

Wrocław, 25 stycznia 2023

Dr hab. inż. Mirosław Klinkowski, prof. IŁ-PIB
Instytut Łączności – Państwowy Instytut Badawczy
Oddz. we Wrocławiu

RECENZJA

rozprawy doktorskiej mgra inż. Daniela Szostaka
„Prognozowanie ruchu w sieciach optycznych
posiadających ustalony poziom ruchu”
Promotor: prof. dr hab. inż. Krzysztof Walkowiak

1. Cel, zakres i charakter rozprawy

Recenzowana praca dotyczy sieci teleinformatycznych, w szczególności problematyki przewidywania (predykcji) ruchu w sieciach szkieletowych opartych na technologii elastycznych sieci optycznych (ang. elastic optical network, EON). Sieci EON pozwalają na realizację połączeń dopasowanych do wymagań ruchowych dzięki zastosowaniu techniki transmisji na wielu nośnych (ang. multi-carrier transmission) umożliwiając elastyczne tworzenie kanałów transmisyjnych w ramach ustalonych, dyskretnych poziomów przepływności dostępnych z określoną ziarnistością. W tym kontekście, głównym celem rozprawy jest opracowanie i analiza metod służących krótkookresowej oraz długookresowej predykcji poziomów ruchu dla potrzeb efektywnego zarządzania i optymalizacji takich sieci. Obszar zastosowań predykcji krótkookresowej to m.in. sterowanie obciążeniem łączy transmisyjnych i przeciwdziałanie stanom przeciążenia, które mogą prowadzić do pogorszenia jakości usług w sieci, jak i wykrywanie nieprzewidzianych przepływów ruchu wynikających z nieprawidłowego działania sieci. Predykcja długookresowa może z kolei wspierać operatora sieci w procesie planowanej rekonfiguracji/rozbudowy sieci oraz być wykorzystywana w mechanizmach routingu. Ze względu na specyfikę sieci EON, proponowane metody mają na celu predykcję predefiniowanych poziomów ruchu, natomiast nie służą dokładnemu przewidywaniu natężenia ruchu. Opracowanie takich metod nie było tematem prac badawczych w literaturze.

Tematykę rozprawy oceniam jako aktualną i istotną dla sektora ICT. Rozprawa ma charakter zarówno teoretyczny jak i użyteczny. W zakresie rozważań teoretycznych, mgr inż. Daniel Szostak zaproponował i opracował nowe strategie i metody krótkookresowej i długookresowej predykcji poziomów ruchu w sieciach EON wykorzystujące algorytmy uczenia maszynowego (ang. machine learning, ML) jak i oparte na szeregach czasowych (ang. time series, TS). Aspekt użyteczny rozprawy jest związany z implementacją proponowanych metod w środowisku Python, z zastosowaniem dostępnych w tym środowisku bibliotek (m.in. scikit-learn), oraz przeprowadzeniem badań i analiz numerycznych z wykorzystaniem opracowanych syntetycznych i rzeczywistych zbiorów danych ruchowych w oparciu o istniejącą i zaproponowaną metrykę jakości predykcji. Badania te pozwoliły na ocenę i porównanie skuteczności analizowanych strategii i algorytmów przy krótkookresowej i długookresowej predykcji poziomów ruchu sieciowego w różnych scenariuszach ruchowych.

WPLYNĘŁO

09-02-2023

RDN-IT/30/2023

2. Zawartości rozprawy

Rozprawa składa się z sześciu rozdziałów, które można podzielić na trzy części: wprowadzającą (rozdziały 1 i 2), prezentującą i analizującą proponowane metody (rozdziały 3–5) oraz podsumowującą (rozdział 6).

- Rozdział 1 stanowi wstęp przedstawiający uzasadnienie tematu rozprawy, tezę doktoratu, główne osiągnięcia, oraz zakres pracy. W części wstępnej zawarta jest też lista publikacji uzyskanych w trakcie realizacji pracy doktorskiej.
- W rozdziale 2 przedstawiono podstawy teoretyczne zagadnień związanych z tematem rozprawy, w tym dotyczące monitoringu ruchu w sieciach teleinformatycznych, technologii elastycznych sieci optycznych, jak i metod uczenia maszynowego i szeregów czasowych. W tym zakresie dokonano też przeglądu literaturowego związanego z problematyką predykcji ruchu oraz zastosowaniem metod uczenia maszynowego w sieciach optycznych.
- Rozdział 3 zawiera sformułowanie problemu badawczego wraz ze szczegółowym opisem modelu sieci, zbiorów danych, proponowanych strategii predykcji ruchu oraz rozważanych metryk oceny jakości predykcji.
- Rozdziały 4 i 5 prezentują wyniki badań numerycznych związanych z rozważanym problemem krótkookresowej i długookresowej predykcji poziomów ruchu.
- Rozdział 6 podsumowuje rozprawę oraz przedstawia w punktach plan dalszych badań.

