

## INFORMACJA O PROPONOWANEJ DO OTWARCIA ROZPRAWIE DOKTORSKIEJ

### **Metody tworzenia heterogenicznych zespołów modeli predykcyjnych z wykorzystaniem głębokiego uczenia**

(ang. "Methods for creating heterogeneous ensembles of prediction models using deep learning")

Doktorant: mgr inż. Mateusz Piwowarczyk

Promotor: dr hab. inż. Bogdan Trawiński, Prof. PWr.

#### *Uzasadnienie podjęcia tematu*

Duża ilość danych gromadzonych z wielu źródeł w ostatnich latach przyczyniła się do poszukiwania i rozwoju metod pozwalających na ich szybką i zarazem dokładną analizę. Duże znaczenie w procesie tym odgrywać zaczynają metody oparte o głębokie sieci neuronowe [6]. Jednocześnie wciąż w wielu zastosowaniach istotną rolę odgrywają metody oparte o kombinację dużej liczby prostych klasyfikatorów składających się na zespoły modeli uczenia maszynowego [4, 5]. Liczne badania pokazują, że modele złożone z wielu różnorodnych klasyfikatorów potrafią uzyskiwać lepsze rezultaty niż te, wykorzystujące jedynie pojedyncze klasyfikatory (nawet jeśli są one bardzo zaawansowane). Wciąż napływające strumienie danych i potrzeba ciągłego reagowania na nie nie pozwalają jednak na ich zapis i przechowywanie w celu późniejszych analiz, a problematyka zmiany rozkładu tych danych w czasie (concept drift) wymusza konieczność odpowiedniego dostosowywania się algorytmów do nowych sytuacji [7].

#### *Cel rozprawy*

Opracowanie skutecznej, adaptacyjnej metody tworzenia heterogenicznych zespołów modeli predykcyjnych dla dużych strumieni danych z wykorzystaniem technik głębokiego uczenia.

#### *Teza rozprawy*

Zastosowanie metod głębokiego uczenia pozwala budować lepsze (w świetle wybranych kryteriów oceny) heterogeniczne zespoły modeli predykcyjnych dla dużych strumieni danych.

#### *Metodyka badań*

- Analiza różnorodnych metod tworzenia zespołów modeli predykcyjnych,
- Analiza metod przetwarzania dużych strumieni danych
- Analiza wykorzystania metod głębokiego uczenia do konstrukcji zespołów modeli predykcyjnych
- Opracowanie i dobór miar i kryteriów oceny skuteczności zaproponowanych metod
- Badania porównawcze opracowanych metod i algorytmów z istniejącymi algorytmami przetwarzania strumieni danych
- Opracowanie metodologii analizy statystycznej uzyskanych wyników (w tym dobór testów istotności statystycznej)

### Zakres rozprawy

1. Analiza różnych podejść do łączenia w zespoły modeli uczenia maszynowego za pomocą głębokiego uczenia
2. Analiza różnych podejść do tworzenia zespołów modeli uczenia maszynowego z wykorzystaniem metod głębokiego uczenia
3. Opracowanie metod reagowania na zmianę rozkładu danych w strumieniu danych (concept drift)
4. Opracowanie metod zapobiegających przeuczeniu w strumieniach danych
5. Opracowanie metod walidacji modeli uczenia maszynowego dla strumieni danych
6. Zastosowanie opracowanych algorytmów do problemu rozpoznawania lokacji obiektów
7. Propozycja wykorzystania kapsułkowych sieci neuronowych w celu budowy zespołów uczenia maszynowego

### Uzyskane wyniki

Prezentowane w literaturze badania na temat poprawy skuteczności predykcji zespołów modeli uczenia maszynowego wykorzystujących w swoich procesach głębokie sieci neuronowe wykazują ich lepszą skuteczność w porównaniu do prostych klasyfikatorów oraz zespołów modeli opierających się jedynie o takie klasyfikatory [2, 3]. Udowodniono również poprawę jakości modeli uzyskanych przez zastosowanie zespołów modeli uczenia maszynowego w celu optymalizacji hiperparametrów dla głębokich sieci uczenia maszynowego [1]. Potwierdzono wysoką skuteczność kapsułkowych sieci neuronowych w zakresie rozpoznawania pisma (ze szczególnym uwzględnieniem nakładających się znaków) [8].

### Literatura

- [1] Qiu, X., Zhang, L., Ren, Y., Suganthan, P.N., Amaratunga, G.: *Ensemble deep learning for regression and time series forecasting*. CIEL 2014. IEEE Symposium on Computational Intelligence in Ensemble Learning, pp. 1-6 (2014)
- [2] Xiao, Y., et al.: *A Deep Learning-Based Multi-Model Ensemble Method for Cancer Prediction*. Computer Methods and Programs in Biomedicine 153, pp. 1-9 (2018)
- [3] Li, S., Liu, G., Tang, X., Lu, J., Hu, J.: *An Ensemble Deep Convolutional Neural Network Model with Improved D-S Evidence Fusion for Bearing Fault Diagnosis*. Sensors 17(8), 1729 (2017)
- [4] B. Trawinski, Z. Telec, J. Krasnoborski, M. Piwowarczyk, M. Talaga, T. Lasota, and E. Sawitow, 2017. *Comparison of expert algorithms with machine learning models for real estate appraisal*, 2017 IEEE International Conference on Innovations in Intelligent Systems and Applications (INISTA).
- [5] Trawiński, B., Lasota, T., Kempa, O., Telec, Z., Kutrzyński, M.: *Comparison of Ensemble Learning Models with Expert Algorithms Designed for a Property Valuation System*. N.T. Nguyen et al. (Eds.): ICCCI 2017, Part I, LNAI 10448, pp. 317–327, Springer (2017)
- [6] G. Ian, B. Yoshua, C. Aaron, *Deep Learning*. *Współczesne systemy uczące się*, Wydawnictwo Naukowe PWN (2018)
- [7] B. Krawczyk, L. Minku, J. Gama, J. Stefanowski, M. Woźniak, *Ensemble learning for data stream analysis: A survey*, Information Fusion 37, pp. 132–156 (2017)
- [8] Sabour, S., Frosst, N., Hinton, G.E.: *Dynamic routing between capsules*. NIPS 2017, Proceedings of the 31st Conference on Neural Information Processing Systems (2017)