

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Tomáš Drozdík

Název práce Asynchronous Duet Benchmarking

Rok odevzdání 2023

Studijní program Informatika **Studijní obor** Softwarové systémy

Autor posudku Vojtěch Horký **Role** vedoucí

Pracoviště Katedra distribuovaných a spolehlivých systémů

Text posudku:

Cílem práce je rozšíření metody tzv. duetového měření (benchmarkování) tak, aby byla použitelná pro libovolný benchmark bez větších zásahů do jeho kódu. Změny se týkají mechanismu spouštění a sběru dat (implementační část) tak i způsobu vyhodnocování takto sesbíraných dat (teoretičtější část práce).

Duetové benchmarkování je určeno pro detekci regresí ve výkonu softwaru, které se snaží odstranit zásadní problém obvyklého porovnávání: měřené verze běží v různý čas a jsou vystaveny různým vlivům ovlivňující jejich výkon (toto je především patrné v nestabilních prostředích jako je např. cloud). V rámci duetu proto běží obě verze současně, takže je rozumný předpoklad, že vnější vlivy (např. zátěž od ostatních procesů) budou působit na obě verze podobně a relativním porovnáním obou verzí můžeme identifikovat případné regrese.

Původní verze duetového benchmarkování vyžadovala, aby jednotlivé iterace benchmarku byly striktně synchronizované, tj. obě verze měly společné (časové) startovní body (a tedy stačilo doby běhu odpovídajících iterací pouze přímo porovnat). Rozšíření v této práci se zaměřuje na zobecnění této metody, kdy není nutné běhy jakkoliv synchronizovat, jedinou podmínkou je ukládat absolutní časové značky aby bylo možné zajistit pseudo-spárování dat při vyhodnocování.

Práce popisuje obecné možnosti benchmarkování a současný stav implementace duetového měření a motivaci pro zobecnění. Následuje popis implementačních rozhodnutí, přehled zpracování výsledků a rozbor experimentálně naměřených dat. Práce uzavírá výhledem na možná rozšíření implementace. Vlastní implementace je k dispozici i jako repozitář na GitHubu.

Práce je přehledně strukturována, vlastní text obsahově poměrně hutný, ale přesto vcelku dobře čitelný.

Na implementaci je nutné zvláště ocenit velkou snahu o univerzálnost, kdy autor využívá moderní technologie kontejnerů jak pro zvýšení přenositelnosti tak i pro zajištění co nejjednodušších

podmínek měření. Distribuce implementace jako Pythoního balíčku s repozitářem na GitHubu je velmi příjemná na použití.

Text práce je přehledně členěn, po nezbytném úvodu autor pevně usadil téma práce do hlavních otázek, na které hledá odpovědi. Velmi se mi líbí, že motivace pro jednotlivá podtémata je ilustrována praktickými příklady, které motivovaly autora k dalšímu zkoumání.

Text kombinuje detailní informace o softwarových technologiích tak i teoretičtější části o statistických testech, v části věnované matematice je bohužel čtenář „zavalen“ poměrně komplikovaným značením. To je sice dané šíří výsledků, nutí ale čtenáře si pamatovat relativně komplikované vztahy, které ale v důsledku jen popisují to, že porovnáváme porovnatelné.

Oceňuji, že v části o vyhodnocení se autor vždy snažil poskytnout i praktický „návod“, jak s daty pracovat, tj. kolik dat je *obvykle* nutné „zahodit“ apod.

Vlastní implementace se skládá ze dvou částí. Je to „spouštěč“ benchmarků napsaný v Pythonu, který používá Docker (Podman) a pak Pythoní skripty pro zpracování výsledků.

Kód pro ovládání kontejnerů je poměrně přímočarý, spuštění vyžaduje netriviální konfiguraci (v porovnání s podobnými nástroji je ale relativně jednoduchá).

Zpracování výsledků je pak založené na Jupyteru, který je standardní volbou pro tento typ analýzy.

Zvláště oceňuji širí sbíraných dat, kdy autor získal data z různých platforem a po celou práci se snažil poskytnout co nejvěrohodnější obraz variability dat, se kterými je možné se v praxi setkat.

Práce se zaměřuje výhradně na porovnávání stejných verzí, takže chybí praktický pohled na sílu metody pro detekci skutečných regresí (nutno dodat, že je velmi složité získat informace o potvrzené regresi, která je důsledkem přirozeného vývoje a nikoliv uměle zavlečeným problémem).

To je částečně kompenzováno pomocí MDS, kdy autor ukazuje předpokládané hranice použitelnosti dané metody.

Celkově je práce na velmi dobré úrovni. Autor prokázal znalosti použitých technologií a především velmi dobře zhodnotil praktičnost zjištěných výsledků a použitelnost celé metody.

Práci proto doporučuji k obhajobě.

V Praze dne 29. ledna 2023

Podpis: