INSTITUTIONEN FÖR DATAVETENSKAP | LUNDS TEKNISKA HÖGSKOLA | PRESENTERAD 2021-05-07

EXAMENSARBETE Finding company logos through their websites: A multimodal approach
STUDENTER Emil Wihlander, Jesper Berg
HANDLEDARE Marcus Klang (LTH)
EXAMINATOR Flavius Gruian (LTH)

Hitta företagsloggor med maskininlärning genom text och bild

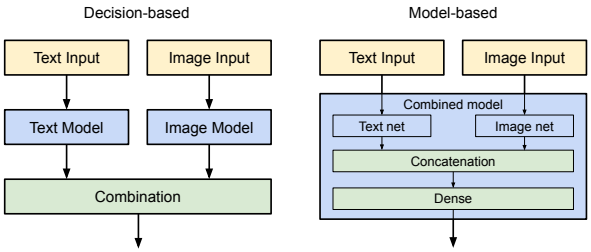
POPULÄRVETENSKAPLIG SAMMANFATTNING **Emil Wihlander, Jesper Berg**

Att automatiskt hämta ett företags logga från deras hemsida kan vara användbart i flera sammanhang. Men hur vet en dator vilken av alla bilder på hemsidan som är loggan? Detta arbete utforskar olika tekniker för att identifiera loggan, inklusive *multimodalitet* vilket använder både bild och text data för att hitta rätt.

Ibland finns flera typer av indata tillgängligt när något ska tolkas, till exempel använder människor både munrörelser och ljud när de tolkar tal. Detta kallas för multimodalitet. Multimodalitet har länge varit intressant i samband med maskininlärning och flera fall finns där det varit lyckat att använda sig av det. Detta eftersom de olika modaliteterna kan innehålla kompletterande information. Målet med detta arbete är att låta datorn avgöra vilka bilder på en hemsida som är företagets logga – men varför är detta problem multimodalt? Jo, bilder på hemsidor har, utöver bilden i sig, även en sökväg som webbläsaren använder sig av för att läsa in bilden. Detta innebär att det finns både bilden i sig samt en text som representerar varje hemsidebild.

I examensarbetet jämförs olika metoder för att låta datorn avgöra bilderna och texterna separat. De metoderna som testas är enkla tekniker så som "innehåller sökvägen ordet *logo*" samt maskininlärning för både bild och text. Efter detta kombinerades modaliteterna genom att kombinera resultatet från de individuella metoderna med ett viktbaserat medel och med maskininlärning. En stor modell som tar både bilderna och texterna som indata testades också.

Bilden nedan visar hur de olika sätten att kombinera modaliteterna ser ut. Den vänstra visar när resultatet från individuella modeller kombineras med en beslutsmodell och den högra hur en stor modell med båda indata ser ut.



Resultatet visar att textklassificering, både genom enkla metoder så väl som maskininlärning, lyckas hitta loggor med rätt hög precision. Bildklassificering presterar sämre då den har problem med kontext och har svårt att koppla ihop rätt logga med rätt hemsida. De multimodala lösningarna visar överlag liknande resultat som textklassificeringen, men en viktbaserad beslutsmodell presterar bättre än någon modell med endast en modalitet. Kompletteringstester mellan modaliteterna visade också att multimodalitet hade väldig hög potential. Dessa två resultaten indikerar på att multimodalitet är ett bra angreppssätt i detta kontext.