

Posudek oponenta bakalářské práce

Student: Gráca Martin

Téma: Využití veřejných obchodních informací pro automatický trading (id 19001)

Oponent: Pichot Oldřich, Ing., Ph.D., UPGM FIT VUT

1. Náročnost zadání **průměrně obtížné zadání**

Zadání práce požaduje po studentovi zvládnutí základních technik strojového učení a záleží na studentovi, jak se k řešení zadaného problému postaví. Zadaný problém byl již mnohokrát řešen za použití různých přístupů a jedná se tak o aplikování dalších známých technik strojového učení na dobře prozkoumaný problém.

2. Splnění požadavků zadání **zadání splněno s vážnými výhradami**

V textu práce mi chybí podrobnější diskuze, kterou vyžaduje bod 6 zadání. Student sice navrhne automatický obchodní systém, který generuje příkazy, nicméně se zde nediskutuje propojení na brokera. Možná úskalí která plynou z obchodování na reálném trhu za aktuální ceny, které se liší od závěrečných cen použitých pro trénování, by měly být diskutovány při návrhu obchodního systému.

3. Rozsah technické zprávy **je v obvyklém rozmezí**

Text práce je obvyklého rozsahu.

4. Prezentací úroveň předložené práce **60 b. (D)**

Kapitoly práce na sebe logicky navazují a jejich obsah je pro znalého čtenáře pochopitelný. Obecně by bylo vhodné citovat příslušnou literaturu u pojmů, které nemusí být čtenáři jasné z jejich krátkého nebo žádného popisu (futures, opce, CFD, RSI, Bollinger Band, MACD).

V úvodní kapitole, či v navazující kapitole mi chybí podrobnější rešerše literatury a přehled systémů, které se zabývají stejným nebo podobným problémem. Kapitoly 5 a 6 by mohly být sloučeny do jedné, jelikož jsou obě dvě velice krátké a popis dat a jejich zpracování logicky patří do návrhu systému.

Další poznámky

- kap. 2.2.1: Pojem svíčka je zde použit aniž by bylo definováno co to je svíčkový graf.
- kap. 2.2.3: Stochastic -> mělo by být zřejmě stochastic oscillator
- kap. 2.2.3: Bolling Band -> mělo by být Bollinger Band
- kap 3. Pokud se vysvětluje pojem učení s učitelem, mohl by být stručně vysvětlen i pojem učení bez učitele.
- Nenašel jsem v textu odkaz na celostránkový obrázek 6.1.
- str. 13 chyba v citaci [?]

5. Formální úprava technické zprávy **75 b. (C)**

Po jazykové a typografické stránce je dokument více méně v pořádku. Text obsahuje několik překlepů a gramatických chyb. Například str. 14,19,21,23 : nepřenosti, časovým rámce, zakonponovat, historickách datech

6. Práce s literaturou **55 b. (E)**

Jak již bylo uvedeno výše, autor měl provést rešerši daného problému a uvést odkazy na příslušnou literaturu. Jelikož práce používá řadu ekonomických pojmů, tak by měl autor uvést odkazy do relevantní ekonomické literatury. Práce obsahuje naprostou většinu citací do internetových zdrojů.

7. Realizační výstup **40 b. (F)**

Po kontrole zdrojových souborů bylo odhaleno několik problémů. Zřejmě chybou z nepozornosti jsou pro trénování modelů použita data, která se budou překrývat s testovacími daty. (trénování: řádek 101 skriptu `rnn_with_sec.py`, testování ve skriptu `mg_trade_strategy.py`, funkce `get_sec_data()` se používá pro úplně stejný soubor pro výběr testovacího vzorku dat. Tento fakt vede k tomu, že dosažené výsledky a závěry práce se budou muset opravit. Současné výsledky nemohou být považovány za relevantní.

Dále by bylo dobré zřetelně označit kód, který psal sám autor od kódu převzatého z internetu z různých příkladů k použitým knihovám. Jedná se například o tyto části:

části kódu v `parse_SEC_fillings.py`

<https://github.com/greedo/python-xbrl/blob/master/examples/gaap.py>

<http://stackoverflow.com/questions/12412994/use-lxml-to-parse-text-file-with-bad-header-in-python/12431072>

části kódu v `rnn_with_sec.py`

<https://gist.github.com/lukovkin/9ce8db594a9dd7d2a2ac>

8. Využitelnost výsledků

Výsledky a závěry práce musí být přepracovány po opravení problému s trénovacími a testovacími daty. Softwarový balík, který zpracovává data a trénuje neuronové sítě je dobře využitelný pro další výzkum, jelikož se jedná a celkem elegantní a lehce pochopitelný kód.

9. Otázky k obhajobě

- Při obhajobě prezentujte opravené výsledky
- V textu se kromě obrázku 7.1 vůbec nezmiňujete o použití dropouts jako regularizační techniky. Okomentujte důvody proč zde byla tato technika použita a pokuste se prezentovat jaký vliv má na trénování a na výslednou úspěšnost systému.
- Z kódu je zřejmé, že mícháte data pro trénování rekurentních modelů. Jelikož tyto modely pracují s delší historií a tento přístup porušuje kontinuitu dat, tak by bylo vhodné analyzovat chování systému i bez zamíchání trénovacích dat. Pokuste se analyzovat chování systému s a bez zamíchání trénovacích dat.
- Při návrhu systému neuvažujete, že obchodní systém bude muset nakupovat za reálné intradenní ceny. Ty mohou být zatíženy značnou volatilitou a spreadem. Pokuste se zanalyzovat, nejlépe na vzorku intradenních dat, jaký vliv by mělo výše uvedené na ceny použitých titulů a výkonnost obchodního systému.

10. Souhrnné hodnocení

55 b. dostatečně (E)

Celkové hodnocení ovlivnily zejména chyby vyskytující se v programové realizaci a neúplný popis systému, který přesně neodpovídá zdrojovému kódu. Prezentační úroveň a práce s literaturou jsou spíše podprůměrné. Celkové hodnocení E je možno přehodnotit na lepší známku při obhajobě, pokud student předvede vynikající prezentaci opravených výsledků a prokáže dobré znalosti a porozumění danému problému.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 1. června 2016

.....
podpis