Kommittén för teknologisk innovation & etik

Komet kommenterar

Maskininlärning – korta faktablad om aktuell teknik

Komet kommenterar 2020:26, publicerad 2020-11-04

Kort om maskininlärning – för beslutsfattare och andra som är nyfikna på hur aktuell teknik påverkar samhället.

Kommenterad rapport

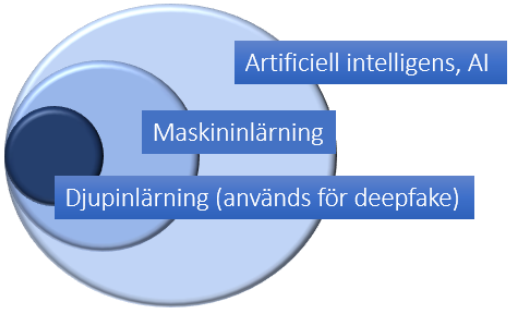
Robinson A and Herbert-Voss A. "Technology Factsheet: **Machine Learning**." Editor Belei B. Paper, Belfer Center for Science and International Affairs, Harvard Kennedy School, June 2019.[1]

Korta faktablad om aktuell teknik

Belfer Center vid Harvard University ger ut en serie faktablad om aktuella teknikområden. Serien är riktad till politiska beslutsfattare i USA i syfte att ge överblick och förståelse av ny teknik. Komet Kommenterar gör en svensk uppföljning av serien.

Belfer Center for Science and International Affairs är del av Harvard Kennedy School of Government. Belfer arbetar bland annat med hur ny teknik kan komma till nytta i samhället.

Komet:s kommentarer



- Maskininlärning är en del av artificiell intelligens, AI (läs gärna faktablad om AI, Komet kommenterar 2020:25).
- En viktig omständighet av etisk betydelse är urval av data för träning av algoritmer vid maskininlärning. Data behöver vara så rättvisande som möjligt och det gäller att vara vaksam på sned-vridning av data för att minska risken för systematiska fel. En bildbehandlingsmodell som enbart tränats på foton av kvinnor riskerar att göra fel om den sedan tillämpas på bilder som visar män.
- Ett välkänt experiment är *The moral machine,* som undersöker moraliska dilemman för självkörande fordon[2]. I experimentet studeras hur etiska principer till grund för styrning av maskiner varierar mellan olika länder och kulturer och om ställningstagande i ett dilemma skiljer sig åt beroende på kön, ålder eller inkomst. Experimentet illustrerar vikten av insyn i vilka data som använts för att träna en algoritm, för att kunna bedöma hur maskinen kommer att bete sig i ett skarpt läge. Det kan finnas utmaningar med system som byggs i ett land med vissa preferenser, säljs i ett annat och sedan används i ett tredje. Ett exempel är beslutsstöd till vården - kommer rekommendation om lämplig vårdinsats för en patient färgas av värderingar i det land träningsdata är hämtade från?[3]
- Dagens system för maskininlärning är begränsade till väl specificerade problem eftersom systemen ännu saknar känslighet för kontext, till skillnad från människor där redan små barn har en förståelse för sammanhangets betydelse. Beslutsfattare bör vara medvetna om begränsningarna och att maskininlärning lämpar sig bäst för på distinkta och smala problem.[4]

1. www.belfercenter.org/sites/default/files/2019-06/TechFactSheet/machinelearning%20-%204.pdf
2. www.nature.com/articles/s41586-018-0637-6 samt www.moralmachine.net/
3. https://smer.se/wp-content/uploads/2019/06/Smer-konferensrapport_2_webb.NY-REV.pdf, se sidan 11-14
4. https://brie.berkeley.edu/sites/default/files/governing_ai_wp5_99.pdf

Sammanfattning av originalrapporten

Författarna lyfter fram att tillvägagångssättet för att träna modeller (och även utformningen av själva modellerna) vid maskininlärning varierar, beroende på vilken typ av problem man försöker lösa. Inlärningen kan delas upp i tre kategorier:

- Övervakad inlärning, där en människa tränar algoritmen genom att använda data som innehåller ”rätt svar”. Ett exempel är att lära ett e-postprogram att sortera bort skräppost genom att gå igenom inkomna meddelanden och markera om de är skräppost eller inte.
- Oövervakad inlärning, där algoritmen tränas på data som inte innehåller något ”rätt svar” och där det inte heller finns någon människa som ger återkoppling. Här är syftet i stället att algoritmen ska lära sig att känna igen mönster i data och kunna bedöma nya uppgifter, baserat på hur väl nya data passar ihop med något tidigare känt mönster. Ett exempel är att träna algoritmen genom en uppsättning data över villapriser på olika adresser i en stad, för att sedan få ett automatiskt förslag på lämpligt utgångspris när ett nytt hus ska läggas ut till försäljning.
- En kombination av ovanstående, där en människa tränar algoritmen på en liten datamängd (genom att ge återkoppling om rätt och fel svar), varefter algoritmen får fortsätta träningen på egen hand med en större mängd data. Detta har blivit en populär metod eftersom det går åt mindre arbetstid, vilket håller nere kostnaden.

En särskild sorts maskininlärning är så kallad djupinlärning, där inlärningen sker genom positiv återkoppling (belöning) inom ramen för en avgränsad miljö. Djupinlärning har liknats vid det sätt den mänskliga hjärnan bearbetar information från sinnesorganen för att över tid utveckla en färdighet, till exempel att lära sig förstå ett språk. I tekniska applikationer används djupinlärning bland annat inom robotik, navigation och för att lösa komplexa strategispel. Se även faktablad om deepfakes, manipulerade filmer som tas fram med hjälp av djupinlärning (Komet kommenterar 2020:28).

Kort om tekniken

Maskininlärning består av tre huvudsakliga komponenter: en uppsättning data för träning och inlärning, en modell samt en algoritm.*

Vid maskininlärning matas träningsdata in i modellen och algoritmen jobbar sig fram till bästa möjliga lösning för det problem som ska undersökas, till exempel att hitta en tumör i en röntgenbild.

Själva inlärningen är en upprepad process, där algoritmen stegvis förändrar inställningarna i modellen lite i taget. Efter varje steg följer algoritmen upp hur bra utfallet blev – för att sedan gå tillbaka, göra en liten förändring, göra om och undersöka om det gick att bli ännu lite bättre.

Tanken är att träna på kända data, för att sedan kunna använda modellen i nya situationer. Om modellen först lärt sig att känna igen tumörvävnad genom att gå igenom data från ett stort antal röntgenbilder (där det finns ett korrekt svar efter bedömning av läkare), så ska den sedan kunna identifiera en tumör i en helt ny bild som ett stöd vid diagnostik av cancer.

**) Algoritm är enligt Nationalencyklopedin en systematisk procedur som i ett ändligt antal steg anger hur man utför en beräkning eller löser ett givet problem.*

Om Komet Kommenterar
Komet kommenterar aktuella internationella rapporter som rör regelverk, teknikutveckling och innovation.
Syftet är att ge ett svenskt perspektiv, sätta information i ett sammanhang och göra underlaget lätt tillgängligt.

Kommittén för teknologisk innovation och etik (Komet) arbetar för att skapa goda förutsättningar för innovation och konkurrenskraft samtidigt som utvecklingen och spridningen av ny teknik sker tryggt, säkert och med ett långsiktigt samhällsperspektiv.

Detta informationsblad är sammanställt av kommitténs sekretariat.
Läs gärna mer på vår webb: kometinfo.se
e-post: n.komet@regeringskansliet.se