

Warszawa, 27 sierpnia 2026

prof. dr hab. inż. Jarosław Arabas

Politechnika Warszawska

Wydział Elektroniki i Technik Informacyjnych

Instytut Informatyki

RECENZJA

rozprawy doktorskiej mgr Sławomira Smyła

pt. „Methods for improving performance
of neural network-based forecasting models”

1. Zakres tematyczny, cel i teza rozprawy

Rozprawa jest poświęcona metodom podniesienia efektywności rekurencyjnych sieci neuronowych, przede wszystkim stanowiących rozszerzenie modelu LSTM, w zadaniach prognozowania szeregów czasowych, ze szczególnym uwzględnieniem prognozowania zapotrzebowania na energię elektryczną.

W rozprawie Autor bazuje na własnych osiągnięciach, polegających na odpowiednim ustawieniu, zestrojeniu i niezbędnych modyfikacjach metod maszynowego uczenia połączonych z podejściami statystycznymi. W efekcie tych zabiegów Autor w roku 2018 wygrał międzynarodowy konkurs M4, stanowiący najważniejsze wydarzenie weryfikujące użyteczność metod prognozowania szeregów czasowych. Wynik ten sam w sobie stanowiłby przesłankę do wystąpienia o stopień doktora, gdyż stanowi istotne osiągnięcie konstrukcyjne. Rozprawa jednak nie zatrzymuje się w roku 2018, lecz przedstawia pogłębioną analizę wykonalności metod prognozowania zapotrzebowania na energię elektryczną ze szczególnym uwzględnieniem sieci rekurencyjnych jako narzędzia pozwalającego na analizę

złożonych zależności czasowych, z oddziaływaniem na moment prognozy wielkości znacznie oddalonych w historii.

Rozwiązania dopasowujące „czyste” sieci neuronowe do rozważanych zagadnień Autor przedstawia w formie tezy, która głosi (parafrazując), że modele neuronowe będą bardziej efektywnie prognozować gdy zostaną użyte po właściwym przetworzeniu wstępnym danych oraz zostanie właściwie dobrana architektura i metoda uczenia, a także, gdy wykorzystane zostanie podejście mieszaniny ekspertów, polegające na zbudowaniu wielu niezależnych modeli i późniejszej agregacji generowanych przez nie prognoz. Teza ta wydaje się dość oczywista w kontekście twierdzenia No Free Lunch dla maszynowego uczenia, ale jej oczywistość staje się nieoczywista w świetle praktyki stosowanej nierzadko w literaturze przedmiotu. Znaczny odsetek publikacji koncentruje się na próbie wytworzenia modelu dostatecznie uniwersalnego (zatem działającego w trybie prawie idealnej czarnej skrzynki), aby mógł efektywnie dokonywać prognozy dla szerokiego spektrum danych, pracując na danych surowych. Jednak ceną, jaką płaci się za próbę stworzenia takiego ogólnego modelu, jest jego ogromny stopień komplikacji, co czyni go nienauczalnym chociażby z powodu niedostatecznie ogromnej ilości danych, z którą nierzadko mamy do czynienia w przypadku prognoz szeregów czasowych np. energii elektrycznej.

Autor idzie inną drogą – zamiast usiłować stworzyć idealny model prognozowania dla danych surowych, stara się właściwie przetworzyć dane, aby ułatwić proces modelowania (a zatem bliżej mu raczej do podejścia szarej skrzynki). Podejście to jest w mojej ocenie bardzo inżynierskie i zasługuje na uznanie. Moje własne doświadczenia praktyczne związane z modelowaniem z użyciem sieci neuronowych, a w szczególności prognozowaniem szeregów czasowych, pozwalają mi w pełni podpisać się pod sformułowaną w rozprawie tezę.

Tematyka rozprawy dotyczy zagadnień modelowania z użyciem metod maszynowego uczenia, a w szczególności rozwoju architektury sieci LSTM, oraz konstrukcji potoków przetwarzania z ich użyciem. Z tych powodów mogę jednoznacznie stwierdzić, że tematyka ta mieści się w dyscyplinie Informatyka Techniczna i Telekomunikacja.

