

MVE420: Maskininlärning och etik

Vilhelm Verendel
2018-04-20

Låt säga att du just nu skriver ett inlägg på Facebook

- Tror du att systemet omedelbart delar detta, längst upp i flödet, med alla dina vänner?

Forskning från 2013

Karrie Karahalios på University of Illinois genomförde en undersökning¹ om Facebooks's nyhetsalgoritm (n=40), och upptäckte att

- 62% av deltagarna var omedvetna om att Facebook påverkar nyhetsflödet
- De trodde att systemet omedelbart delar ut allt de postar med sina vänner

Vad tror ni händer istället?

¹Victor Luckerson, *Here's How Facebook's News Feed Actually Works*, 2015, <http://time.com/collection-post/3950525/facebook-news-feed-algorithm/>

Hur skulle ni göra själva?

Hur skulle ni göra om ni fick programmera en liknande algoritm?

Hur skulle ni göra själva?

Hur skulle ni göra om ni fick programmera en liknande algoritm?

- Visa allt för alla i slumpmässig ordning?
- Presentera allt i tidsordning?
- Någon annan strategi?

Hur skulle ni göra själva?

Hur skulle ni göra om ni fick programmera en liknande algoritm?

- Visa allt för alla i slumpmässig ordning?
- Presentera allt i tidsordning?
- Någon annan strategi?

Ett alternativ till att själv programmera detaljerna är att använda *maskininlärning*

Exempel på maskininlärning

- Problemet kallas sedan länge för *binär klassificering* och förekommer även inom maskininlärning: Här handlar det om att välja en funktion som till en vektor av attribut x beskriver relationen:

$$f(x) \rightarrow \{ \text{"intressant"}, \text{"ointressant"} \}$$

Exempel på maskininlärning

- Problemet kallas sedan länge för *binär klassificering* och förekommer även inom maskininlärning: Här handlar det om att välja en funktion som till en vektor av attribut x beskriver relationen:

$$f(x) \rightarrow \{ \text{"intressant"}, \text{"ointressant"} \}$$

- Några möjligheter: Klassificera som "intressant" för dem som
 - Gillar eller kommenterar dina inlägg
 - Själva har skrivit liknande inlägg
 - Liknar dig på andra sätt
 - Har en nära relation till dig
 - ...
 - Kanske en komplicerad blandning av dessa saker?

Exempel på maskininlärning

- Problemet kallas sedan länge för *binär klassificering* och förekommer även inom maskininlärning: Här handlar det om att välja en funktion som till en vektor av attribut x beskriver relationen:

$$f(x) \rightarrow \{ \text{"intressant"}, \text{"ointressant"} \}$$

- Några möjligheter: Klassificera som "intressant" för dem som
 - Gillar eller kommenterar dina inlägg
 - Själva har skrivit liknande inlägg
 - Liknar dig på andra sätt
 - Har en nära relation till dig
 - ...
 - Kanske en komplicerad blandning av dessa saker?
- Att automatiskt söka och välja ut sådana blandningar av bidragande faktorer är en vanlig fråga inom maskininlärning

Programmering kontra maskininlärning

Att välja kriterier för hand är det klassiska sättet att utforma algoritmer som klassificerar: Låt programmerarna välja vad som verkar bra och systemet sätts sedan i produktion, och utvärdering sker gradvis

Programmering kontra maskininlärning

Att välja kriterier för hand är det klassiska sättet att utforma algoritmer som klassificerar: Låt programmerarna välja vad som verkar bra och systemet sätts sedan i produktion, och utvärdering sker gradvis

- Big data-ekonomin leder till nya utmaningar: Stora högdimensionella datamängder som beskriver snabba förändringar kommer in i hög takt

Programmering kontra maskininlärning

Att välja kriterier för hand är det klassiska sättet att utforma algoritmer som klassificerar: Låt programmerarna välja vad som verkar bra och systemet sätts sedan i produktion, och utvärdering sker gradvis

- Big data-ekonomin leder till nya utmaningar: Stora högdimensionella datamängder som beskriver snabba förändringar kommer in i hög takt
- Vi behöver metoder för att snabbt och automatiskt hitta mönster i denna flod av data: Programmerarens roll blir att sätta upp detta?

