dostępnych na serwerze Uniwersytetu z Shenzhen. Zbiory te reprezentują problemy diagnozowania chorób nowotworowych układu krwionośnego (zbiór *Leukemia* zawierający 3571 atrybutów numerycznych i 72 rekordy oraz zbiór *Mixed Lineage Leukemia (MLL)* zawierający 2474 atrybuty numeryczne i 72 rekordy), jelita grubego (zbiór *Colon* zawierający 1096 atrybutów numerycznych i 62 rekordy) oraz układu limfatycznego (zbiór *Lymphoma* zawierający 4026 atrybutów numerycznych i 62 rekordy). W przypadku danych medycznych, dokonano oceny efektywności grupowania zarówno dla tzw. grupowania bazującego na próbkach jak i - co istotniejsze - bazującego na genach. W tym ostatnim przypadku wykorzystano bazę genów wzorcowych DAVID (ang. *Database for Annotation, Visualization and Integrated Discovery*) oraz współpracujące z nią narzędzia do statystycznej analizy funkcyjnej genów DAVID Functional Annotation Tool. Narzędzie to posłużyło do wykonania dokładnych testów Fishera dla badanych grup genów - patrz szczegóły w rozprawie. Przeprowadzono również obszerną analizę porównawczą z szeregiem innych, do pewnego stopnia alternatywnych, metodologii grupowania danych (łącznie rozważono 5 różnych podejść) wskazując na istotne zalety proponowanego w rozprawie podejścia, m.in. w efektywnym wykrywaniu grup genów o spójnym profilu ekspresji (ang. *coexpressed genes*) odpowiedzialnych za powstawanie określonych chorób.

Reasumując, należy stwierdzić, że rozprawa wnosi istotny wkład w rozwój współczesnych metod sztucznej inteligencji zarówno w obszarze narzędzi teoretycznych jak i oryginalnych i różnorodnych ich zastosowań ze szczególnym uwzględnieniem zastosowań w badaniach medyczno-genetycznych.

Część wyników zawartych w rozprawie przedstawiana była na renomowanych międzynarodowych konferencjach naukowych takich jak *2016 World Congress on Computational Intelligence* w Vancouver - pozycja [4] (patrz poniższy wykaz) oraz *International Conference on Artificial Intelligence and Soft Computing 2014 i 2015* - pozycje [1], [2].

Wyniki innych, nie zawartych w rozprawie, zastosowań proponowanego narzędzia teoretycznego przedstawiono na *2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology* w Gijón - pozycja [3] (grupowanie dokumentów WWW). Ponadto, na *2018 IEEE International Symposium on Industrial Electronics* zgłoszono pracę [5], prezentującą zastosowanie proponowanego narzędzia do grupowania szeregów czasowych opisujących obciążenie systemu energetycznego przez różne grupy odbiorców energii elektrycznej.

Wykaz publikacji związanych z rozprawą:

- [1] Marian B. Gorzałczany, Jakub Piekoszewski, Filip Rudziński, "Generalized Tree-Like Self-Organizing Neural Networks with Dynamically Defined Neighborhood for Cluster Analysis", in *Artificial Intelligence and Soft Computing - ICAISC 2014*, ser. Lecture Notes in Computer Science, L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, and J. M. Żurada, Eds., Berlin: Springer-Verlag, 2014, vol. 8468, pp. 713–725, DOI: 10.1007/978-3-319-07176-3\_62.
- [2] Marian B. Gorzałczany, Jakub Piekoszewski, Filip Rudziński, "Microarray *Leukemia* Gene Data Clustering by Means of Generalized Self-organizing Neural Networks with Evolving Tree-Like Structures", in *Artificial Intelligence and Soft Computing - ICAISC 2015*, ser. Lecture Notes in Computer Science, L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, and J. M. Żurada, Eds, Berlin: Springer-Verlag, 2015, vol. 9119, pp. 15–25, DOI: 10.1007/978-3-319-19324-3\_2.
- [3] Marian B. Gorzałczany, Filip Rudziński, Jakub Piekoszewski, "Generalized SOMs with Splitting-

Merging Tree-Like Structures for WWW-Document Clustering", in *2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (IFSA-EUSFLAT-15)*, J. M. Alonso, H. Bustince, and M. Reformat, Eds., Advances in Intelligent Systems Research, vol. 89, Gijón, Spain: Atlantis Press, June 30 – July 3 2015, pp. 186–193, DOI: 10.2991/ifsa-eusflat-15.2015.29.

- [4] Marian B. Gorzalczany, Jakub Piekoszewski, Filip Rudziński, "Gene expression data clustering using tree-like SOMs with evolving splitting-merging structures", in *2016 IEEE International Conference on Neural Networks (IJCNN)*, 25-29 July 2016, Vancouver, Canada, 2016, pp. 3666–3673 [part of *2016 World Congress on Computational Intelligence (WCCI)*], DOI: 10.1109/IJCNN.2016.7727671.
- [5] Marian B. Gorzalczany, Jakub Piekoszewski, Filip Rudziński, "Electricity consumption data clustering for load profiling using generalized self-organizing neural networks with evolving splitting-merging structures", *2018 IEEE International Symposium on Industrial Electronics (IEEE ISIE)*, 12-15 June 2018, Cairns, Australia, submitted.



.....  
podpis Promotora/ki