Rozprawa została przygotowana na podstawie cyklu 8 artykułów opublikowanych w międzynarodowych czasopismach oraz dwóch referatów konferencyjnych. Artykuły te są ściśle powiązane tematycznie, a w rozprawie Autor zamieścił jeszcze krótki tekst w języku angielskim stanowiący komentarz i uzupełnienie treści zawartych w artykułach, zwany dalej tekstem wstępnym. Spełniona jest zatem przesłanka Art 187.3 Prawo o szkolnictwie wyższym i nauce, tekst jednolity, Dz.U. 2024/1571, stanowiąca o dopuszczalnej formie rozprawy. Należy przy tym wzmiankować, że miejsca publikacji są bardzo poważane w środowisku maszynowego uczenia – są to IEEE Transactions on Neural Networks and

Learning Systems oraz International Journal of Forecasting (po 2 prace), Information Fusion, Neural Networks, International Joint Conference on Neural Networks oraz Neural Information Processing (po 1 pracy).

Problemem naukowym, o którego oryginalnym rozwiązaniu jest mowa w ustawie w Art. 187.2. Prawo o szkolnictwie wyższym i nauce, tekst jednolity, Dz.U. 2024/1571, jest metodyka konstrukcji potoku przetwarzania danych, w celu dokonania efektywnego prognozowania szeregu czasowego, oraz modyfikacje rekurencyjnej sieci neuronowej bazującej na LSTM. Rozwiązanie to jest oryginalne, gdyż w literaturze nie ma innych rozwiązań zgodnych z zaproponowanymi przez Autora, a jego rozwiązanie wygrało konkurs M4.

2. Krytyczna analiza stanu wiedzy

Autor dokonuje przeglądu literatury przedmiotu w rozdziale 3 tekstu wstępnego. W bardzo elegancki sposób i ze swadą omawia rozwój sieci neuronowych, wychodząc od perceptronu. Szczególnie podoba mi się podrozdział poświęcony alternatywnym sposobom uwzględnienia dynamiki: przez stan i przez oknowanie, z atencją. Mechanizm uwagi jest wprowadzany bardzo oryginalnie, bo w odseparowaniu od architektury transformera, z którą w większości literatury jest syjamsko zrośnięty. Autor wyjaśnia również powody, dla których mechanizm uwagi zastosowany wprost nie znajduje zastosowania w prognozowaniu szeregów czasowych. Autor nie zapomina o tradycyjnych statystycznych modelach szeregów czasowych, takich jak ARIMA, a także o innych niż sieci neuronowe modelach nieliniowych, takich jak SVM oraz drzewa regresyjne. W przeglądzie literatury z uznaniem odnotowałem fakt, że autor wskazuje na korzyści z użycia jako funkcji straty takiej miary jakości, według której model będzie finalnie oceniany (np. przyjęcie jako funkcji straty wartości kosztów pomyłek), jednocześnie odnotowując fakt, że w literaturze jest raczej rzadkie.

Przegląd literatury świadczy o bardzo rozległej wiedzy Autora w obszarze dotyczącym tematyki rozprawy oraz wycuciu stopnia istotności, możliwości i ograniczeń różnorodnych pomysłów związanych z modelowaniem szeregów czasowych w obszarze maszynowego uczenia.

3. Osiągnięcia przedstawione w rozprawie

Praca ma charakter konstrukcyjny. Poza tekstem wstępnym, każdy z artykułów stanowiących rozprawę ma podobną, typową dla publikacji naukowej strukturę.

Artykuł [D1] opisuje rozwiązanie, które było zastosowane w wygranym przez Autora konkursie M4. Składają się na nie:

1. usunięcie trendów i zmian sezonowych poprzez zastosowanie wygładzania wykładniczego

Zastosowane rozwiązanie zakłada usuwanie trendu i sezonowości nie jednorazowo (jak to się dzieje typowo w przypadku usuwania trendów podczas prognozowania modelami liniowymi), ale dla każdej epoki procesu uczenia. W mojej ocenie jest to trafny wybór, gdyż pozwala na obsłużenie sytuacji, gdy trend lub sezonowość ulegają zmianom w czasie.

Usuwanie trendu jest wykonywane poprzez wygładzanie wykładnicze danych oryginalnych, natomiast usuwanie sezonowości – poprzez wygładzanie wykładnicze danych oryginalnych podzielonych przez współczynnik sezonowości. Dane są dzielone przez wartość będącą wynikiem wygładzania wykładniczego (metoda Holta-Wintersa).

