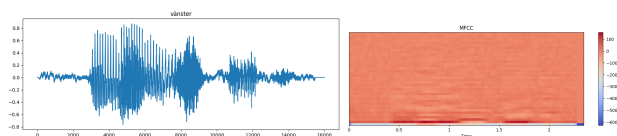


EXAMENSARBETE Domain Adapting English Speech Recognition to Swedish**STUDENT** Michael Hansen**HANDLEDARE** Pierre Nugues (LTH)**EXAMINATOR** Jacek Malec (LTH)

Konsten att lära en engelskförstående maskin svenska

POPULÄRVETENSKAPLIG SAMMANFATTNING **Michael Hansen**

Taligenkänning används mycket och väl idag, men det kräver mycket data för att fungera bra. I detta projekt utvecklades ett system för att träna om en engelsk taligenkänningsmodell till svenska med ett insamlat dataset 100 gånger mindre än det ursprungliga.



Automatisk taligenkänning handlar om att få en dator att översätta mänskligt tal till text. Nyckelordssökning är en fördjupning av detta, där man förenklar uppgiften till att urskilja mellan några specifika ord.

Många av dagens lösningar för taligenkänning kräver massvis med inspelningar, vilket leder till invecklade maskininlärningsmodeller som kräver egna servrar för att behandla ljudet snabbt nog för användning. Genom att minska antalet möjliga ord till ett bestämt antal kan man förenkla modellen så att den kan användas direkt på mobila enheter, vilket eliminerar behovet av internetuppkoppling. Denna sortens modeller behöver dock fortfarande en massa data att träna på.

I 2017 släppte Google ett dataset med $\sim 106'000$ engelska ljudklipp av 35 olika ord. Med hjälp av dessa inspelningar kan forskare utveckla taligenkänningslösningar som presterar bra, visat med en hög träffsäkerhet på deras gissningar. Så stora dataset tar lång tid att samla in och kategorisera. Språk som svenska har inte möjligheten att samla in data av samma storlek, vilket begrän-

sar framsteg och prestationen inom området på hemmaplan. Det leder också till att forskningen blir beroende av de engelska företagen för sådan datainsamling.

I detta projekt har jag utvecklat ett system som använder det engelska datasettet för att träna en taligenkänningsmodell som jag sedan anpassar till svenska ord, med ett mindre svenskt dataset som jag själv samlat in. Med denna metod nådde jag $\sim 86\%$ i träffsäkerhet med en datamängd på $\sim 1\%$ av den engelska. Jämför det med träffsäkerheten på 51% som fås om jag tränar modellen utan att ha använt den engelska datan innan.

Domänanpassning har använts i andra områden för maskininläring, och detta projektet visar att dess resultat är även användbart inom taligenkänning.

Med detta system kan specialiserade program för taligenkänning byggas utan att man behöver samla in en stor mängd data, vilket gör det lättare att, till exempel, skapa mobilapplikationer med taligenkänningsfunktioner. Med detta projekt som utgångspunkt kan man lättare träna taligenkänningsmodeller för nya språk, och det skulle vara intressant att se hur systemet fungerar med språk som har andra ursprung än svenska/engelska.