



Posudek oponenta bakalářské práce

Student: Ondřej Šrámek

Název práce: Pokročilé architektury neuronových sítí pro analýzu dat z finančních trhů

Předkládaná práce zkoumá možnosti využití neuronových sítí pro predikci vývoje časových řad z finančních trhů. Text práce je rozdělen do tří kapitol. První obsahuje úvod do strojového učení, popis neuronových sítí a učícího algoritmu. Druhá kapitola se věnuje problému přetrénování, a jak mu zamezit pomocí regularizace nebo jiných technik. Samotná aplikace na časové řady je popsána ve třetí kapitole.

Autor vhodně použil rekurentní neuronovou síť neboť tento typ sítí je dobře vhodný právě pro zpracování časových řad nebo textu. Navrhl postup pro předzpracování dat, který by měl urychlit konvergenci učící fáze. Výstupem sítě je jeden ze tří možných vývoju časové řady a sice růst, pokles a stagnace. Z výsledků vyplývá, že neuronová síť výrazně preferuje stagnaci před růstem a poklesem. Skoro se mi zdá, že jde o možné zkreslení (bias) modelu. Napadá mě, jestli to není způsobeno standardizací dat, kterou autor používá. Nicméně je jasné, že predikce finančních řad je velice náročná úloha a těžko zde lze očekávat i jen nadprůměrnou přesnost odhadů.

Téma práce chápu stejně spíše jako zajímavou úlohu pro seznámení se s neuronovými sítěmi a základy strojového učení. V tomto směru bych řekl, že student mohl jít o něco více do hloubky. Text se mi zdá být místy trochu povrchní. Například vztah (1.4) je uveden bez zdůvodnění nebo odvození. Na straně 11 se píše "Ta (funkce pro časový plán) má za úkol postupně vhodně snižovat učící čas." Tento výklad je mi nesrozumitelný a bylo by potřeba význam této funkce lépe vysvětlit. Stejně tak mám pocit, že tvar ztrátové funkce (1.10), kde zřejmě chybí závorka, mohl být popsán a okomentován podrobněji.

V textu je dále řada sice drobných chyb, ale je jich poměrně dost. Následně uvádím jejich seznam:

1. Abstrakt: "... pomocí klasifikačních LSTM neuronových strukturách."
2. Str. 7: Je zde uvedena zkratka LSTM bez vysvětlení.
3. Str. 13: "... se pouze uvnitř sítě..", "... důležitosti *i-té*vlastnosti.", dále zde nechápu význam výrazu (1.12), tečka by měla být až za ním, a jak se došlo k výrazu (1.13).
4. Str. 14: Odkaz na vztah 1.15 není v závorce.
5. Str. 15: "... vah se s vstupujícím signálem...", "**Není** v neuronových sítích **nebývá**..."
6. Str. 16: Nechápu druhou rovnost v prostření řádku ve výrazu (1.24).

7. Str. 17: "Průběh rovnic (1.24) je graficky...", " W_{k+1} spojující ..."
8. Str. 19: "... i v neuronových sítích..."
9. Str. 29: V tabulce 2.2 jsou uvedeny pojmy jako multinomiální penalizace a různé řešiče, které ale nejsou popsány. Také by v tabulce mělo být napsáno "Řešiče" a ne "Solvey".
10. Str. 32: "... se provádí v **bash** ... "
11. Str. 33: "**Volený** časový úsek byla **volena** jedna hodina."
12. Str. 35: "Kód třídy je uveden ve výpisu kódu", zde buď chybí odkaz (chybí zde totiž tečka za větou) nebo věta nedává příliš smysl.
13. Str. 45: Místo popisku "accuracy" by zřejmě mělo být "přesnost". To samé platí i pro další obrázky.
14. Str. 46: Popisky "True label" a "Predicted label" by také měly být česky.
15. Str. 47: "... podsekcí 3.2.5"

Je vidět, že autor měl textu věnovat více pozornosti. Jak již jsem zmínil, myslím si, že v některých oblastech mohl jít autor více do hloubky. Celý text vlastně obsahuje jen dvě věty s důkazem a podrobnější popis algoritmu zpětného šíření (back propagation), ale není zde jeho varianta pro rekurentní sítě. V třetí kapitole navíc není úplně jasné, zda předzpracování dat autor pouze někde převzal nebo jde o jeho vlastní nápad. Možná bych i zde očekával vyzkoušení různých metod. Na autora mám tedy následující dotaz:

1. Není možné, že neuronová síť preferuje stagnaci časové řady právě kvůli zvolené standardizaci dat? Zkoušel autor i jiné způsoby předzpracování dat?

Předkládaná práce splnila všechny body zadání. Přesto mám pocit, že od práce na úrovni MM bych čekal o trochu více. Buď mohl jít autor více do hloubky v teoretické části, tj. kapitola první a druhá, nebo mohl otestovat např. více metod předzpracování dat. Takto se zdá, že jen zkoušel různé počty vrstev a neuronů v neuronové síti. Autor navíc použil již existující knihovny, což naprosto dává smysl, ale není možné ocenit to, že by třeba některý typ neuronových sítí naimplementoval sám. Vzhledem k výše uvedenému navrhuji bakalářskou práci ohodnotit známkou **C** tedy **dobře**.

V Praze, 11. srpna 2021.

Ing. Tomáš Oberhuber, Ph.D.
 katedra matematiky
 Fakulta jaderná a fyzikálně inženýrská
 ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
 Trojanova 13
 120 00 PRAHA 2