

Warszawa, 20 marca 2020 r.

Dr hab. inż. Jacek Malinowski
Instytut Badań Systemowych
Polskiej Akademii Nauk
01-447 Warszawa, ul. Newelska 6

Recenzja rozprawy doktorskiej mgr. inż. Artura Maździarza pt. „Wykorzystanie ciągów binarnych w metodach korelacji alarmów w mobilnych sieciach telekomunikacyjnych”.

Recenzowana praca dotyczy analizy ciągów binarnych reprezentujących sekwencje alarmów generowanych przez zdarzenia niepożądane (awarie) występujące w sieciach telefonii komórkowej. Doktorant wykorzystuje w niej współczynniki podobieństwa Dice'a, oraz metodę k-średnich stosowaną w analizie skupień, do określania związków przyczynowo-skutkowych między zdarzeniami generującymi sygnały diagnostyczne (alarmy) w systemach telekomunikacyjnych. Rozprawa składa się z siedmiu rozdziałów. Pierwszy z nich to wstęp, w którym opisano pokrótce jej treść, określono cel pracy badawczej, oraz sformułowano tezę. W tej części moją uwagę zwróciło następujące zdanie: „Korelacja alarmów jest procesem grupowania alarmów, które odnoszą się do tego samego problemu w celu wyodrębnienia alarmów będących przyczyną źródłową awarii”. Otóż alarmy nie są przyczyną awarii, ale jej symptomami.

Drugi rozdział zawiera wprowadzenie w tematykę rozprawy. W trzech pierwszych podrozdziałach (2.1, 2.2 i 2.3) autor przedstawił w skondensowanej formie zasady budowy i działania systemów telefonii komórkowej, ze szczególnym uwzględnieniem ich diagnostyki. W dalszych dwóch podrozdziałach (2.4 i 2.5) zawarł teoretyczne podstawy analizy wzajemnych relacji zdarzeń skutkujących alarmami. Cały rozdz. 2 bazuje głównie na cytatach z literatury źródłowej, ale mimo to zawiera szereg nasuwających wątpliwości sformułowań, których przykłady podaję poniżej.

Str. 19: Ostatni akapit w niepełny i niejasny sposób tłumaczy podział alarmów na aktywne i skasowane, oraz różnicę między systemami w pełni stanowymi i bezstanowymi.

Str. 22: W przedostatnim akapicie jest zdanie: „W większości przypadków uszkodzony element sieci wykrywa awarię i wysyła alarm do systemu zarządzania siecią”. Autor powinien tu podać przykład urządzenia wykrywającego własną awarię i wysyłającego alarm, gdyż wiele urządzeń traci swą funkcjonalność w przypadku awarii.

Str. 24: Drugi akapit zaczyna się od zdania: „Topologia sieci to dziedzina technologii ...”. Otóż topologia jest strukturą połączeń w sieci lub zależności funkcjonalnych między jej komponentami, ale nie dziedziną technologii.

Str. 27: W drugim akapicie jest zdanie: „Skierowane grafy definiują strukturę sieci bayesowskiej, podczas gdy nieskierowane grafy definiują strukturę pola losowego Markowa”, z którego wynika, że rola grafów ogranicza się do wymienionych dwóch zastosowań. Oprócz tego, opis metod szacowania topologii jest tak lakoniczny, że nie sposób wywnioskować, jaki jest cel tych metod i na czym one polegają. Z kolei w definicji łańcucha topologii zabrakło jasnego określenia, czy jest on jednowymiarową podstrukturą topologii hierarchicznej (drzewiastej). Dopiero rys. 3.4.1 znajdujący się na str. 90 wyjaśnia tę wątpliwość.

Str. 33, 34, 39, 40: Błędne użycie kwantyfikatorów we wzorach (2.4.3), (2.4.5), (2.4.6), (2.4.15) i (2.4.17).

Str. 35: Niejasna definicja szkieletu relacji ścisłego porządku. Błędne użycie operatora implikacji we wzorze (2.4.9).

Str. 37: Wbrew temu, co pisze autor, w grafie z rys. 2.4.1 zdarzenie 2 jest zdarzeniem pośrednim, a nie jest takim zdarzenie 5.

