

Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Samuel Michalik	Samuel Michalik	
Název práce	Deep learning and visualization of models for image captioning and multimodal translation	
Rok odevzdání	2020	
Studijní program	Informatika	
Studijní obor	Obecná informatika	
Autor posudku	Mgr. Jindřich Helcl	Vedoucí
Pracoviště	Ústav formální a aplikované lingvistiky	

K celé práci

lepší OK horší nevyhovuje

	lepší	OK	horší	nevyhovuje
Obtížnost zadání	X	X		
Splnění zadání		X		
Rozsah práce <i>... textová i implementační část, zohlednění náročnosti</i>	X	X		
<p>Práce se zabývá vizualizací a trénováním neuronových sítí pro generování popisků obrázků a jejich překladu. Je rozdělená na experimentální a implementační část. Experimentální část tvoří trénování modelů a evaluace na veřejně dostupných obrázkových datasetech. Cílem implementační části bylo vytvořit webové rozhraní, kde bude možné vybrat natrénovaný model a pro daný vstupní obrázek vizualizovat proces generování popisku. Pro generování se používá algoritmus beam search. V n-tém kroku algoritmu je možné zobrazit si přes obrázek tzv. attention heat map, která ukazuje, na které části obrázku se neuronová síť nejvíce zaměřovala během generování n-tého slova.</p> <p>Podle mého názoru jde rozsahem i obtížností zadání o náročnější práci, kterou autor zvládl splnit bez závažnějších nedostatků.</p>				

Textová část práce

lepší OK horší nevyhovuje

	lepší	OK	horší	nevyhovuje
Formální úprava <i>... jazyková úroveň, typografická úroveň, citace</i>	X	X		
Struktura textu <i>... kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu</i>	X	X		
Analýza		X		
Vývojová dokumentace		X		
Uživatelská dokumentace		X		
<p>Práce je psaná dobrou angličtinou, je strukturovaná jasně a přehledně. Moje poznámky jako vedoucího práce autor do textu průběžně zapracovával a vznikl tím text, s jehož kvalitou jsem spokojen. Jedna připomínka k textu se týká citování – články, které vyšly ve sborníku konference nebo v časopise a zároveň jako preprint na arxivu je lepší citovat s odkazem na daný sborník/časopis, nikoliv na arxiv. Na tohle jsem ale autora během psaní práce neupozornil a nepovažuji to u bakalářské práce za výraznou chybu.</p>				

Implementační část práce

lepší OK horší nevyhovuje

Kvalita návrhu ... architektura, struktury a algoritmy, použité technologie	X			
Kvalita zpracování ... jmenné konvence, formátování, komentáře, testování	X	X		
Stabilita implementace		X		

Do hodnocení implementační části zahrnuji i hodnocení části experimentální (tj. kolonky návrh a zpracování výše).

Z hlediska implementace, předložený program obsahuje všechny funkce zmíněné v zadání, čili zejména vizualizaci beam searche a attention map pro každý krok algoritmu. Dále je zde možné vkládat datasety a upravovat způsoby předzpracování dat a další.

Co se týče části experimentální, původní plán bylo zkoumat vliv zahuštění dat pomocí různého předzpracování obrázků (ořez, zvětšení, mírný posun) na zlepšení kvality popisků. Během řešení se ale ukázalo, že předzpracování nemá na popisky valný vliv, zřejmě kvůli tomu, že použitá konvoluční síť, kterou používáme na výpočet spojitě reprezentace obrázku, je už vůči těmto drobným změnám v obrázcích robustní, a tak produkuje podobné reprezentace.

Místo toho se autor rozhodl prozkoumat různé změny v architektuře neuronového dekodéru, který generuje popisek. Prozkoumávané architektury volil autor sám podle toho, co se používá v soudobé literatuře. Jedna ze zkoumaných architektur (tzv. scaled general attention) je původní autorovou inovací, kterou pokud je mi známo nikdo dříve netestoval. Výsledky tohoto modelu však nebyly příliš odlišné od ostatních. Velmi kladně zde hodnotím iniciativu vymýšlet nové experimenty a relativně poctivé otestování všech navržených modelů.

Celkové hodnocení Výborně**Práci navrhuji na zvláštní ocenění** Ne

Datum

Podpis