



Porozumění řeči založené na neuronových sítích

Adam Frémund¹

1 Úvod

Zpracování řeči a přirozeného jazyka je oblastí umělé inteligence. Jedná se o obor, jehož cílem je naučit stroje zpracovávat jazyk tak, jako to umí člověk. Jinými slovy řečeno jde o schopnost jazykovému projevu porozumět a také sdělení v přirozeném jazyce vyprodukovat. V dnešní době, kdy roste popularita hlasových asistentů (Amazon Alexa, Google Home, atd.), se zpracování přirozeného jazyka stává velice perspektivní disciplínou. Mezi typické úlohy zpracování přirozeného jazyka patří převod řeči na text, syntéza řeči, analýza sentimentu, detekce pojmenovaných entit a mnoho dalšího. Poslední dvě zmíněné úlohy jsou předmětem experimentů této diplomové práce.

2 Formulace úloh

Pro strojové zpracování přirozeného jazyka se v posledním desetiletí začaly hojně používat neuronové sítě. Jedním z nejmodernějších přístupů je tzv. *Transformer* architektura (viz. Vaswani (2017)). Právě modelům s touto architekturou je věnována celá diplomová práce. Modely používané pro porozumění řeči bývají shodné s těmi, které se používají pro zpracování přirozeného jazyka v textové podobě. Proto jsou *Transformer* modely nejprve použity pro několik úloh porozumění textu. V rámci této diplomové práce jsou zkoumány modely typu BERT a T5 (Text-To-Text Transfer Transformer). T5 model představuje zcela nový přístup, jenž umožňuje použití jednoho modelu pro několik specifických úloh zpracování přirozeného jazyka, aniž by bylo potřeba měnit jeho architekturu.

Prvním cílem této diplomové práce bylo provést srovnání obou modelů, tj. BERT a T5 modelu. V první řadě byly obě architektury použity pro analýzu sentimentu anglických a českých recenzí z internetových filmových databází. Druhou úlohou, na které byly porovnány výkonnosti obou modelů, je detekce pojmenovaných entit *Czech Named Entity Corpus* (CNEC, Sevcikova (2007)) textového korpusu.

Poslední zkoumanou úlohou bylo porozumění řeči na úrovni sémantických konceptů. Tato úloha ukazuje použití T5 modelu s nedokonalým výstupem systému automatického rozpoznávání řeči. Jedná se o klíčovou úlohu při implementaci zmíněných hlasových asistentů. Druhým hlavním cílem této práce bylo zjistit, zda je T5 model v této úloze konkurence schopný pro model navržený speciálně pro danou úlohu.

3 Dosažené výsledky

Analýza sentimentu spočívá v identifikaci a extrakci vztahu autora k předmětu textu. Názor je subjektivní informací, která je zanesena v textu. Samotný sentiment lze rozdělit do

¹ student navazujícího studijního programu Aplikované vědy a informatika, obor Kybernetika, specializace Umělá inteligence, e-mail: afremund@students.zcu.cz

	Metrika F1				Metrika Acc	
	IMDB	ČSFD	CNEC		HHTT+TIA	
			7 entit	42 entit	validační dataset	testovací dataset
anglický BERT _{BASE}	0.9378	-	-	-	-	-
anglický T5 _{BASE}	0.9523	-	-	-	-	-
multilingual BERT _{BASE}	-	0.9302	0.8184	0.8225	-	-
český BERT _{BASE}	-	0.9315	0.8363	0.8246	-	-
český T5 _{BASE}	-	0.9470	0.7871	0.7495	0.6970	0.7425
model pro úlohu detekce sémantických entit	-	-	-	-	0.6970	0.7451

Tabulka 1: Nejlepší dosažené výsledky pro modely typu BERT a T5. Pro korpusy IMDB, ČSFD a CNEC se jedná o výsledky dosažené na testovacích datech.

několika tříd. Pro účely této práce byly uvažovány pouze dvě třídy - *negativní* a *pozitivní*. Z tabulky 1 je patrné, že pro úlohu detekce sentimentu dosáhl T5 model výrazně lepších výsledků než modely typu BERT. A to jak pro anglické (IMDB), tak i české (ČSFD) filmové recenze.

Detekce pojmenovaných entit v textu je nedílnou součástí zpracování přirozeného jazyka. Jedná se o jednu z nejtěžších úloh především z důvodu nejednoznačnosti jazyka. Cílem je detekovat a správně zařadit slova nebo slovní druhy do předem stanovených tříd. Pro účely této úlohy byl použit textový korpus CNEC. Tento korpus obsahuje anotované texty obsahující entity typu *číslo*, *osobní jména*, *časové výrazy*, *geografické názvy*, atd. Tato úloha byla rozdělena na detekci 7 hlavních (*osobní jména*, *časové výrazy*, *geografické názvy*, ...) a 42 detailních (*jména*, *příjmení*, *akademické tituly*, *státy*, *města*, ...) pojmenovaných entit. V této úloze dosáhly modely typu BERT lepších výsledků ve zvolené metrice než model T5.

V poslední úloze bylo cílem zjistit, jak si T5 model povede při rozpoznávání sémantických entit v řečovém korpusu (HHTT+TIA). Zde byl T5 model porovnán s modelem navrženým speciálně pro danou úlohu. V této úloze jsou použita data z nedokonalého systému automatického rozpoznávání řeči. V této úloze dosáhl T5 model obdobných výsledků jako model navržený speciálně pro tuto úlohu. Předtrénovaný T5 model na rozdíl od navrženého modelu není primárně určen pouze k jedné úloze. Navíc se jedná o základní verzi této architektury bez jakýchkoliv úprav.

4 Závěr

Lze říci, že T5 model svou architekturou přináší nové poznatky do zpracování přirozeného jazyka i porozumění řeči. Dosažené výsledky v rámci této diplomové práce poukazují na možný potenciál této nové architektury. Vzhledem k tomu, že doposud nebyl vydán žádný článek, který by se věnoval použití T5 modelu pro úlohy zpracování českého jazyka, lze tuto práci považovat za inovativní. Práce se snaží přispět k obohacení dosavadních poznatků a zároveň poskytuje námět k dalším experimentům s touto architekturou.

Literatura

- Ashish Vaswani and Noam Shazeer and Niki Parmar and Jakob Uszkoreit and Llion Jones and Aidan N. Gomez and Lukasz Kaiser and Illia Polosukhin (2017) *Attention Is All You Need*.
- Magda Ševčíková and Zdeněk Žabokrtský and Oldřich Krůza (2007) *Named Entities in Czech: Annotating Data and Developing NE Tagger* vol. 4629, pp. 188–195. Springer, Heidelberg.