Praca została napisana w języku angielskim na poziomie nie stwarzającym większych trudności w zrozumieniu jej treści oraz prezentowanych rozwiązań. Ogólna struktura pracy jest prawidłowa, z jedną uwagą dotyczącą tytułu rozdziału 3 („Problem formulation”), który w mojej ocenie powinien zostać rozszerzony (np. „Research methodology and proposed methods”) ze względu na zakres i charakter przedstawionej w nim treści. Osobna uwaga dotyczy jakości prezentacji pracy, która w kilka aspektach może zostać poprawiona, do czego w szczegółach odnoszę się w punkcie dotyczącym uwag krytycznych.

3. Poprawność i oryginalność postawionej tezy

Teza pracy jest sformułowana w następujący sposób:

There exist methods for short-term and long-term traffic forecast in optical networks, where transmission bases on predefined traffic levels.

która można przetłumaczyć następująco:

Istnieją metody krótkoterminowej i długoterminowej predykcji ruchu w sieciach optycznych realizujących transmisję z predefiniowanymi poziomami przepływności.

W mojej ocenie teza pracy jest oryginalna i została sformułowana poprawnie. Mgr inż. Daniel Szostak na podstawie przeglądu literaturowego i własnej wiedzy prawidłowo określił zakres swojej pracy, koncentrując się na istotnych aspektach związanych z problemem predykcji poziomów ruchu w elastycznych sieciach optycznych, w tym na zagadnieniu predykcji krótkookresowej oraz długookresowej.

Teza jest wykazana w rozprawie poprzez:

1. Opracowanie trzech różnych strategii wykorzystujących algorytmy uczenia maszynowego i szeregów czasowych w celu krótkookresowej i długookresowej predykcji poziomów ruchu.
2. Opracowanie zbiorów danych, w oparciu o syntetyczne i rzeczywiste dane przepływów ruchu sieciowego, dla potrzeb badanych algorytmów ML i TS.
3. Zdefiniowanie metryki oceny jakości predykcji dla rozważanego problemu predykcji poziomów ruchu.
4. Implementację metod i przeprowadzanie badań numerycznych wraz z analizą i porównaniem efektywności proponowanych strategii dla różnych algorytmów ML i TS w różnych scenariuszach ruchowych.

Moim zdaniem mgr inż. Daniel Szostak rozwiązał postawiony problem naukowy stosując prawidłowe metody badawcze. Postawione cele zostały w pracy osiągnięte. Dla wykazania tezy pracy Doktorant posłużył się środowiskiem Python, w którym zaimplementował proponowane strategie i metody predykcji poziomów ruchu i następnie dla opracowanych scenariuszy badawczych przeprowadził badania numeryczne. Metodologia przeprowadzonych badań została dokładnie opisana, natomiast wyniki zawarte w pracy są opatrzone odpowiednią analizą i dyskusją.

4. Analiza źródeł (w tym literatury światowej i stanu techniki) świadczącej o dostatecznej wiedzy autora w danej dyscyplinie naukowej

Doktorant dokonał szczegółowego przeglądu bibliograficznego związanego z tematem rozprawy, na podstawie którego w rozdziale 2 zaprezentował wprowadzenie do zagadnień monitoringu ruchu sieciowego, transmisji w szkieletowych sieciach optycznych, metod i algorytmów uczenia maszynowego i szeregów czasowych oraz metryk stosowanych przy ocenie modeli ML. Dodatkowo, przedstawił przegląd literaturowy dotyczący problematyki predykcji ruchu w sieciach teleinformatycznych oraz obszarów zastosowania metod uczenia maszynowego w sieciach optycznych. Lista odniesień do literatury przedmiotu przedstawiona w rozprawie zawiera 139 pozycji.

Odniesienia w tekście do źródeł są odpowiednie i świadczą o dobrej znajomości współczesnej literatury związanej z dyscypliną informatyka techniczna i telekomunikacja. Proponowane przez Doktoranta rozwiązania dla postawionego problemu badawczego świadczą o zaawansowanej wiedzy o charakterze szczegółowym związanej z przetwarzaniem i analizą danych oraz zastosowaniem metod uczenia maszynowego i szeregów czasowych.

