

EXAMENSARBETE Quality Assurance of Generative Dialogue Models used for Language Practice

A Test Framework used for Measuring the Quality of Generative Dialogue Models in an Automated Fashion

STUDENT Johan Bengtsson**HANDLEDARE** Markus Borg (LTH), Alexander Hagelborn (NordAxon)**EXAMINATOR** Emelie Engström (LTH)

Testramverk för att mäta kvaliteten hos Dialoggenererande modeller

POPULÄRVETENSKAPLIG SAMMANFATTNING Johan Bengtsson

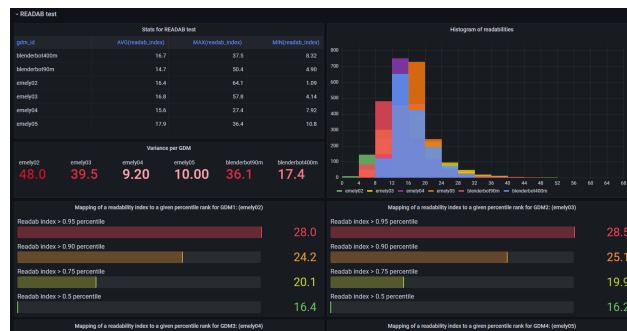
För att lära sig ett språk krävs det bland annat att du talar språket mycket under en längre tid. För detta ändamål utvecklar NordAxon Emely, som baserad på dialoggenererande modeller pratar svenska med användaren. Detta arbete har resulterat i ett testramverk som kan användas för att utvärdera kvaliteten hos sådana produkter.

Vid flytt till ett nytt land står du inför olika utmaningar, och att kunna tala det lokala språket underlättar. Men att lära sig ett nytt språk är inte helt enkelt. Det är viktigt att tala språket mycket, något som kräver att man har en samtalspartner vilket inte är enkelt att hitta. För att förenkla detta så utvecklar NordAxon en produkt som kallas för Emely. Det är en produkt som baseras på Dialoggenererande modeller (GDM) och som går att prata svenska med. GDMer är maskininlärningsbaserade modeller som svarar med människolika svar på det du säger. Men hur mäts kvaliteten på en sådan produkt?

Detta arbete inleddes med att intervjua SFI-lärare för att förstå vad som är viktigt hos samtalspartnern. Därefter skickades formulär ut för att samla in deras åsikter kring tidigare forskningsresultat på området. Sammanlagt så är de viktigaste egenskaperna att språket ska vara koherent, att det inte ska vara ovänligt och att nivån ska justeras efter språkstudenten.

Jag har skapat ett program där två GDMer producerar över tusen konversationer, som sedan analyseras. För att analysera koherensen och nivån av ovänligt språk användes två olika maskininlärningsmodeller på varje meddelande. För

att analysera nivån på språket användes frekvensordlistor och läsbarhetsindex (LIX), som tillsammans antas kunna visa vilken språklig nivå GDMen är på. Genom de genererade konversationerna fås en bild av vilket vokabulär GDMen har, och hur pass frekvent det är. Till det beräknades LIX för varje meddelande.



Resultaten presenteras i en dashboard som ska hjälpa NordAxon att veta hur bra deras produkt är. Sammanlagt tyder resultaten på att NordAxon lyckats göra Emely alltmer koherent, vänlig och till att oftare ha ett enklare språk. Allt som allt kommer detta bidra till bättre versioner av Emely, som i förlängningen kommer att hjälpa nyanlända till Sverige att snabbare integreras in i samhället.