

## POSUDEK OPONENTA NA DIPLOMOVOU PRÁCI

### MARTINA KŮSOVÁ: MODELLING AND PREDICTION OF DATA IN LIMIT ORDER BOOKS

Autorka programuje sestavení časového průběhu stavů knihy objednávek z dílčích zpráv EOBI, na reálných datech (futures na FGDL z 2.12.2019) bojuje s volbou stochastického modelu, v závěru přebírá predikční model z [4].

#### Splnění cílů práce

- nadstandardně
- velmi dobře
- splněny
- s výhradami
- nebyly splněny

#### Odborný přínos práce

- nové výsledky
- netradiční postupy
- zpracování výsledků z různých zdrojů
- shrnutí výsledků z různých zdrojů
- bez přínosu

#### Matematická (odborná) úroveň

- vynikající
- velmi dobrá
- průměrná
- podprůměrná
- nevyhovující

#### Věcné chyby

- téměř žádné
- vzhledem k rozsahu přiměřený počet
- méně podstatné, větší množství
- podstatnější, větší množství
- závažné

#### Grafická, jazyková a formální úroveň

- vynikající
- velmi dobrá
- průměrná
- podprůměrná
- nevyhovující

Ač při prvním prolistování vypadá práce čtivá a budí dojem, že problematiku čtenáři dobře přiblíží, při bližším pohledu zjistíme, že řada věcí je popsána jen povrchně, neúplně, nedostatečně, že části přebírané z různých zdrojů nejsou provázané a že práce, bohužel, obsahuje i podstatné chyby.

- Při rekonstrukci knihy limitních objednávek postrádám popis metodiky (souvislost zpráv knihovního systému s algoritmem konstrukce knihy z 3.1) a soudě podle přiloženého programového kódu se zdá, že rekonstrukce knihy je pak naprogramována *chybně* (už jen to, že při zprávě typu OrderModifySamePrio pravděpodobně opomíná upravit údaje dotčené objednávky v knize). *Tím je správná rekonstrukce zhacena* a odvozené veličiny, na nichž je založeno mj. pozdější stochastické modelování, jsou tak spočteny špatně.
- V 4.4. se na datech ověřují předpoklady o rozdělení dob mezi příchody objednávek a o rozdělení doby jejich setrvání v knize objednávek, které se patrně týkají nějakého náhodného *procesu*, který však není v textu nikde popsán, natož vysvětlen. Nejprve je u obou veličin zkoumáno exponenciální rozdělení, jediným vodítkem je zde QQ-graf, způsob jeho zobrazení (str. 33) však neumožňuje posoudit tvar rozdělení. Je škoda, že se např. nedozvíme žádné popisné statistiky. Následuje test s mocninnými rozděleními, ale proč bylo zvoleno zrovna to (příhodnější rozdělení jsou přitom popsána v literatuře, z níž diplomantka

cituje, a potkala je i během studia)? Není uvedena metoda odhadu parametrů, prezentované *hodnoty odhadů* u mocninného rozdělení pokládám za *zcela nesmyslné*, čímž jsou výsledky dalšího testování irelevantní. Kromě toho je k testování použit Kolmogorovův-Smirnovův test místo Lillieforsova testu, přestože rozdíl v jejich použití autorka předtím v teoretické části popisuje.

- Předpoklady na datech autorka naopak neověřuje v 4.6 u modelu s homogenním markovským řetězcem, který popsán byl a který je v závěru použit ke konstrukci obchodní strategie (obojí převzato z [4]). Čekal bych např., že reálně frekvence příchodu objednávek přinejmenším závisí na denní době, což nabourává předpoklad o homogenitě. Model je použit v lehce *jiné podobě*, než ho prezentovaly vzorečky v *metodologické části* 3.2.2, i odhady intenzit jsou počítány jinak, než uvádí vzorec na str. 21. Způsob prezentace, včetně programových kódů a zobrazení výsledku, víceméně kopíruje [12], v zásadě jsou jen dosazena jiná data. V předloženém textu nejsou ani uvedeny odhadnuté hodnoty parametrů modelu (omezuje se jen na výsledné dílčí podmíněné pravděpodobnosti), výsledky zůstávají bez dalších komentářů či porovnání s výsledky zdrojové práce.

V textu je i spousta menších chyb. Od zmateného výkladu o cenách futures a dlouhopisů (který se v dalším nijak nepoužije), přes překlepy, opomenutí (namátkou: otočená nerovníka na str. 12, vypadlá  $t \pm$  před  $\Delta$  v intervalech na str. 20, chybějící předpoklad o spojitosti rozdělení u Kolmogorovova-Smirnovova testu, nevyslovený předpoklad homogenity na str. 19), nešikovné formulace, špatně zvolené typy grafu (např. str. 31), až po úsměvná vysvětlování (poslední odstaveček na str. 30 či 42). Nehezky vypadá, že za rovnicemi je ukončován odstavec (i uprostřed věty). Nelíbí se mi také práce s citacemi — většinou nelze rozeznat, kdy citační značka maskuje „slepení textu“ z úryvků zdrojové literatury (pár slov, věta, odstavec?) a kdy naopak regulérně odkazuje na nějaký podpůrný argument či podrobnější výklad.

Úlohu najít vhodný stochastický model pro budoucí stav knihy limitních objednávek (a cenu aktiva) považuji za obtížnou, proto mi až tak nevádí, že některé cíle práce stanovené v zadání nebyly úplně dosaženy, tím spíše, že pro získání dat bylo nejprve potřeba se zorientovat v mechanismu obchodování a struktuře zpráv. Co mi vadí a co mě vede k výslednému hodnocení, je jednak neúplný popis, jednak popsání chyby v klíčových bodech práce a jednak malá přidaná hodnota u markovského modelu.

Práci hodnotím známkou *dobře*.

Otázky do rozpravy

1. Proč se v programovém kódu při rekonstrukci knihy objednávek např. při zprávě `OrderAdd` řeší případ, kdy by limitní cena vkládané kupní objednávky nebyla menší než nejlepší prodejní cena dostupná v knize?
2. Popište srozumitelným způsobem a kompletně (ne jen dvěma příklady jako na str. 41) výslednou obchodní strategii (obr. 4.23 na str. 42). Šlo by i přímo pravděpodobnosti z tab. 4.7 prezentovat v přehlednější struktuře, resp. vyslovit k dosaženým hodnotám v tabulce nějaká další komentáře?

MICHAL FRIEŠL

Plzeň, 12. června 2023.