Posudek oponenta diplomové práce

Autor/autorka práce: **Bc. František Koleňák**

Název práce: **Vysvětlitelná umělá inteligence**

Obsah práce

Cílem práce bylo prostudovat problematiku vysvětlitelné umělé inteligence a seznámit se s knihovnami, které tuto problematiku implementují. Dále měl diplomant navrhnout a implementovat minimální životaschopný produkt, který by obsahoval řešení jak klasickou, tak vysvětlitelnou umělou inteligencí.

Kvalita řešení a dosažených výsledků

Vzhledem k tomu, že nejsem příliš znalý problematiky vysvětlitelné inteligence, zajímali mně výsledky této oblasti a i to, jakým způsobem bude strojově provedeno vysvětlení u klasifikačního problému z oblasti ručně psaných znaků. Jako vstupní datová sada byla v práci použita sada MNIST obsahující ručně psané číslice, jako klasifikátor pak byla použita konvoluční neuronová síť. Bohužel jsem z předložené práce asi úplně nepochopil interpretaci výsledků. V práci jsou sice prezentovány výstupy použitých knihoven ve formě map, co ale tyto mapy znamenají, není úplně zřejmé. Dle mého názoru byl nevhodně zvolen příklad, na kterém je činnost knihoven pro „vysvětlení“ demonstrována. Zvolit pro začátek datovou sadu obsahující ručně psané znaky, které jsou z hlediska klasifikace problematické a nechat je klasifikovat konvoluční sítí, která si sice s klasifikací velice dobře poradí, ale z hlediska funkce není příliš průhledná, považuji za nevhodné. Proč například nebyla zvolena datová sada IRIS (která je v práci také zmiňován a která je mnohem jednodušší z hlediska interpretace výsledků) a jako klasifikační algoritmus nebyl např. zvolen rozhodovací strom, popř. vícevrstvý perceptron. Na těchto jednodušších problémech by možná bylo lépe vidět jak „vysvětlující“ knihovny opravdu fungují.

Formální úroveň

Po formální stránce má práce vcelku slušnou úroveň a její rozsah splňuje požadavky na diplomovou práci. Je napsaná anglicky a až na některé části je napsaná vcelku přehledně. Logicky ji lze rozdělit do tří částí. V první části (kap. 1-5) se diplomant zabývá popisem různých algoritmů z oblasti strojového učení a umělé inteligence (AI) a dále popisuje teorii vysvětlitelné umělé inteligence (XAI). Dále se diplomant zabývá popisem knihoven, které pro oblast XAI existují. V této části mám výhrady pouze ke kapitole 2, která je napsaná poměrně chaoticky. Diplomant zde popisuje problematiku neuronových sítí, kdy začíná od popisu umělého neuronu, následuje popis algoritmu zpětného šíření bez uvedení architektur, ve kterých se tento algoritmus používá, pak následuje ukázka architektur sítí, z nichž většina jsou převážně „Deep“ modely, ale o hlubokém učení se čtenář dozví informace až o několik podkapitol dále, mezitím se dozvíte informace o používaných aktivačních funkcích funkcí u umělých neuronů atd. Napříště by bylo mnohem lepší nejprve rozmyslet strukturu kapitoly a pak teprve začít psát.

Druhá část (kap. 6) se zabývá vlastní realizací. Jak již bylo řečeno, diplomant si zvolil datovou sadu MNIST a jako klasifikátor zvolil konvoluční síť. O nevhodné volbě jsem se již zmiňoval v předchozí části posudku. Dále zde považuji za nevhodně zvolený obrázek, který demonstruje, co knihovna MNIST obsahuje. Na webu existuje mnohem více vhodnějších obrázků, ze kterých je lépe patrné, jak vypadají jednotlivé varianty ručně psaných znaků.

Ve třetí části (kap. 7) diplomant diskutuje dosažené výsledky a demonstruje, jak pracují „vysvětlující“ knihovny. Jak jsem se již zmínil v předchozí části, prezentované obrázky ve mně nezanechaly valný dojem o tom, že použité knihovny jsou výrazným přínosem k vysvětlení toho, jak se klasifikátor rozhodoval. Otázkou zůstává, jestli šlo o nevhodnou volbu problému, na kterém byla jejich činnost demonstrována nebo jsou implementované knihovny pouze začátkem a jejich další rozvoj teprve přinese odpovídající výsledky.

Práce dále obsahuje 4 přílohy (seznam zkratek, tabulky s porovnáním vlastností XAI knihoven, uživatelský manuál aplikace a přílohu ve které je popis verzí knihoven použitých v aplikaci). Součástí práce není příloha toho, co lze nalézt na přiloženém CD disku.

Kódy zmiňované v práci jsou napsané v jazyce Python a jsou funkční.

Práce s literaturou

Práce obsahuje 110 odkazů, vesměs se jedná o online články, technické zprávy a manuály. V citacích je i několik knih z oblasti strojového učení a umělé inteligence. Literaturu považuji za relevantní vzhledem ke zpracovávanému tématu.

Splnění zadání

Zadání práce bylo splněno.

Dotazy k práci

1. Jak souvisí volba modelu s vysvětlujícím algoritmem? Zvolím-li např. jiný klasifikátor, mohu dostat jiné vysvětlení u výsledného rozhodnutí?
2. Jaké důvody Vás vedly k tomu, že byla použita pouze jediná trénovací epocha u modelu na str. 50 (tabulka 6.1), kdy při použití 10 epoch se procentuálně zvýšila úspěšnost klasifikátoru a doba výpočtu nebyla v tomto případě příliš kritická (cca 6 minut)?

Jak již bylo řečeno, diplomant splnil zadání diplomové práce a prokázal, že je schopen samostatné inženýrské práce. Práci doporučuji k obhajobě a navrhuji hodnotit známkou

dobře.

V Plzni 25. 8. 2020 Ing. Pavel Mautner, Ph.D.