

# Posudek oponenta bakalářské práce

Autor/autorka práce: Jindřich Pouba

Název práce: **Systém pro paralelní spuštění obecného genetického algoritmu**

## Obsah práce

Cílem práce je vytvoření systému pro řízení distribuovaného výpočtu genetických algoritmů. Teoretická část rozebírá problematiku genetických algoritmů, s důrazem na různé operátory výběru, křížení a mutace. Stručně popisuje problém paralelizace a následně návrh a implementaci aplikace.

V textu práce je několik problematických míst. Popis souběhu (3.2.1) ve skutečnosti nevysvětluje princip souběhové chyby, ale jen normální chování paralelně prováděného programu. Popis gaussovské mutace (s. 37) je nejasný. Tabulka 8.1 obsahuje řádek se součty hodnot, které nejsou k ničemu užitečné (sčítají se pravděpodobně počty provedených výpočtů fitness funkce na dvou různých počítačích ve dvou různých testech).

## Kvalita řešení a dosažených výsledků

Samotné programové řešení vypadá funkčně, a nakolik jsem byl schopen ho otestovat je jednoduše použitelné. Program nabízí řadu operátorů pro práci s populací a umožňuje implementaci vlastní fitness funkce, kterou automaticky distribuuje na jednotlivé výpočetní uzly.

Při testování jsem narazil na několik problémů. Po opakovaném spuštění algoritmu se GUI nerestartuje, hodnoty nejvyšší dosažené fitness funkce a počtu výměn zůstávají nezměněny. Nevím, jestli doopravdy dochází k opětovnému spuštění, protože i při změně fitness funkce se u výpisu „Fitness name“ objevuje původní hodnota, algoritmus proběhne za mnohem kratší dobu a ve výpisu výsledků zůstávají původní výsledky. Nepodařilo sem i program opětovně spustit bez restartu všech procesů a nového nastavení celého procesu, pak ale funguje normálně. Po nastavení algoritmu není možné se k němu vrátit a ověřit si jaké bylo, je možné jen nastavit nový algoritmus (dialog pro nastavení je vždy nastaven na výchozí hodnoty). U jednotlivých workerů není nijak aktualizován jejich pokrok – „progress“ obsahuje vždy hodnotu 0 nebo 100. Není mi moc jasné, co obsahuje pole „number of exchanges“. Při nastavení ukončovací podmínky na 60 s a výměny jednou za 10 s napočte správně 6 výměn. Při nastavení ukončovací podmínky na 100 generací a výměny jednou za 10 generací program napočte do hodnoty 101. Přes značnou uživatelskou nepříjemnost je program použitelný pro spuštění distribuovaného genetického výpočtu, byla by ale potřeba řada úprav a dodělávání, aby jeho použití nebylo oproti ručnímu spuštění spíš na překážku. Drobným opomenutím je nepřiložení souborů s ikonami na CD s prací.

Nejvýraznější výhrady mám k testování (kapitola 8). Zvolená metodika mi připadá nevhodná – počet jedinců v populaci je velmi malý (2), ukončovací podmínka je nastavena na dobu běhu pro měření výkonu program počítá počet výpočtů fitness funkce za tuto dobu. Namísto využití čítače je počet běhů zaznamenáván přes výpis znaku „.“ do konzole a počítání jeho výskytu, což považuji za zbytečně komplikované a neefektivní řešení. Fitness funkce použitá při testech ve skutečnosti nic nepočítá, vrací konstantní hodnotu, ve skutečnosti výpočet genetického algoritmu neprobíhá. Měření probíhalo pouze na jednom počítači, jednou se dvěma a jednou se čtyřmi jádry, tabulky v příloze E jsou ale vytvořené až pro 8 procesů – v okamžiku kdy je procesů víc než jader, nemá měření výkonu

valný smysl. Z testování mám dojem, že bylo prováděno ve velké časové tísní a bez patřičné pozornosti, která mu měla být věnována.

### Formální úroveň

Text práce je dobře strukturovaný a srozumitelně popisuje jak teoretickou problematiku, tak návrh aplikace. I zdrojové texty působí přehledně, dekompozice programu je provedena dobře a umožňuje snadné doplnění vlastní fitness funkce. Překvapila mě prakticky naprostá absence komentářů, alespoň na úrovni základních funkcí jednotlivých tříd.

### Práce s literaturou

Veškerá citovaná literatura je relevantní, zejména shromáždění informací o genetických algoritmech bylo věnováno velké úsilí.

### Splnění zadání

První tři body zadání považuji za splněné, i když použitelnosti grafického rozhraní mohla být věnována větší péče. Čtvrtý bod je splněn, ale metodika je nevhodná a data nejsou příliš vypovídající. Naměřené údaje ukazují na superlineární urychlení distribuovaného výpočtu, které autor práce na s. 21 sám označuje jako příznak možného problému s implementací paralelního algoritmu. Důvody dosažení takového urychlení ale v práci diskutovány nejsou.

### Dotazy k práci

K práci mám následující dotazy:

Podle čeho jste navrhl metodiku pro testování v kapitole 8?

Proč jste se rozhodl umožnit snadné doplnění fitness funkce, ale ne operátorů křížení a mutace? Jak náročné by bylo program rozšířit tak, aby to umožnil?

Jaká topologie propojení jednotlivých uzlů je použita?

Navrhuji hodnocení známkou **dobře** a práci doporučuji k obhajobě.

V Plzni 26.5.2015

Ing. Richard Lipka, Ph.D.