2. użycie sieci neuronowej dla danych transformowanych

Autor dokonuje prognozowania zlogarytmowanych wartości szeregu, po czym prognozy są retransformowane przez funkcję wykładniczą. W ten sposób niwelowane jest naturalne dla sieci LSTM ograniczenie zakresu generowanych prognoz, wynikające z zastosowania na wyjściu funkcji sigmoidalnych.

3. wykonanie prognozy na podstawie sieci LSTM „z dylatacją” (dilated LSTM)

Autor wykorzystuje stos sieci LSTM z dylatacją, czyli takich, w których wartość wejściowa modyfikująca stan pochodzi sprzed kilku chwil czasowych szeregu, a nie z ostatniej chwili.

Tu komentarz polemiczny recenzenta – rozwiązanie takie świadczy chyba o niepełnym zaufaniu do jakości metod usuwania sezonowości. Gdyby usunięcie sezonowości było doskonałe, wprowadzenie dylatacji byłoby nadmiarowe.

4. użycie funkcji oceny pochodzącej z benchmarku jako funkcji straty

Wybór ten był już przeze mnie odnotowany z uznaniem – jestem przekonany po konieczności postępowania zgodnie z tą zasadą. Jednak rozwiązanie takie może powodować trudności obliczeniowe, jeśli używany jest w uczeniu algorytm gradientowy, gdyż nie dla każdej funkcji może być dostępny gradient. Autor nie odnosi się do tej kwestii w

artykule, a korzysta z uczenia metodą stochastycznego gradientu. Ponieważ funkcją straty jest „pinball loss”, jej różniczkowalność jest zapewniona prawie wszędzie.

5. wykorzystanie komitetu modeli

Autor miesza modele, z których każdy powstał jako wynik uczenia z innym ziarnem generatora liczb pseudolosowych, lub jako tzw. mieszanina specjalistów, czyli zbiór modeli wytworzonych dla różnych, ale podobnych do siebie zadań prognozowania. Prognoza powstaje poprzez agregację prognoz indywidualnych, np. poprzez uśrednienie.

Komentarz polemiczny- podejście mieszaniny specjalistów jest w pełni uzasadnione, gdy mówimy o udziale w konkursie z dużą liczbą szeregów czasowych do prognozowania. Jednak nierzadko, w praktyce, taka duża liczba jest niedostępna, i wówczas podejście to wydaje się nie mieć zastosowania.

Artykuł [D2], wcześniejszy o 2 lata od [D1], dotyczy prognozowania szeregu czasowego zapotrzebowania na energię elektryczną. Wykorzystywana jest mieszanina trzech modeli – lasu losowego, *gradient boosting*, oraz perceptronu wielowarstwowego z połączeniami skrótowymi (ResNet). Zadanie polega na prognozowaniu pełnego zapotrzebowania miesięcznego, w rozdzielczości godzinowej. W przypadku sieci neuronowej, zostało to przeformułowane na 168 prognoz dla każdej godziny, każdego dnia tygodnia. Wejściem modelu były statystyki poprzedniego roku i poprzedniego miesiąca, a także informacja o kalendarzu oraz dniach świątecznych. Sposób wstępnego przetworzenia danych oraz sformułowania zadania prognozowania jest w mojej ocenie najciekawszym aspektem artykułu. Wynikowy system zajął 2. miejsce w konkursie Global Energy Forecasting Competition 2017.

Uwaga polemiczna – w przypadku dni świątecznych, zapotrzebowanie zazwyczaj jest niższe. Skutkować to może obniżeniem wartości prognozy, która bazuje na wartościach zapotrzebowania z dnia świątecznego (efekt bazy). Zabrakło mi odniesienia się do tego efektu.

Artykuł [D3] omawia wynik zastosowania metodyki z pracy [D1] oraz jej nieznaczonej modyfikacji, do prognozowania zapotrzebowania na energię elektryczną. Dane pochodzą z 35 krajów europejskich. Modyfikacje dotyczyły uwzględnienia w funkcji celu, opartej na *pinball loss*, kary za zbyt dużą dynamikę (*wiggleness loss*) oraz uproszczonej architektury LSTM. Bardzo wartościowym elementem pracy jest porównanie jakości wyników, uzyskanych za pomocą zmodyfikowanej metody zastosowanej do M4, z wynikami uzyskanymi przez wiele różnych metod statystycznych i maszynowego uczenia (w tym: ARIMA, warianty regresji kNN, regresja jądrowa, perceptron wielowarstwowy, prosty LSTM).

Zaproponowana technika modelowania uzyskała najmniejszą wartość rozstępu międzykwartylowego, przy jednocześnie niskiej wartości mediany średniego błędu procentowego.

Efektywności metody opisanej w [D3] może być zwiększona poprzez umiejętne wstępne przetwarzanie szeregu czasowego, o czym traktuje praca [D4]. Analizowane wstępne przetwarzanie polega na odjęciu wartości średniej i podzieleniu przez standardowe odchylenie wartości średniego zapotrzebowania miesięcznego, które jest wykorzystywane w usuwaniu trendów sezonowych. Zarówno średnia, jak i standardowe odchylenie, są obliczane dla 12 próbek obejmujących rok kalendarzowy.

Artykuł [D5] dotyczy zagadnienia prognozowania zapotrzebowania na energię elektryczną w trybie prognozy dnia następnego. W stosunku do wcześniejszych prac, główna zmiana w metodyce prognozowania dotyczy innego sposobu podziału danych w szeregu, uwzględniającego sezonowość tygodniową. Z kolei w architekturze modelu wprowadzona jest zmiana w sposobie uwzględnienia dylatacji w sieci LSTM.

Praca [D6] pogłębia analizę autorskiej architektury LSTM z dylatacją, poprzez wprowadzenie mechanizmu uwagi. Uwaga jest realizowana przez ustalenie wag związanych z kolejnymi pozycjami wektora wejściowego modelu, i jest realizowana poprzez dedykowaną architekturę LSTM, której celem jest nauczenie się wartości wag uwagi. Testy proponowanego podejścia zostały przeprowadzone w sposób analogiczny jak w artykule [D5]. W wyniku analizy empirycznej potwierdzono poprawę efektywności prognozowania w stosunku do [D5].

Artykuł [D7] rozszerza badania przedstawione w [D6] poprzez wprowadzenie mieszanki modeli zgodnych z opracowanymi w [D6], przy sposobie testowania i modelach porównawczych analogicznych do użytych w [D6].

Kończąca cykl praca [D8] wzbogaca rozwijane podejście o dane generowane losowo, będące kwantylami rozkładu wartości szeregu obserwowanych na wejściu. Podejście to, zaproponowane przez innych badaczy, zostało użyte do obliczenia wartości funkcji straty nie tylko dla prognozowanej wielkości, ale również dla jej modyfikacji o wartość kwantyla. W ten sposób możliwa jest dokładniejsza wycena pomyłek systemu prognostycznego. Podobnie jak w poprzednich tekstach, efektywność podejścia została pozytywnie zweryfikowana dla szeregów czasowych zapotrzebowania w 35 europejskich krajach.

Zaprezentowane wyniki, sposób ich uzyskania poprzez staranną i statystycznie wiarygodną analizę empiryczną, a także wprowadzanie kolejnych ciekawych modyfikacji, bardzo dobrze świadczą o warsztacie Kandydata. Wraz z prawidłowo wykonanym przeglądem literatury, świadczą one o tym, że recenzowana rozprawa prezentuje ogólną wiedzę teoretyczną

Kandydata w dyscyplinie oraz umiejętność samodzielnego prowadzenia pracy naukowej (por. wymagania ustawy Prawo o szkolnictwie wyższym i nauce, tekst jednolity, Dz.U. 2024/1571, Art. 187. 1).

4. Wkład rozprawy w rozwój nauki

Podstawowy wkład w rozwój nauki wiąże się z przygotowaniem i publikacją efektywnej metody prognozowania szeregów czasowych.

Meritum rozprawy uważam za bardzo wartościowe. Na uznanie zasługuje fakt, że Autor jest zwycięzcą lub stanął na podium międzynarodowych konkursów poświęconych prognozowaniu. Miejscami publikacji wyników są czasopisma i konferencje, w których ocena na etapie recenzji jest rygorystyczna, a uzyskanie publikacji trudne.

5. Wniosek końcowy

Rozprawę oceniam pozytywnie, zarówno w świetle obowiązującej ustawy Prawo o szkolnictwie wyższym i nauce, jak i w świetle wymagań zwyczajowych. Wnoszę o dopuszczenie Doktoranta do dalszych etapów postępowania doktorskiego. Jednocześnie, ze względu na znaczące osiągnięcia Autora związane z jego udziałem w konkursach oraz wyróżniające się publikacje, wnoszę o wyróżnienie rozprawy.