5. Pozycja rozprawy w stosunku do stanu wiedzy i stanu techniki reprezentowanych przez literaturę światową

Tematyka rozprawy doktorskiej mgr. inż. Daniela Szostaka dotyczy ważnej problematyki predykcji ruchu w sieciach teleinformatycznych oraz zastosowania metod analizy danych i uczenia maszynowego w celu optymalizacji działania optycznych sieci szkieletowych. W szczególności, poświęcona jest opracowaniu dedykowanych metod predykcji ruchu w rozwijanych w ostatnich latach elastycznych sieciach optycznych. Zaprezentowany przez Doktoranta przegląd literaturowy związany z tematyką pracy wskazuje na aktualność i istotność tej problematyki, a jednocześnie pokazuje, że predykcja ruchu w sieciach EON nie była tematem głębszych badań w literaturze światowej. W szczególności, wcześniejsze prace dotyczące sieci teleinformatycznych skupiały się na predykcji dokładnych rozmiarów ruchu sieciowego, podczas gdy w kontekście sieci EON potrzebne są rozwiązania

uwzględniające ziarnisty charakter dostępnych przepływności w kanałach transmisyjnych, co ma znaczenie np. przy dynamicznym i elastycznym wymiarowaniu połączeń. Stąd też, w mojej ocenie proponowane przez Doktoranta zastosowanie metod uczenia maszynowego oraz szeregów czasowych do krótkookresowej i długookresowej predykcji poziomów ruchu celem ich wykorzystania w sieciach EON jest nowatorskie i wzbogacające aktualny stan wiedzy reprezentowany przez literaturą światową w obszarze sieci teleinformatycznych.

6. Znaczenie uzyskanych wyników dla danej dyscypliny naukowej

Jako najważniejsze oryginalne osiągnięcia rozprawy doktorskiej mgr. inż. Daniela Szostaka należy wymienić:

1. Opracowanie, implementację i analizę metod przetwarzania danych historycznych przepływów ruchu dla potrzeb predykcji poziomów ruchu w sieciach optycznych.
2. Opracowanie zbiorów danych testowych dla celów analizy proponowanych metod w oparciu o syntetyczne i rzeczywiste dane ruchu sieciowego.
3. Opracowanie trzech strategii (LB, RVB i LVB) wykorzystujących algorytmy uczenia maszynowego i szeregów czasowych dla krótkoterminowej i długoterminowej predykcji poziomów ruchu.
4. Opracowanie statycznej i dynamicznej strategii dla długoterminowej predykcji poziomów ruchu.
5. Opracowanie nowej metryki (TLPQ) służącej do oceny jakości predykcji algorytmów, dedykowanej do rozważanego problemu predykcji poziomów ruchu.
6. Analizę i porównanie efektywności proponowanych metod w różnych scenariuszach ruchowych.

Należy podkreślić, że opracowane koncepcje oraz uzyskane wyniki mają znaczenie praktyczne. Doktorant zdefiniował i następnie rozwiązał realny i aktualny problem badawczy związany z przewidywaniem poziomów ruchu w sieciach szkieletowych opartych na technologii EON. Opracowane metody mogą znaleźć zastosowanie m.in. w mechanizmach routingu jak i w algorytmach optymalizacyjnych związanych z elastyczną rekonfiguracją połączeń.

7. Główne wady rozprawy, słabe stron wraz z krytycznymi uwagami szczegółowymi

Uwagi ogólne

1. W mojej ocenie najsłabszą stroną rozprawy jest jakość prezentacji pracy, która w przedstawionej formie w wielu fragmentach sprawia trudność w dokładnym zrozumieniu proponowanych rozwiązań, przeprowadzonych badań, oraz głównych osiągnięć rozprawy. Wiąże się to w dużej mierze z licznnością przeprowadzonych analiz, wykorzystanych metryk, zastosowanych strategii i algorytmów, przyjętych scenariuszy badań oraz zaprezentowanych wyników liczbowych. W szczególności, jakkolwiek wprowadzenie do tematyki rozprawy i prezentacja podstaw teoretycznych w rozdziale 2 została wykonana na bardzo dobrym poziomie pod względem usystematyzowania i klasyfikowania prezentowanych pojęć i metod (m.in. w formie list punktowych), zastosowanie podobnego podejścia zostało ograniczone w dalszej części dot. proponowanych rozwiązań i zrealizowanych badań. Taka forma listy

