

## Posudek oponenta diplomové práce

Diplomant : Bc. Jan Hanousek  
Název práce : Aproximace histogramu na kruhu  
Vedoucí práce: Mgr. Jaroslav Marek, Ph.D.

Diplomant měl vytvořit aplikaci, která vizualizuje soubor dat s periodickým rozdělením jednak ve formě histogramu a jednak aproximací spojitého rozdělení. Jako zdrojová data byly použity informace o výjezdech zdravotnické záchranné služby za rok 2009 uložené ve formátu Excel. Aplikace má pro zadaný časový úsek určit a zobrazit rozložení četnosti výjezdů v průběhu 24 hodin.

Pro získání průběhu aproximace hustoty pravděpodobnosti byly využity dva přístupy. První přístup je založen na neparametrické metodě jádrových odhadů, kdy se počítají přímo odhady bodů aproximace. Druhý přístup využívá von Misesovo rozdělení závislé na třech parametrech a výsledné rozdělení daného souboru dat aproximuje směsí (součtem) tří nezávislých rozdělení. Devět neznámých parametrů se určuje minimalizací účelové funkce – Pearsonovy  $\chi^2$  statistiky. Pro vyhledání minima jsou použity dvě metody – algoritmus E-M a genetický algoritmus.

Diplomant vytvořil a realizoval ve formě programu v Javě algoritmy pro výpočet bodů aproximace spojitého rozdělení dvěma různými způsoby, přičemž způsob využívající von Misesovo rozdělení vyžaduje i realizaci E-M algoritmu a genetického algoritmu hledání hodnot parametrů minimalizujících účelovou funkci. Dále vytvořil grafickou aplikaci umožňující volbu rozsahu dat použitých pro vykreslení histogramu a aproximace spojitého rozdělení včetně volby způsobu výpočtu aproximace. V programu jsou využity dvě externí knihovny funkcí – jedna pro načítání dat ze souboru ve formátu Excel a druhá matematická obsahující při výpočtech potřebné modifikované Besselovy funkce prvního druhu.

Funkčnost aplikace diplomant prokazuje vizuální shodou histogramu se spojitou aproximací rozdělení vypočítanou jako směs tří von Misesových rozdělení s parametry určenými E-M algoritmem při 1000 iteracích. Výsledky – četnost výjezdů v průběhu jednoho dne – jsou komentovány pro data zpracovaná jednak po měsících a jednak po dnech v týdnu.

Teoretické základy, ze kterých práce vychází, nejsou triviální a jejich uvedení by určitě přesahovalo rámec diplomové práce. Diplomant se proto rozumně soustředil pouze na popis použitých metod z pohledu jejich praktického použití a tato část (Kapitola 1) tvoří teoretickou část práce. Nicméně popis a formulace uvedené v DP nejsou moc jasné a připouští různý výklad. Teprve po opakovaném čtení lze odhadnout, jak to autor asi myslel. Kapitola 2 obsahuje popis Integrovaného záchranného systému a je z pohledu DP nevýznamná. Kapitola 3 je převážně komentovaná interpretace rozložení četnosti výjezdů během dne v různých obdobích. Užitečná je pouze Kapitola 3.4, kde jsou porovnány na základě dosažené hodnoty  $\chi^2$  statistiky jednotlivé způsoby výpočtu z pohledu rychlosti a kvality aproximace měření. Kapitola 4 se pak zabývá popisem vytvořené aplikace z hlediska používání. Informace o programovém řešení jsou uvedeny v příloze ve formě UML diagramu a informace o použitých externích knihovnách a použitém vývojovém prostředí je v Kapitole Závěr. Součástí DP je i CD s funkční aplikací a zdrojovými soubory.

V DP jsem si všiml několika nejasností, které by mohl diplomant stručně komentovat:

- a) pro jakou kombinaci znamének čitatele a jmenovatele platí prostřední výpočet v rovnicích (2.6), (2.18) a (2.24)
- b) co chtěl diplomant vyjádřit na str. 14 vzájemně si oporujícími větami „Křivku ... není možné integrovat. K výpočtu integrálů byla použita Simpsonova metoda“?
- c) proč diplomant používá problematický slovní popis algoritmu, když existuje standardní způsob (strukturogram, vývojový diagram). Např. v popisu algoritmu na str. 16 je nejasný význam symbolu  $\theta$  (pole o 100 položkách) v kroku 5, který je důsledkem kombinace popisu jednoho cyklu indexem a druhého cyklu slovně „opakování pro všechny  $\theta$ “
- d) pokud označení  $^{-1}$  ( $\tan^{-1}$ ) znamená inverzní funkci – rovnice (2.6), (2.18) a (2.24) – jaký význam má totéž označení  $2(n_{\text{KML}})^{-1}$  v rovnici (2.12). Proč diplomant nepoužil místo  $\tan^{-1}$  jednoznačné označení např.  $\text{atan}$ ?
- e) orientaci v popisech též komplikuje použití symbolu  $\pi$  jak ve významu Ludolfova čísla, tak ve významu jednoho z parametrů von Misesova rozdělení.
- f) proč diplomant vytváří vlastní způsob bibliografické citace a nerespektuje normu ČSN ISO 690 (01 0197) dle Směrnice UPa č. 9/2012 „Pravidla pro zveřejňování závěrečných prací a jejich základní jednotnou formální úpravu“? Obdobně proč jsou číslovány rovnice až na následujícím řádku a nikoliv na řádku s rovnicí?

Navrhuji, aby při obhajobě diplomant zodpověděl následující otázky:

- a) **jaký jiný způsob (než po zadaném počtu iterací) se používá pro ukončení vyhledávání extrému funkce**
- b) **jak si vysvětluje fakt, že při použití E-M algoritmu se dosahuje významně jiných hodnot účelové funkce ( $\chi^2$  statistiky) než při použití genetického algoritmu**

Přes výše uvedené ponejvíce formální připomínky navrhuji hodnocení

**== výborně-m ==**

v Pardubicích 25. 5. 2015

.....  
doc. Ing. František Dušek, CSc