

Univerzita Karlova
Matematicko-fyzikální fakulta

ZÁZNAM O PRŮBĚHU OBHAJOBY
DISERTAČNÍ PRÁCE

Název práce: Neural Network Based Named Entity Recognition

Jazyk práce: angličtina

Jméno studenta/studentky: RNDr. Jana Straková

Studijní program: Informatika

Studijní obor: 4I3 Matematická lingvistika

Školitel: prof. RNDr. Jan Hajič, Dr.

Oponenti: Ing. Miloslav Konopík, Ph.D. (přítomen)
doc. Dr. Ing. Jan Černocký (omluvil se)

Členové komise:

doc. Ing. Zdeněk Žabokrtský, Ph.D. (přítomen)
doc. RNDr. Markéta Lopatková, Ph.D. (přítomna)
prof. PhDr. Jarmila Panevová, DrSc. (přítomna)
prof. PhDr. Eva Hajičová, DrSc. (přítomna)
RNDr. Pavel Pecina, Ph.D. (přítomen)
RNDr. Jan Cuřín, Ph.D. (přítomen)
doc. Ing. Pavel Ircing, Ph.D. (přítomen)
doc. RNDr. Vladimír Petkevič, CSc. (přítomen)

Datum obhajoby: 12. 6. 2017

Průběh obhajoby:

Předseda komise konstatoval, že všichni členové komise jsou přítomni, omluvil se pouze jeden z oponentů doc. Černocký, komise je tedy usnášeníschopná. Dále konstatoval, že všechny podmínky pro uskutečnění obhajoby byly splněny, jak z ohledu studia uchazečky, tak dosavadního průběhu odevzdání práce a dodání posudku. Předseda komise představil uchazečku a popsal průběh jejího studia; ocenil její experimenty a zmínil i komerční uplatnění systému rozpoznávání pojmenovaných entit, který uchazečka vypracovala.

Uchazečka poté představila svou práci, přičemž na začátku ukázala důležitost rozpoznávání pojmenovaných entit (NER) pro další úlohy zpracování jazyka NLP. Její práce samotná popisuje systém NER založený na neuronových sítích, jeho implementaci a vyhodnocení. Uchazečka připomenula dostupná anotovaná data pro NER v angličtině a češtině a rozdíly mezi nimi. Dále ukázala východisko před započítáním práce na své disertaci – systém založený na support vector machines, který implementovala v dřívější době. Osvětlila, že cílila na systém založený na neuronových sítích, ale v době implementace neexistovaly žádné jednoduché frameworky neuronových sítích (jako TensorFlow, Theano) a musela tedy implementaci vytvořit od základu. Dál popisovala, jak neuronovou architekturu svým výzkumem rozšířila a jak se tímto rozšířením zlepšila

Pokyny pro předsedy nebo místopředsedy komisi:

Práce v elektronické podobě musí být studentem vložena do SIS. Formulář vyplňte ve všech bodech v elektronické podobě. V bodě Členové komise se uvedou všichni členové komise a za jejich jména se uvede „přítomen“ nebo „(nepřítomen)“. Předseda nebo místopředseda komise je jejím členem. V bodě Průběh obhajoby by měly být uvedeny alespoň čtyři věty vystihující průběh obhajoby. Po vyplnění formuláře ho vytiskněte, dole formulář ještě vlastnoručně podepište a přiložte k zápisu o státní závěrečné zkoušce. Současně vložte formulář v elektronické podobě (bez vlastnoručního podpisu) do SIS.

úspěšnost. Výsledná neuronová síť je rekurentní, tj. pracuje s kontextem věty, a nepotřebuje ručně vytvořené klasifikační rysy (využívá však v některých konfiguracích lemmatizaci a morfologické značky). Klasifikační síť používá parametrické jednotky ReLU a výstupní softmaxovou vrstvu. Vstupní rekurentní vrstva sítě extrahuje automaticky vytvořené rysy z textu. Slova vstupního textu jsou zde reprezentovány embeddingy (vektory čísel) na úrovni slov i znaků, dále i 1-hot kódováním a případně i dalšími rysy (morfologie). Uchazečka porovnávala jednotlivé reprezentace a jejich úspěšnost na úloze NER – čím bohatší reprezentace, tím vyšší – a ukázala, že dosáhla state-of-the-art výsledků pro češtinu čistě na základě embeddingů a morfologie. Na angličtině srovnala výkon systému s podobnými architekturami vyvinutými ve stejné době, které dosahují o něco málo lepších výsledků, protože jejich architektura více zapojuje kontext (díky použití conditional random fields jako výstupní vrstvy).

Vedoucí práce prof. Hajič přečetl svůj doporučující posudek a zdůraznil publikační činnost uchazečky. Oponent dr. Konopík přečetl posudek a zdůraznil zejména své dotazy – k anotaci korpusu NER, podrobnostem trénování neuronové sítě a embeddingů. Zdůraznil, že ač práce postrádá radikální nové postupy, je velice kvalitní. Poznamenal, že uchazeččin systém NER úspěšně použil i ve vlastní výuce, ač se zde jednalo o jinou doménu, než na které byl systém trénován. To dokazuje robustnost systému. Práci jednoznačně doporučil k obhajobě. Předseda komise přečetl posudek oponenta doc. Černockého. Oponent kladně hodnotí veřejnou dostupnost systému i trénovacích dat a přes některá kritická doporučení prací považuje za významný přínos k výzkumu v oboru a jasně ji doporučuje k obhajobě. Připojil několik dotazů (entity linking, zapojení širšího kontextu, redukce dimenzionality, systém dostupný online atd.), které uchazečka oponentovi již zodpověděla dříve písemně a zároveň své odpovědi zde prezentovala přítomným. Oponent dr. Konopík byl s odpověďmi spokojen, dodal ještě doplňující otázky k dotazům doc. Černockého (k anotaci korpusu NER, entity linking), které se v diskusi s uchazečkou vyjasnily.

Předseda komise dále vysvětlil způsob anotace českého NER korpusu, na níž se také podílel a otevřel veřejnou rozpravu, do které se přihlásil dr. Pavel Straňák (dotaz evaluace NER na PDT2.5). Předseda komise konstatoval, že uchazečka uspokojivě zodpověděla všechny vznesené dotazy. Následovala neveřejná diskuse členů komise a tajné hlasování.

Komise jednohlasně schválila udělení titulu Ph.D.

Počet publikací: 12

Výsledek hlasování:

Počet členů s právem hlasovacím: 8

Počet přítomných členů: 8

Odevzdáno hlasů kladných: 8

Odevzdáno hlasů neplatných: 0

Odevzdáno hlasů záporných: 0

Výsledek obhajoby: prospěl/a neprospěl/a

Předseda nebo místopředseda komise: doc. Ing. Zdeněk Žabokrtský, Ph.D.

Pokyny pro předsedy nebo místopředsedy komisi:

Práce v elektronické podobě musí být studentem vložena do SIS. Formulář vyplňte ve všech bodech v elektronické podobě. V bodě Členové komise se uvedou všichni členové komise a za jejich jména se uvede „(přítomen)“ nebo „(nepřítomen)“. Předseda nebo místopředseda komise je jejím členem. V bodě Průběh obhajoby by měly být uvedeny alespoň čtyři věty vystihující průběh obhajoby. Po vyplnění formuláře ho vytiskněte, dole formulář ještě vlastnoručně podepište a přiložte k zápisu o státní závěrečné zkoušce. Současně vložte formulář v elektronické podobě (bez vlastnoručního podpisu) do SIS.