punktowej, wraz z krótkim opisem (definicją) każdego elementu, mogłaby zostać zastosowana dla przykładu przy definicji odmian metryki TLPO, podczas wprowadzenia strategii statycznej i dynamicznej dla predykcji długookresowej, czy też przy prezentacji rozważanych algorytmów ML. Jednocześnie, rozdziały 4 i 5 mogłyby zostać wzbogacone we wstępie o opis specyfiki badanego w nich problemu (tzn. predykcji krótko- i długo-okresowej) oraz przedstawionej w punktach metodyki badań wraz z wyjaśnieniem zasadności (potrzeby) realizacji poszczególnych kroków. W odniesieniu do struktury pracy, treść rozdziału 3 wydaje się wykraczać poza tytuł nagłówka („problem formulation”), gdyż zawiera wiele dodatkowych informacji dotyczących m.in., modelu sieci, zbiorów danych ruchowych wraz z ich analizą, rozważanych zbiorów cech (features), algorytmów ML, metryk oceny metod, czy też definicje proponowanych i badanych w rozprawie strategii predykcji poziomów ruchu. Także, część szczegółowych wyników analiz numerycznych prawdopodobnie mogłaby zostać przedstawiona w formie aneksu dla „odciążenia” zasadniczej części pracy. Bardziej szczegółowe uwagi związane z prezentacją pracy zamieszczono poniżej.

2. Odnosząc się do treści merytorycznej rozprawy, proponowane dla sieci EON strategie predykcji poziomów ruchu zostały przebadane numerycznie z wykorzystaniem zastosowanych metryk, co niewątpliwie pozwala na ocenę ich efektywności i porównanie ich skuteczności w badanych scenariuszach ruchowych. Niemniej jednak, w podsumowaniu pracy brak jest jednoznacznej konkluzji lub zestawienia metod, które na podstawie wyników przeprowadzonych badań można określić jako najodpowiedniejsze do zastosowania w sieciach EON.
3. Praca nie obejmuje wyników implementacji proponowanych metod w środowisku sieci EON (dla przykładu w symulatorze sieci), co pozwoliłoby na pokazanie skuteczności tych metod w praktycznych zastosowaniach. W tym zakresie, w badaniach nie odniesiono się też do metod referencyjnych z literatury (opisanych w rozdziale 2.3), co nie pozwala dokonać oceny skuteczności proponowanych metod na tle innych rozwiązań. Taki kierunek badań można uwzględnić w dyskusji dotyczącej przyszłych prac.

Uwagi szczegółowe

1. Model sieci w sekcji 3.1 został przedstawiony ze stosunkowo niewielkim rygiorem formalnym. Jakkolwiek topologia sieci została zdefiniowana wykorzystując standardową notację stosowaną w teorii grafów, pozostałe elementy modelu (takie jak interwał czasowy (TI), rozmiar ruchu generowanego pomiędzy parą węzłów, (maksymalny) poziom ruchu w przedziale czasowym, etc.) są zdefiniowane jedynie w formie opisowej. Dla przykładu, prezentacja koncepcji poziomów ruchu zilustrowana na Rysunku 8 mogłaby być wsparta opisem matematycznym definiującym jednoznacznie ten parametr.
2. Pojęcie przepływu ruchu oraz bazowego przepływu ruchu (str. 34) nie jest zdefiniowane w jasny sposób. Czy przepływy ruchu to przepływy zagregowanego ruchu generowanego tylko pomiędzy parą węzłów (tzn. wchodzącego w węzeł źródłowy i wychodzącego w węzeł docelowy), czy uwzględniają one też ruch tranzytowy przesyłany pomiędzy tą parą węzłów?
3. Jak opisano w sekcji 3.2, wybór 5 konkretnych zbiorów danych ruchowych (dataset_1, ..., dataset_5) wynika z różnej wartości metryki MAPE, którą te zbiory się charakteryzują. Czy wartość MAPE dla dataset_5 to wartość maksymalna wśród dostępnych 756 par węzłów i przyjęte do analizy zbiory obejmują cały przedział możliwych wartości MAPE?

