

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Incorporating Language Models into Non-autoregressive Neural Machine Translation
Jméno autora:	Zdeněk Kasner
Typ práce:	diplomová
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačů
Vedoucí práce:	Jindřich Helcl
Pracoviště vedoucího práce:	Ústav formální a aplikované lingvistiky (MFF UK)

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání <i>Hodnocení náročnosti zadání závěrečné práce.</i>	náročnější
--	-------------------

Splnění zadání <i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	splněno
---	----------------

Zvolený postup řešení <i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	vynikající
---	-------------------

Odborná úroveň <i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	A - výborně
---	--------------------

Formální a jazyková úroveň, rozsah práce <i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	A - výborně
--	--------------------

Výběr zdrojů, korektnost citací <i>Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.</i>	A - výborně
---	--------------------

Další komentáře a hodnocení <i>Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.</i>	
--	--

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Diplomová práce Zdeňka Kasnera se zabývá začleněním n-gramových jazykových modelů do modelů neautoregresivního překladu trénovaných pomocí Connectionist Temporal Classification (CTC).

Text práce je rozdělen do sedmi kapitol. V úvodu sdělí autor motivaci pro svůj výzkum (zrychlení dekodování z neuronových modelů při zachování kvality překladu) a vytyčí cíle práce. Druhá kapitola shrnuje vývoj a metody strojového překladu až k překladu pomocí neuronových sítí (NMT). Třetí kapitola je věnována neautoregresivním modelům pro NMT. Poslední ze třech prezentovaných přístupů založený na výpočtu hodnoty ztrátové funkce pomocí CTC je přístupem, ze kterého vychází experimenty popisované ve zbytku práce. Ve čtvrté kapitole je prezentován původní výzkum, jímž se tato práce zabývá – zakomponování n-gramových jazykových modelů a dalších rysů do prohledávacího algoritmu u modelů trénovaných s CTC. Pátá kapitola popisuje použité nástroje a implementační detaily řešení experimentů. V šesté kapitole je popsáno experimentální nastavení – data, varianty modelů a parametry trénování. Dále je zde prezentována analýza výsledků a diskuse. V závěrečné kapitole je podáno shrnutí a návrhy na další výzkum.

Práce má 63 stran a je psána anglicky. Příložené jsou použité skripty a nástroje v odpovídající verzi.

Text práce má přehlednou strukturu a je psán velmi dobrou angličtinou. Obecný přehled je adekvátní a vzhledem k relativně úzkému zaměření nepostrádá řešerše související literatury žádné významné příspěvky. Za zmínku by možná stály insertion-based metody dekodování a metody dekodování v konstantním čase (při paralelizaci a vzhledem k délce cílové věty), většina z nich však byla publikována natolik nedávno (některé až po odevzdání této práce), že lze jejich absenci v řešerši lze pochopit.

Práce je rozšířením našeho předchozího výzkumu neautoregresivního překladu s CTC (Libovický & Helcl, 2017). Autorovi se podařilo ukázat, že obětováním velmi malého množství rychlosti překladu lze získat výrazné zlepšení kvality překladu (řádky „End-to-End w/ CTC“ a „Ours, beam 5“ v tabulkách 6.5 a 6.6). Přestože totiž beam search pracuje sekvenčně, cena za jeho provedení je v mnohem nižší než cena autoregresivního generování, při kterém musí být několikrát spouštěn celý dekodér. Je však patrné, že kvalita překladu tohoto neautoregresivního modelu pořád nedosahuje kvality autoregresivních modelů.

Velmi pozitivně hodnotím, že v rámci diplomové práce vznikl článek, který jsme poslali na konferenci ACL (CORE ranking A*). Přestože na tuto konferenci přijat nebyl, je to velmi cenný výstup diplomové práce. Momentálně je článek v recenzním řízení na konferenci EMNLP (A). Práce na článku jde dál rozšířit, zejména o použití tzv. back-translation, tj. metody využití jednojazyčných dat v samotném překladovém modelu, nejen pro trénování jazykového modelu.

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **A - výborně**.

Datum: 13.6.2019

Podpis: Mgr. Jindřich Helcl