INSTITUTIONEN FÖR DATAVETENSKAP | LUNDS TEKNISKA HÖGSKOLA | PRESENTERAD 2022-06-03

EXAMENSARBETE Investigating and Mitigating Effects of Quantization on Algorithmic Bias
STUDENTER Oscar Andersson, William Isaksson
HANDLEDARE Flavius Gruian (LTH), Axel Berg (Arm), Felix Johnny Thomasmathibalan (Arm)
EXAMINATOR Jacek Malec (LTH)

Påverkan av kvantisering på algoritmiskt bias i deep learning

POPULÄRVETENSKAPLIG SAMMANFATTNING **Oscar Andersson, William Isaksson**

Kvantisering av neurala nätverk gör det möjligt att använda sig av maskininlärning där det annars inte vore möjligt som t.ex. mikroprocessorer. I detta arbete undersöker vi hur kvantiseringen påverkar prestanda olika för olika grupper av datan. Vi föreslår också hur man kan lindra dessa icke önskvärda effekter.

Deep learning har blivit en allt mer populär typ av maskininlärning. Deep learning utgörs av artificiella neurala nätverk som är inspirerade av den mänskliga hjärnan på en ytlig nivå. Dessa nätverk kan lära sig mönster genom att träna på väldigt stora mängder data som t.ex. bilder. De kan t.ex. lära sig att avgöra vilken hårfärg en person har. Ett sådant nätverk lär sig att avgöra hårfärg enbart genom att titta på bilder med associerad hårfärg. Den kommer lära sig trender i dessa bilder, som kanske inte är önskvärda. T.ex. om nästan alla bilder på blonda personer består av kvinnor, riskerar nätverket att lära sig att män inte kan vara blonda. Detta är ett exempel på algoritmiskt bias. Igenkänning av hårfärg är inte särskilt problematiskt, men i andra fall, som t.ex. igenkänning av olika typer av cancer, kan sådant algoritmiskt bias vara problematiskt.

Deep learning har traditionellt sett varit resurskrävande, men tack vare bättre processorer och bättre mjukvara, kan man idag använda deep learning även i resursbegränsade miljöer, så som en smartwatch eller till och med en mikrovågsugn. Neurala nätverk använder sig av en stor mängd parametrar (eller sparade nummer), som tar upp mycket datorminne och datorkraft. Dessa nummer tar upp olika mycket minne beroende på dess precision, t.ex. hur många decimaler som sparats. Att reducera precisionen av ett nätverk kallas kvantisering av nätverket. Kvantisering är i vissa fall nödvändigt för att ett nätverk ska kunna användas i en digital enhet. Kvantisering kan påverka prestandan på vissa grupper oproportionerligt mycket, t.ex. skulle nätverket kunna bli mycket sämre på att avgöra om en person är brunett, men i övrigt vara opåverkat.

I detta examensarbete har vi undersökt hur prestandan på olika hårfärger påverkas olika av kvantisering. Vi undersöker också om prestandan för en hårfärg påverkas olika mellan män och kvinnor samt gamla och unga personer. Detta undersöks på fyra nätverk och vi kommer fram till att underrepresenterade härfärger påverkas mer negativt än andra hårfärger. Vi undersöker också hur olika nätverk och olika inställningar av nätverk påverkas olika av kvantisering.

För att lindra dessa icke önskevärda effekter, föreslår vi ett par strategier. Vår första strategi är att se till datan som nätverket lär sig av är balanserad. Vår andra strategi ser till att nätverket påverkas mindre av kvantisering och således påverkar algoritmiskt bias mindre också. Det tredje alternativet handlar om att kompensera för förändringarna i det kvantiserade nätverket.