

**EXAMENSARBETE** Correction of grammatical errors in Swedish**STUDENTER** Joel Ehnroth, Yoonjoo Park**HANDLEDARE** Pierre Nugues (LTH), Filip Bolling (NordAxon AB)**EXAMINATOR** Jacek Malec (LTH)

# Korrigerings av grammatiska fel med hjälp av maskininlärning

POPULÄRVETENSKAPLIG SAMMANFATTNING Joel Ehnroth, Yoonjoo Park

Verktyg som erbjuder hjälp med stavning och grammatik är viktiga i det digitala samhället. I detta examensarbete presenterar vi maskininlärningsmodeller som presterar bättre än traditionella regelbaserade system.

Detta examensarbete utforskar maskininlärningsmetoder för rättning av grammatiska fel i svensk text, såsom stavfel, böjningsfel, felaktiga ordval, med mera. Resultaten visar att maskininlärningsmetoderna presterar lovvärt och kan korrigerar fel till en högre grad än tidigare regelbaserade system.

System för automatisk rättning av olika fel i skriven text har blivit en del av vardagen för många. De återfinns i ordbehandlingsprogram, när vi skriver epost eller SMS, eller när vi gör en sökning via en sökmotor i webbläsaren.

Att kunna formulera sig korrekt och precist är viktigt i dagens samhälle, och att kunna få hjälp av ett verktyg som uppfattar felaktigheter i våra formuleringar är därför av intresse. Särskilt relevant är det möjligen för andraspråkselever som vill underlätta lärandet på egen hand, när en mänsklig pedagog inte finns att tillgå.

Trots att automatiska system för rättning av språkfel har funnits i många år, domineras området av de stora internationella språken, främst engelskan. Detta beror i stor utsträckning på att datatillgång är en central del vid framtagning av maskininlärningsmodeller.

Modellerna som har utvärderats under detta ar-

bete har tränats på stora mängder text från svenska webbsidor. Modellerna kan genom att observera text skapa en intern modell av hur språket fungerar. Denna interna modell kan sedan med fördel användas för en rad olika syften, exempelvis rättning av grammatiska fel.

Gemensamt för maskininlärningsmodeller är dock att de presterar dåligt på uppgifter som de inte specifikt har tränats för. Detta innebär att modellerna ofta måste tränas på ytterligare data som är särskilt avsedd för den specifika uppgiften. I detta fall är uppgiften rättning av grammatiska fel. Att ta fram ett välfungerande system för det svenska språket kräver alltså särskilda resurser för just svensk grammatik, något som i dagsläget är en bristvara. Under arbetets gång utforskas därför möjligheten att använda syntetisk data för att utvidga de tillgängliga resurserna.

Sammanfattningsvis har vi under examensarbetet tagit fram modeller som avgör om en mening är grammatiskt korrekt eller inte, samt modeller som markerar var i meningen fel finns och vilken typ av fel det rör sig om. Till sist utvecklar vi modeller som försöker korrigerar felen. Modellerna visar goda resultat och det finns stor potential till vidare utveckling och förbättring.