Programmering kontra maskininlärning

Att välja kriterier för hand är det klassiska sättet att utforma algoritmer som klassificerar: Låt programmerarna välja vad som verkar bra och systemet sätts sedan i produktion, och utvärdering sker gradvis

- Big data-ekonomin leder till nya utmaningar: Stora högdimensionella datamängder som beskriver snabba förändringar kommer in i hög takt
- Vi behöver metoder för att snabbt och automatiskt hitta mönster i denna flod av data: Programmerarens roll blir att sätta upp detta?
- Sådana metoder förekommer ofta inom maskininlärning

Two categories of machine learning

We think of a machine getting some sequence of input data x_1, x_2, x_3, \dots : It could be the values for pixels in images, numbers representing sound waves, text, or some other kind of information, such as products in a shopping cart or events on social media.

Två kategorier av maskininlärning

Vi tänker oss att en maskin får någon sekvens av indata x_1, x_2, x_3, \dots : Det skulle kunna vara värden för pixlar i bilder, tal som representerar ljudvågor, text, eller någon annan sorts information, så som produkter i en varukorg eller händelser på sociala medier.

Det passar ofta att dela upp maskininlärning i två kategorier:

Två kategorier av maskininlärning

Vi tänker oss att en maskin får någon sekvens av indata x_1, x_2, x_3, \dots : Det skulle kunna vara värden för pixlar i bilder, tal som representerar ljudvågor, text, eller någon annan sorts information, så som produkter i en varukorg eller händelser på sociala medier.

Det passar ofta att dela upp maskininlärning i två kategorier:

- 1 *Övervakat lärande*: Maskinen får även ett antal önskade utdata y_1, y_2, y_3, \dots kopplat till indata: Målet för maskinen är att hitta en representation av sambandet $y_i = f(x_i)$ som ger rätt utdata för nya indata

Två kategorier av maskininlärning

Vi tänker oss att en maskin får någon sekvens av indata x_1, x_2, x_3, \dots : Det skulle kunna vara värden för pixlar i bilder, tal som representerar ljudvågor, text, eller någon annan sorts information, så som produkter i en varukorg eller händelser på sociala medier.

Det passar ofta att dela upp maskininlärning i två kategorier:

- 1 *Övervakat lärande*: Maskinen får även ett antal önskade utdata y_1, y_2, y_3, \dots kopplat till indata: Målet för maskinen är att hitta en representation av sambandet $y_i = f(x_i)$ som ger rätt utdata för nya indata
- 2 *Oövervakat lärande*: Systemet får ingen information om vad som är rätt respons (om det ens är meningsfullt): Men maskinen kan ändå tjäna på att hitta mönster i data för att ha en (i) Effektiv representation av indata, (ii) Prediktera framtida indata, (iii) Gruppera och summera indata.

Målet med den här föreläsningen

Relatera maskininlärning till etik

- 1 Förklara vad maskininlärning är och illustrera några av de tekniska möjligheterna, men också några potentiella problem
- 2 Introducera begreppet “moraliskt dilemma”
- 3 Diskutera några etiska problem relaterat till maskininlärning sett som moraliska dilemman

Maskininlärning: Övervakat lärande

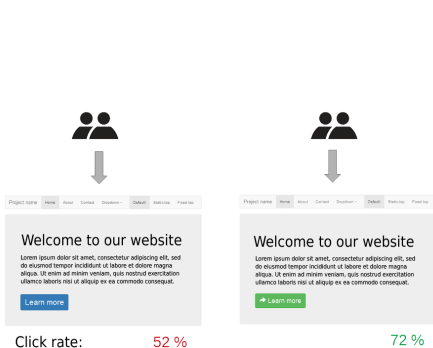
- Den mest välstuderade formen av maskininlärning: Observationer uppmärkta med resultat (x_i, y_i) möjliggör sökning efter samband:

$$y_i = f(x_i)$$

Vi känner inte till f , men söker uppskatta f med en funktion \hat{f}

- Övervakad maskininlärning kan ses som många metoder för att hitta bra, men ibland även mycket komplicerade \hat{f} ...
- Vad är en relevant funktion att uppskatta? Beror på tillämpningen:
 - Visa meddelanden (x_i) relaterat till längden på användarens uppmärksamhet (y_i) som proxy för intresse ...
 - Visa annonser (x_i) som har en hög chans att få klick (y_i) ...
 - ...

