

ZÁZNAM O PRŮBĚHU OBHAJOBY
DISERTAČNÍ PRÁCE

Název práce: Unsupervised Dependency Parsing

Jazyk práce: angličtina

Jméno studenta/studentky: RNDr. David Mareček

Studijní program: Informatika

Studijní obor: I-3 Matematická lingvistika

Školitel: doc. Ing. Zdeněk Žabokrtský, Ph.D., ÚFAL MFF UK

Oponenti: Anders Søgaard, Ph.D., University of Copenhagen
Ing. Mgr. Filip Jurčíček, Ph.D., ÚFAL MFF UK

Předseda komise: doc. RNDr. Markéta Lopatková, Ph.D., MFF UK

Členové komise: doc. RNDr. Markéta Lopatková, Ph.D., předsedkyně, ÚFAL MFF UK
prof. PhDr. Jarmila Panevová, DrSc., místopředsedkyně, ÚFAL MFF UK
prof. PhDr. Eva Hajičová, DrSc., ÚFAL MFF UK
Ing. Mgr. Filip Jurčíček, Ph.D., ÚFAL MFF UK
RNDr. Pavel Pecina, Ph.D., ÚFAL MFF UK
Mgr. Roman Neruda, CSc., Ústav informatiky, AV ČR (nepřítomen)
RNDr. Jiří Havelka, Ph.D., IBM ČR
Martin Plátek, CSc., KTIML MFF UK

Datum obhajoby: 26. 9. 2012

Průběh obhajoby:

Obhajoba proběhla ve středu 26. 9. 2012 od 10:00 hodin v budově MFF UK v místnosti S1, Malostranské nám. 25, Praha 1.

Přítomní členové: viz prezenční listinu.

Hosté: viz prezenční listinu.

Celá obhajoba probíhala v angličtině. Předsedající doc. RNDr. Markéta Lopatková, Ph.D., přivítala všechny přítomné, představila členy komise a konstatovala, že komise je usnášeníschopná. Předsedající dále stručně představila uchazeče a zkontrolovala úplnost dokladů vyžadovaných k obhajobě. Poté konstatovala, že byly splněny všechny podmínky pro konání obhajoby a že nedošly žádné písemné připomínky. Přítomní

Pokyny pro předsedy nebo místopředsedy komisí:

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.

dostali možnost nahlédnout do výtisků předkládané disertační práce a do seznamu publikovaných prací uchazeče.

Školitel, doc. Ing. Zdeněk Žabokrtský, Ph.D., vyjádřil své stanovisko k průběhu doktorského studia uchazeče a seznámil přítomné se svým hodnocením obhajované disertační práce. Uchazeč dostal slovo a výstižně představil svou práci – motivaci pro neřízený závislostní parsing (unsupervised dependency parsing), metodu počítání vypustitelnosti (reducibility) včetně názorných příkladů v angličtině, použitý generativní model, metodu Gibbs sampling a dosažené výsledky.

Předsedající požádala oponenty o přednesení posudků. Anders Søgaard, Ph.D., kladně hodnotil práci i její prezentaci a položil dotaz ohledně vymezení metod přípustných v neřízeném parsingu, zda se mohou využívat morfologické značky získané řízeným učením a jak hledat hyperparametry. Ing. Mgr. Filip Jurčíček, Ph.D., ocenil výběr tématu a přístupu, označil práci jako dobře napsanou a přehlednou, ale uvedl, že by uvítal více detailů, například vysvětlení pojmu *neprojektivní strom* pro čtenáře z jiných oborů. Zeptal se na rozdíl mezi Dirichletovou distribucí a procesem ve vztahu k Chinese restaurant process a konečnému počtu parametrů. Další dotaz se věnoval vymezení pojmu *historie* v Gibbsově samplingu.

Uchazeč se detailně vyjádřil ke všem poznámkám a otázkám oponentů. Uvedl, že hledání optimálních hyperparametrů pro každý jazyk zvlášť by sice vedlo k lepším skóre, ale odporovalo by to požadavku na neřízenost učení a budoucí využití pro jazyky, ke kterým neexistují anotovaná data. Princip historie v Gibbsově samplingu vysvětlil s odkazem na větu o zaměnitelnosti (exchangeability). Oba oponenti vyjádřili svou spokojenost s odpověďmi uchazeče.

Předsedající otevřela veřejnou rozpravu. Anders Søgaard, Ph.D., navrhl použití EM algoritmu s náhodnými restarty. RNDr. Jiří Havelka, Ph.D., se zeptal, co se myslí pojmem unsupervised part-of-speech tags a uchazeč odpověděl, že se jedná o word clustering, citoval autory použité metody a dodal, že jednou z možností budoucího výzkumu je pokusit se o současný tagging a parsing. Anders Søgaard, Ph.D., se zeptal na cílové aplikace. RNDr. Jiří Havelka, Ph.D., položil otázku ohledně samplingu neprojektivních stromů. Uchazeč vysvětlil, že jeho současný přístup sampluje projektivní stromy, ale závěrečná fáze používá algoritmus MST, takže může vzniknout i neprojektivní strom. Dále doplnil, že v předchozím přístupu zkoušel i sampling neprojektivních stromů, ale ten vedl k horším výsledkům.

Na závěr předsedající shrnula průběh obhajoby a konstatovala, že uchazeč výstižně přednesl výsledky své práce a že uspokojivě odpověděl na dotazy, které byly vzneseny v diskusi. Následovala neveřejná diskuse členů komise a tajné hlasování. Komise jednohlasně schválila udělení titulu Ph.D.

Zapsal: Mgr. Martin Popel, 26. 9. 2012

Počet publikací: 20

Výsledek hlasování:

Počet členů s právem hlasovacím: 8
Počet přítomných členů: 7
Odevzdáno hlasů kladných: 7
Odevzdáno hlasů neplatných: 0
Odevzdáno hlasů záporných: 0

Výsledek obhajoby: Prospěl

Předseda nebo místopředseda komise:

doc. RNDr. Markéta Lopatková, Ph.D., MFF UK

Pokyny pro předsedy nebo místopředsedy komisí:

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.