

VILNIAUS UNIVERSITETAS
INFORMATIKOS INSTITUTAS
PROGRAMŲ SISTEMOS

Sustiprinto mokymosi taikymas žaidimo agento valdymo programos kūrimui

Application of reinforcement learning to the software development for game agent management

Bakalauro baigiamojo darbo planas

Atliko: Jokūbas Rusakevičius (parašas)

Darbo vadovas: <reikia parašyti pareigas> Virginijus Marcinkevičius (parašas)

Darbo recenzentas: (parašas)

Vilnius – 2020

Tyrimo objektas ir aktualumas

Darbo Tikslas, keliami uždaviniai ir laukiami rezultatai

Šio darbo **tikslas** –

Darbui išskirti **uždaviniai**:

1. Paruošti eksperimentinę aplinką ir įrašų rinkinį eksperimentui.
2. Aptikti ir analizuoti anomalijas įrašų rinkinyje naudojantis „MacroBase“.
3. Palyginti aptinkamąs anomalijas, keičiant įrašų rinkinio gerinimo kriterijus.
4. Pateikti rekomendacijas IPSS duomenų anomalijų aptikimui naudojantis „MacroBase“.

Darbo metu laukiami **rezultatai**:

1. Paruošta eksperimentinė aplinka ir paruoštas darbui tinkamas įrašų rinkinys.
2. Aptiktos anomalijos, naudojantis „MacroBase“.
3. Pakeisti duomenys pagal tam tikrus kriterijus ir gautos tikslesnės anomalijos.
4. Pateiktos rekomendacijos IPSS duomenų anomalijų aptikimui naudojantis „MacroBase“.

Tyrimo metodas

Darbui atlikti bus naudojami šie tyrimo metodai:

1. **Mokslinės literatūros analizė.**
2. **Eksperimentas.**

Kiekybinis metodas – aptikti pirmines anomalijas, jų priklausomybes nuo kitų atributų. Atliekamas, su visais turimais įrašų rinkiniais, ir ieškoma didžiausio anomalijų kiekio tarp turimų rinkinių.

Kokybinis metodas – su turima atrinkta (arba visa) įrašų rinkinio dalimi, atliekamos įvairios manipuliacijos ir stebimi pakitimai gaunamų anomalijų tikslume priklausomai nuo padarytų pakeitimų.

3. **Gautų duomenų ir rezultatų analizė.**

Numatomas darbo atlikimo procesas

Kaip jau minėta anksčiau, darbo eksperimentas bus atliekamas per dvi dalis, tačiau visas darbo procesas susidės iš daugiau dalių:

1. Visų pirma, eksperimentui atlikti bus paruošta eksperimentinė aplinka (tikėtina Ubuntu operacinė sistema įdiegta virtualioje mašinoje).
2. Duomenų paruošimas bus atliekamas, atrenkant svarbius tyrimui atributus ir padarant rinkinį pasiekiamą „MacroBase“ įrankiui.
3. Jei visas įrašų rinkinys bus per didelis analizei, bus ieškoma didžiausios anomalijų dalies bandymų keliu.
4. Su šiuo punktu, prasidės pagrindinė eksperimento dalis. Bus ieškoma ir eksperimentuojama su metodais, kurie pagerintų anksčiau gautus rezultatus.
5. Galutinė analizė visų atliktų tyrimų ir gautų rezultatų.

6. Pateikiama rekomendacija, kaip pagerinti anomalijų aptikimą IPSS įrašų rinkiniui naudojantis „MacroBase“ analitinį įrankį.

Darbui aktualūs literatūros šaltiniai

1. **[BGM⁺17]** – Standfordo Universiteto išleistas straipsnis apie jų vystomą „MacroBase“ analitinį įrankį bei jo apžvalgą. Šiame leidinyje yra pristatomas pats įrankis, paaiškinamas kiekvienas iš jo naudojamų darbo proceso aspektų bei principų. Tai svarbus šaltinis naudojantis „MacroBase“.
2. **[BGR⁺17]** – Standfordo Universiteto išleistas straipsnis. Šiame leidinyje yra rašoma daugiau apie pačius principus ir teoriją, o ne apie „MacroBase“ analitinį įrankį.
3. **[RB17]** – Standfordo Universiteto straipsnis apie laiko intervalu fiksuojamus duomenis bei jų išlyginimą, padaryma lengviau skaitomais. Straipsnyje tai vadinama ASAP, analitiniu operatoriumi, kuris automatiškai išlygina įrašų rinkinį, taip padidinant analizuojamų duomenų gautų rezultatų tikslumą, bei sumažinant skaičiavimams reikalingą greitį.

Literatūra

- [Ast16] Anthony Asta. Observability at twitter: technical overview, part i. 2016. URL: https://blog.twitter.com/engineering/en_us/a/2016/observability-at-twitter-technical-overview-part-i.html (tikrinta 2019-03-10).
- [BGM⁺17] Peter Bailis, Edward Gan, Samuel Madden, Deepak Narayanan, Kexin Rong ir Sahaana Suri. Macrobaze: prioritizing attention in fast data. *SIGMOD'17- Proceedings of the 2017 ACM International Conference on Management of Data*, p. 541–556, Chicago, Illinois, USA. Stanford Infolab ir Massachusetts Institute of Technology, ACM New York, 2017. ISBN: 978-1-4503-4197-4.
- [BGR⁺17] Peter Bailis, Edward Gan, Kexin Rong ir Sahaana Suri. Prioritizing attention in fast data: principles and promise. *CIDR 2017, 8th Biennial Conference on Innovative Data Systems Research*. Stanford Infolab, 2017.
- [HA17] Makrufa Sh. Hajirahimova ir Aybeniz S. Aliyeva. About big data measurement methodologies and indicators. *International Journal of Modern Education and Computer Science*, 9:1–9, 2017.
- [PFT⁺15] Tuomas Pelkonen, Scott Franklin, Justin Teller, Paul Cavallaro, Qi Huang, Justin Meza ir Kaushik Veeraraghavan. Gorilla: a fast, scalable, in-memory time series database. *VLDB Endowment - Proceedings of the 41st International Conference on Very Large Data Bases*, p. 1816–1827, Kohala Coast, Hawaii. Facebook Inc., 2015. (Tikrinta 2019-03-10).
- [RB17] Kexin Rong ir Peter Bailis. Asap: prioritizing attention via time series smoothing. *Proc. VLDB Endow.*, 10(11):1358–1369, 2017-08. ISSN: 2150-8097. DOI: 10.14778/3137628.3137645. URL: <https://doi.org/10.14778/3137628.3137645>.
- [rrtar18] Lietuvos Respublikos ryšių reguliavimo tarnyba. Interneto prieigos stebėsenos sistema ipss. 2018. URL: <https://opendata.rrt.lt/ipss> (tikrinta 2019-03-08).
- [Rus18] Jokūbas Rusakevičius. *Didelių duomenų srautų analizė, anomalijų aptikimas*. Kursinis darbas, Vilniaus Universitetas, Matematikos ir Informatikos Fakultetas, Vilnius, 2018.
- [Tec18] Techopedia. Anomaly detection. 2018. URL: <https://www.techopedia.com/definition/30297/anomaly-detection> (tikrinta 2019-03-07).
- [Woo15] Alex Woodie. Kafka tops 1 trillion messages per day at linkedin. 2015. URL: <https://www.datanami.com/2015/09/02/kafka-tops-1-trillion-messages-per-day-at-linkedin/> (tikrinta 2019-03-10).