

Posudek vedoucího diplomové práce

Téma diplomové práce:	Identifikace svazkových částic pomocí detektorů CEDAR a strojového učení na experimentu COMPASS v laboratoři CERN
Řešitel:	Bc. František Voldřich
Vedoucí práce:	Ing. Martin Zemko
Rok:	2022

Slovní hodnocení práce

Cílem této práce bylo seznámit se s částicovými detektory typu CEDAR a současnými metodami používanými pro analýzu dat na experimentu COMPASS. Student měl prozkoumat různé metody strojového učení pro identifikaci částic svazku na základě dat z detektorů CEDAR. Vybraná metoda měla být optimalizována na existujících datech, řádně otestována a integrována do stávajícího systému pro analýzu dat.

Předkládaná práce se skládá z pěti kapitol. V první kapitole autor nejprve stručně představuje experiment COMPASS, jeho systém pro sběr dat a softwarové nástroje pro analýzu dat. Druhá kapitola se zabývá popisem činnosti detektorů CEDAR a problémů, které vznikají při analýze dat. Zde je krátce představena metoda, která byla doposud používána.

Třetí kapitola je věnována metodám strojového učení, jejich vlastnostem a specifickým. Autor postupně rozebírá a definuje základní pojmy, aktivační funkce, účelové funkce a metody optimalizace. Dále zde uvádí různé typy neuronových sítí, jejich definice a principy.

Kapitola čtvrtá se podrobněji zabývá nedostatky předešlých analytických metod a vysvětluje příčiny těchto problémů. Autor zde navrhuje tři různé přístupy strojového učení pro řešení uvedených problémů a popisuje jejich vlastnosti. Dále pak rozebírá přípravu testovacích dat, jejich význam a původ. Rovněž se věnuje posouzení kvality dat a chybám, které se v nich nacházejí.

Poslední, pátá kapitola popisuje samotnou implementaci řešení. Autor představuje vlastní nástroj pro přípravu dat a trénování, který využívá genetického algoritmu pro optimalizaci struktury neuronové sítě. Celkem byly otestovány tři typy neuronových sítí a jejich charakteristiky byly porovnány. Ve všech případech autor verifikoval a zhodnotil jejich efektivitu za použití různých vstupních dat zahrnujících fyzikální data, Monte Carlo data, i šum. Cílem byla kompenzace nedostatku fyzikálních dat pro trénink neuronové sítě. Pozornost byla také věnována simulaci nedokonalostí reálných dat v Monte Carlo datech. V závěru kapitoly je popsána integrace výsledného řešení do existujícího nástroje pro analýzu dat a příklad použití.

Text je srozumitelný, obsahuje minimum překlepů. V práci jsou použity standardní prostředky softwarového inženýrství jako jsou UML diagramy. Autor prokázal, že umí programovat v jazyku Python na pokročilé úrovni a dokáže efektivně využívat existujících knihoven při řešení problému.

Návrh hodnocení, závěr

Autor se řešení úkolu věnoval průběžně po celý akademický rok a pravidelně se mnou konzultoval aktuální stav své práce. Předložená práce popisuje a implementuje kvalitní systém pro identifikaci částic svazku. Protože student splnil všechny body zadání, navrhuji ohodnotit práci známkou **A (výborně)**.