Znaczna część podrozdziału 2.5 poświęcona jest sieciom Bayesa jako probabilistycznym modelom zależności między zdarzeniami. Zawiera ona czysto formalne, ogólnikowe i mało zrozumiałe omówienie teoretycznych podstaw konstrukcji i analizy tych sieci, bez podania choć jednego przykładu ilustrującego mało komunikatywny przekaz. Można odnieść wrażenie, że aparat sieci Bayesa będzie w istotny sposób wykorzystywany w dalszym ciągu pracy, ale, wbrew temu wrażeniu, nigdzie nie jest używany jawnie. Podsumowując cały drugi rozdział, można stwierdzić, że wprowadza on czytelnika w tematykę korelacji zdarzeń i generowanych przez nie alarmów na dość wysokim poziomie abstrakcji, bez odniesienia się do sytuacji spotykanych w praktyce.

Rozdział trzeci stanowi zasadniczą bazę teoretyczną pracy. Jest podzielony na cztery podrozdziały. W pierwszym z nich autor podaje definicje trzech współczynników (Dice, Dice1 i Dice2) określających podobieństwo (lub jego brak) ciągów binarnych, oraz zasady wnioskowania dotyczące korelacji alarmów reprezentowanych przez takie ciągi. Zabrakło tu określenia *explicite* sposobu reprezentowania sekwencji alarmów przez ciąg binarny – można się tylko domyślić, że jedynka albo zero na miejscu k oznaczają odpowiednio alarm albo jego brak chwili t_k , gdzie t_k jest k -tą dyskretną chwilą czasową. Nie wiadomo, czy t we wzorze 3.2.1 jest dyskretną czy ciągłą zmienną czasową i co oznacza, że $t \neq t_1, t_2, t_3$. Zasady pomiaru czasu i konwersji alarmów na ciąg binarny są wyjaśnione dopiero pod koniec drugiego podrozdziału na str. 80. W tymże podrozdziale jest przedstawiony sposób określania związków przyczynowo-skutkowych między alarmami (a właściwie zdarzeniami generującymi alarmy) za pomocą porównywania wartości współczynników Dice, Dice1 i Dice2. Są tu omówione dwa warianty tej metody – bez opóźnienia i z opóźnieniem (ten drugi jest nazywany przez autora „ruchomy Dice”). Pierwszy wariant jest stosowany w sytuacji, gdy zdarzenie skutkowe następuje natychmiast po przyczynowym, a drugi – gdy zdarzenie skutkowe jest opóźnione względem przyczynowego. Opóźnienie to jest wyrażane parametrem τ służącym do wyznaczania pomocniczych współczynników $Dice^{\sim}(\tau)$, $Dice1^{\sim}(\tau)$ i $Dice2^{\sim}(\tau)$, a następnie docelowych współczynników $Dice^{\sim}$, $Dice1^{\sim}$ i $Dice2^{\sim}$ (wzory 3.2.2 - 3.2.4). Powyższe współczynniki kwantyfikują stopień i kierunek korelacji alarmów generowanych przez zdarzenia będące w związku przyczynowo-skutkowym. Dopuszczalny zbiór wartości parametru τ jest określony wzorem 3.2.5 i zależy istotnie od odległości Hamminga między odpowiednimi ciągami binarnymi. Należy tu zauważyć, że doktorant nie podał *explicite* definicji pomocniczych współczynników $Dice^{\sim}(\tau)$, $Dice1^{\sim}(\tau)$ i $Dice2^{\sim}(\tau)$ i pozostaje się domyślać, że są to współczynniki dla ciągu binarnego przesuniętego o τ bitów w lewo lub w prawo. Ponadto, kierunek przesunięcia jest określony w zawiły i nieporadny sposób (patrz drugi akapit na str. 72). Wystarczyło napisać, że τ jest liczbą bitów przesunięcia pierwszego z porównywanych ciągów, a znak τ określa kierunek przesunięcia. Przy tej okazji wypada nadmienić, że tego rodzaju nieporadności językowych jest w pracy niemało. Nie jest również określone, czy dopuszczalne są wartości τ skutkujące utratą jedynek w przesuwanym ciągu, oraz brakuje uzasadnienia dla arbitralnie przyjętej wartości 1,5 występującej we wzorach 3.2.2 – 3.2.4. W dalszym ciągu rozdz. 3.2 podane są reguły wyznaczania

siły i kierunku korelacji ciągów binarnych w oparciu o współczynniki Dice oraz przykłady ilustrujące zastosowanie tych reguł, a na jego końcu są zamieszczone schematy blokowe algorytmu wyznaczania współczynników Dice przy występowaniu opóźnienia i algorytmu ustalania źródłowych przyczyn alarmów (ang. root-cause analysis) wykorzystującego metodę Dice.

