

Review of Bachelor's Thesis

Student: Šurina Timotej
Title: Anomaly Detection in Generated Incident Ticket Volumes (id 21169)
Reviewer: Rychlý Marek, RNDr., Ph.D., UIFS FIT VUT

- 1. Assignment complexity** **more demanding assignment**
Jedná se o obtížnější zadání - problematika detekce anomálií v časových řadách je složitá. Zadání je anglicky, přestože řešení je česky.
- 2. Completeness of assignment requirements** **assignment fulfilled with minor reservations**
Zadání je splněno s drobnými výhradami k věcné stránce, kdy dle zadání má být výstupem detekce anomálií v trendu, zatímco nástrojem generovaná zpráva o anomáliích (obr. 6.1 na str. 35) je v tomto ohledu moc stručná a obsahuje jediný výstup algoritmu a to je očekávaný počet lístků na daný den, nikoliv příznak, že se jedná o anomálii či popis a vztah k trendu (vzestup, pokles, odchylku, trvání atp.).
- 3. Length of technical report** **in usual extent**
Od úvodu po závěr obsahuje technická zpráva 34 vysázených stran a je tedy svým rozsahem v obvyklém rozmezí. Zpráva obsahuje hodně obrázků, z nich ne všechny jsou nezbytné (např. některé grafy) a bylo by vhodné je přesunout do příloh, nebo upravit (např. zbytečně velký obr. 6.1).
- 4. Presentation level of technical report** **60 p. (D)**
Technická zpráva má logickou strukturu sledující experimenty a vývoj realizačního výstupu. V technické zprávě je hodně grafů, což by mohlo být chválihodné, avšak tyto postrádají popisky (např. význam čar různých barev na obr. 5.1 až 5.8), mnohdy i význam (např. téměř shodné 5.7 a 5.8) a v nich zobrazené hodnoty a vyvozené závěry nejsou dostatečně odůvodněné (např. proč je z obr. 5.5 vidět, že je "aj keď" výsledok nie je najvhodnejší už v tejto podobe je podstatnelepší ako metóda ARIMA"; celkově bych doporučoval omezit subjektivní vyjádření typu "na prvý pohľad je jasné"). Kapitola 6.5 Testování mohla být obsáhlejší a otestovat úspěšnost nad více různorodými sadami dat (např. s různými problémy generujícími analyzované incidenty).
- 5. Formal aspects of technical report** **65 p. (D)**
Technická zpráva je psána ve slovenském jazyce, přestože zadání je v jazyce anglickém a ve zprávě se vyskytují často anglické názvy kapitol a počeštěné výrazy (např. "tikety"). Ve zprávě se nachází hodně typografických chyb (např. anglické uvozovky místo českých; nadbytečný spojovník v nadpisu kap. 3.4.2 a 3.4.3; spojovník místo pomlčky kap. 5.1 a 6.1; chybějící tečky za větami str. 19; chybějící mezery str. 33; atp.).
- 6. Literature usage** **85 p. (B)**
Seznam literatury obsahuje 14 položek a jsou zde v podstatné míře zastoupeny odborné publikace k tématu strojového učení a časových řad. Zdroje jsou v textu zprávy řádně citovány je patrný způsob a rozsah jejich použití.
- 7. Implementation results** **55 p. (E)**
Realizačním výstupem je fungující nástroj aplikující do práce převzatý algoritmus TES na dávku s časovou sérií lístků pro trénování a odhad vývoje jejich počtu v daném období. Nástroj je implementován v jednom Python skriptu o 580 řádcích, což není mnoho a je otázkou, jak pracné bylo takový skript napsat. Program není téměř vůbec strukturován (jen dvě pomocné funkce pro výstup, zbytek je přímý nestrukturovaný kód), takže samotný algoritmus je těžko k nalezení, ale je celkem srozumitelně komentován. V případě detekce anomálií (v testech) se používá (i v produkčním/netestovacím kódu) autorem navržená nekonfigurovatelná konstanta 15, která se přičítá ke všem odhadovaným počtům, přičemž ve výsledku je toto skryto (pojmenování "expected tickets" tedy není pravdivé). Realizační výstup mohl být jistě lépe navržen.
- 8. Utilizability of results**
Vzhledem k stručným výsledkům detekce (jen odhady počtu, žádná detekce anomálií, navíc s "magickou konstantou" 15), chybějícímu kontextu výsledných hodnot a velmi špatné kvalitě programového díla je, dle mého názoru, praktické využití značně omezené.
- 9. Questions for defence**
 - Jakým způsobem je patrný výskyt anomálie (detekce) z výstupu programu (počty) a jak toto usnadní odhalení důvodů anomálie (problému) uživateli aplikace?
 - Proč jste zvolil zrovna 15 jako konstantu, která se přičítá k odhadovanému počtu lístků?
- 10. Total assessment** **60 p. satisfactory (D)**
Technická zpráva i realizační výstup vykazují řadu nedostatků (zejména kvalita implementace). Zadání je však

obtížnější a bylo splněno s drobnou výhradou (chybějící detekce anomálií). Navrhuji hodnotit práci na spodní hranici stupně **uspokojivě (D)**.

In Brno 24. May 2019

Rychlý Marek, RNDr., Ph.D.
reviewer