INSTITUTIONEN FÖR DATAVETENSKAP | LUNDS TEKNISKA HÖGSKOLA | PRESENTERAD 2022-02-15

EXAMENSARBETE Arabic Image Captioning using Pre-training of Deep Bidirectional Transformers
STUDENT Jonathan Emami
HANDLEDARE Pierre Nugues (LTH), Ashraf Elnagar (UOS), Imad Afyouni (UOS)
EXAMINATOR Jacek Malec (LTH)

Arabisk bildtextgenerering med hjälp av förtränande transformer-modeller

POPULÄRVETENSKAPLIG SAMMANFATTNING **Jonathan Emami**

Automatisk bildtextgenerering är idag ett utmanande problem inom datorseende och naturlig språkbehandling. Engelsk bildtextgenerering har sett stora framsteg de senaste åren, medan forskning på arabisk bildtextgenerering har hamnat efter. I detta examensarbete har vi utvecklat och utvärderat flera modeller för arabisk bildtextgenerering, alla initierade på förtränade transformer-modeller.

Bildtextgenerering har många olika tillämpningar, exempelvis effektiv bildsökning, auto-arkivering och som stöd för synskadade. De bästa bildtextgeneringsmodellerna idag följer en kodar-avkodar arkitektur:

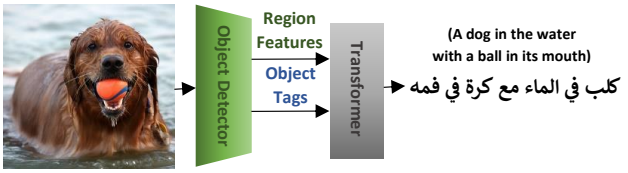
1. Extrahera den viktigaste informationen om bildens olika regioner m.h.a. en objektdetektor, t.ex. en CNN-kodare.

2. Generera en mening från den extraherade vektorn m.h.a. en språkmodell, t.ex. en RNN-avkodare.

I detta examensarbete använde vi förtränade transformer-modeller för att initialisera våra modeller för bildtextgenerering. Därefter finjusterade vi modellerna genom att träna dem på bild-text par med en inlärningsmetod som heter OSCAR. Denna inlärningsmetod använder sig av objekttaggar, detekterade i bilden, som ankarpunkt för att underlätta inlärningen av bild-text semantik.

Vårt examensarbete handlade om att utforska prestandan hos fyra olika transformer-modeller på ett bildtextdataset. De fyrade testade modellerna var *Multilingual BERT*, *AraBERT*, *ArabicBERT* och *GigaBERT*.

Våra resultat visar på bra inlärningsförmåga för alla våra modeller, men att AraBERT fick bättre evalueringspoäng. Figuren visar en bildtext genererad från AraBERT tränad på datasetet.



Dessutom visade vi att det är möjligt att få bra resultat genom att träna flerspråkiga transformer-modeller, som GigaBERT, på arabisk bildtext med engelska objekttaggar. Däremot drar vi slutsatsen att en modell tränad på ett rent arabiskt dataset, med arabiska objekttaggar, presterar bättre.