4. Nie jest jasne dlaczego dane wejściowe algorytmów (oznaczone czarnym kolorem) na rys. 16-18 różnią się, tzn. mają dyskretne wartości dla LB i LVB, podczas gdy są ciągłe dla RVB. Czy dane wejściowe (ruchowe) nie powinny być takie same dla rozważanych metod?
5. Znaczenie „niebieskiego tła” na rys. 18-22 nie jest dobrze wyjaśnione. Co to znaczy, że „autokorelacja staje się nieznacząca” w tym obszarze?
6. W pracy trudno jest znaleźć dokładniejszy opis oraz znaczenie wyników analizy statystycznej dot. testu statystycznego Friedmana, zaprezentowanej na rys. 26 (jak i wszystkich kolejnych rysunkach tego typu). Czym jest test Friedmana? Co ilustrują poszczególne wartości i linie na tego typu rysunku? Także, czym jest wspomniany na początku rozdziału 4 test Nemenyi? Które wykresy/tabele prezentują wyniki otrzymane dla tego testu?
7. W rozdziałach 3-5 trudno jest znaleźć informację o tym jakiego rodzaju algorytmy TS dla metody szeregów czasowych były badane. Czy to był jeden algorytm, czy było ich więcej?
8. Przy opisie strategii „statycznej” i „dynamicznej” predykcji długookresowej (na początku rozdziału 5) wskazano, że różnią się one zbiorem przyjętych cech (odpowiednio F2 i F1). W tym miejscu dobrze byłoby przywołać wyjaśnienie jak te cechy wiążą się z nazewnictwem przyjętym dla tych strategii, tzn. że cechy w zbiorze F2 reprezentują jakiś element statyczny, podczas gdy cechy w zbiorze F1 odpowiadają jakiejś dynamice.
9. Dla wszystkich przypadków zaprezentowanych w Tabeli 21, czasy treningu i predykcji nie przekraczają 30 milisekund. Czasy te wydają się być bardzo krótkie, stąd warto wyjaśnić jakie mają znaczenie i czy są istotne dla praktycznych zastosowań metod.
10. Badania zostały wykonane dla danych ruchowych wygenerowanych w jednej konkretnej topologii sieci (EURO28). W pracy nie podjęto analizy wpływu topologii i rozmiaru sieci na efektywność poszczególnych metod, w szczególności nie badano czy algorytmy będą zachowywały się podobnie w innych sieciach. Wskazana byłaby dodatkowa dyskusja na ten temat i ewentualnie też uwzględnienie takich badań w przyszłych planach.
11. Rozdzielczość części grafik zamieszczonych w pracy jest zbyt niska, co uniemożliwia lub utrudnia odczytanie wartości liczbowych na wykresach i w tabelach (dotyczy to m.in. Rys. 19 – Rys. 23). Także, zastosowano wiele różnorodnych formatów prezentacji wykresów i tabel – ich ujednolicenie do wspólnego formatu poprawiłoby aspekt wizualny pracy.

Uwagi drobne

1. Przywoływana w pracy technologia OFDM jest tylko jednym z rozwiązań proponowanych dla sieci EON. W istocie ta technologia ma swoje ograniczenia, które powodują, że jest trudna w implementacji w sieciach optycznych. Stąd też wskazane byłoby odniesienie w pracy do rozwiązań alternatywnych.
2. Przy definicji modelu sieci zalecana jest reprezentacja liczby węzłów i łączy za pomocą symboli pisanych dużymi literami, natomiast stosowanie małych liter (takich jak „k”, „v”) np. przy indeksacji.
3. Na str. 52 dwukrotnie pojawia się oznaczenie „TS” wśród „tested algorithms”.
4. Na str. 86, stwierdzenie „TLPQ values returned by algorithms are inversely related with traffic MAPE values” precyzyjniej oddaje związek pomiędzy wspomnianymi metrykami niż użyte obecnie „TLPQ values returned by algorithms are correlated with traffic MAPE value”.
5. W pracy można znaleźć drobne literówki i błędy gramatyczne (np. „Figures Figure 15 – Figure 17” -> „Figures 15 – 17”, „(5,)” -> „(5)”, „in this subsection” -> „in this section”, „algorithm get” -> „algorithm gets”, „base on” -> „based on”).

8. Konkluzja

Mgr inż. Daniel Szostak zaprezentował oryginalne rozwiązanie sformułowanego problemu naukowego, wykazując jednocześnie umiejętność samodzielnej realizacji badań naukowych, w tym wykorzystania właściwych metod badawczych oraz poprawnej analizy uzyskanych wyników. Wymienione powyżej sugestie i uwagi krytyczne, głównie związane z prezentacją rozprawy, nie zmieniają mojej pozytywnej opinii dotyczącej osiągnięć pracy, w tym wykazania postawionej tezy badawczej. Stąd też, w świetle przeprowadzonej analizy i oceny rozprawy doktorskiej mgr. inż. Daniela Szostaka, stwierdzam, że przedłożona mi do recenzji praca doktorska spełnia ustawowe kryteria wymagane dla uzyskania stopnia naukowego doktora, zgodnie z ustawą Prawo o szkolnictwie wyższym i nauce, i wnoszę o dopuszczenie jej do publicznej obrony.