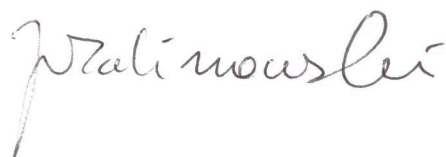
W kolejnych częściach rozdz. 3 (podrozdziały 3.3 i 3.4) autor przedstawia zarys metod analizy skupień, ze szczególnym uwzględnieniem metody k-średnich, a następnie opisuje jej zastosowanie do wykrywania skorelowanych sekwencji alarmów. Dla celów tego zastosowania doktorant usprawnił metodę k-średnich poprzez filtrowanie polegające na uwzględnianiu danych wejściowych pochodzących tylko z jednego łańcucha topologicznego, czyli poddrzewa hierarchicznej struktury reprezentującej zależności funkcjonalne między elementami sieci telekomunikacyjnej. Nie jest przy tym jasne, w jaki sposób wzory z rozdz. 3.4 adaptują się do ciągów binarnych, które, jak można domniemywać, pełnią rolę danych wejściowych dla używanej przez autora metody k-średnich.

Metodyka badania korelacji alarmów opisana w rozdz. 3, została zilustrowana w następnym, czwartym rozdziale trzema przykładami bazującymi na danych pochodzących z pewnej nienazwanej sieci telefonii komórkowej. W dwóch pierwszych przykładach zastosowano metodę „ruchomy Dice”, natomiast w trzecim – metodę analizy skupień w dwóch wariantach, czyli bez filtrowania topologicznego oraz z jego użyciem. Autor przedstawił szczegółowe wnioski z otrzymanych wyników pozwalające na zdiagnozowanie przyczyn awarii i nieprawidłowości występujących w analizowanej sieci.

W rozdziale 5 zaproponowane przez doktoranta rozwiązania, czyli „ruchomy Dice” i analiza skupień z filtrowaniem topologicznym, są porównywane z innymi, opisanymi w literaturze metodami analizy alarmów. Doktorant zwraca tu uwagę na brak wystarczających kryteriów porównawczych, co uzasadnia jakościowy charakter tych porównań. Rozdział 5 zawiera skrótkowe przedstawienie owych metod zakończone tabelą z wartościami sześciu wskaźników charakteryzujących cztery wybrane metody. Wprawdzie dwa z tych wskaźników (dokładność i czas przetwarzania danych) są typu ilościowego, ale zostały wyrażone za pomocą wartości lingwistycznych (wysoka, średnia, niska). Z owej tabeli wynika przewaga metody „ruchomy Dice” nad pozostałymi. Pracę kończą dwa krótkie rozdziały zawierające wnioski podsumowujące i

wskazania dotyczące tematyki przyszłych badań w dziedzinie diagnostyki sieci telekomunikacyjnych.

Podsumowując, stwierdzam, że doktorant przedstawił w swojej rozprawie interesujące i mające nowatorski charakter usprawnienia metod analizy danych diagnostycznych w systemach telekomunikacyjnych, oraz, pomimo przedstawionych tu uwag krytycznych, recenzowana praca spełnia w stopniu akceptowalnym wymagania stawiane rozprawie doktorskiej w dziedzinie nauk inżyniersko-technicznych w dyscyplinie informatyka techniczna i telekomunikacja. Wnoszę zatem o dopuszczenie mgr. inż. Artura Maździarza do kolejnych etapów przewodu doktorskiego.

A handwritten signature in black ink, reading "J. Palimowski". The signature is written in a cursive, flowing style with a long, sweeping underline.