Enkelt eksempel: Design av en websida



Enkelt exempel: Annonsplassering

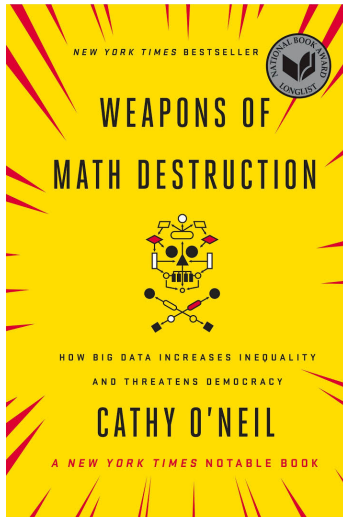
Websidor med hög besöksfrekvens kan göra små variationer i annonser automatiskt, t ex vad gäller

- Färg
- Storlek på teckensnitt
- Andra parametrar så som motiv, beskrivning...

Det stora antalet besökare gör att

- Annonser kan provas snabbt på tusentals personer
- Det är rimligt att förlita sig på statistiska metoder för att välja form av annons
- Gradvis kan den vinnande annonsen väljas ut, eventuellt efter anpassning till användarprofilen

Bok som problematiserar bilden



Weapons of Math Destruction: Förklaring

The math-powered applications powering the data economy were based on choices made by fallible human beings. Some of these choices were no doubt made with the best intentions. Nevertheless, many of these models encoded human prejudice, misunderstanding, and bias into the software systems that increasingly managed our lives. Like gods, these mathematical models were opaque, their workings invisible to all but the highest priests in their domain: mathematicians and computer scientists. Their verdicts, even when wrong or harmful, were beyond dispute or appeal. And they tended to punish the poor and the oppressed in our society, while making the rich richer.

I came up with a name for these harmful kinds of models: Weapons of Math Destruction, or WMDs for short.

Facebook och maskininlärning: Cathy O'Neil skriver

As you know by now, I am outraged by all sorts of WMDs. So let's imagine that I decide to launch a campaign for tougher regulations on them, and I post a petition on my Facebook page. Which of my friends will see it on their news feed?

I have no idea. As soon as I hit send, that petition belongs to Facebook, and the social network's algorithm makes a judgment about how to best use it. It calculates the odds that it will appeal to each of my friends. Some of them, it knows, often sign petitions, and perhaps share them with their own networks. Others tend to scroll right past. At the same time, a number of my friends pay more attention to me and tend to click the articles I post. The Facebook algorithm takes all of this into account as it decides who will see my petition. For many of my friends, it will be buried so low on their news feed that they'll never see it.

En aktuell fråga

Cambridge Analytica

Far more than 87m Facebook users had data compromised, MPs told

Former Cambridge Analytica employee gives evidence before parliamentary committee



▲ Brittany Kaiser said Cambridge Analytica had a suite of personality quizzes designed to extract personal data from the social network. Photograph: HANDOUT/Reuters

Far more than 87 million people may have had their Facebook data harvested by Cambridge Analytica, according to evidence from former employee Brittany Kaiser.

Speaking to the Commons digital, culture, media and sport select committee, Kaiser said Cambridge Analytica had a suite of personality quizzes designed to extract personal data from the social network, of which Aleksandr Kogan's This Is Your Digital Life app was just one example.

Alex Hern

@alexhern

Tue 17 Apr 2018 13:23 BST



< 715

Advertisement

An advertisement for KOMPLETT.se. It features a yellow border and a background of green grass. At the top left is the KOMPLETT.se logo. In the center, the text "VÅRENS bästa FYND" is written in large, bold, green letters. At the bottom, there is a silhouette of a person running on the grass.

Exempel på där maskininlärning används

O'Neils bok visar bland annat hur algoritmer och maskininlärning i ökande grad används inom

- Reklam: Många sidor på Internet
- Politiska kampanjer: Anpassar budskapet efter mottagare
- Arbetsmarknaden: Förstärker ojämlikheter
- Finansiella beslut, Försäkringsbolag, Brottsbekämpande myndigheter...

O'Neil menar: Maskininlärning ger oss fördelar, men även många nya problem (O'Neils bok fokuserar på denna sida av saken)

Relevant för den här kursen

Kunna se och diskutera några exempel på nya problem som uppstår baserat på hur maskininlärning kan tillämpas ute i samhället

Exempel: Privata preferenser²

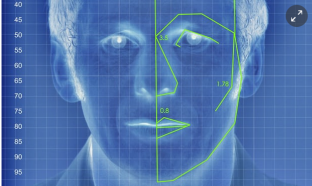
UK world sport football opinion culture business lifestyle fashion environment tech travel

home > tech

Artificial intelligence (AI) New AI can guess whether you're gay or straight from a photograph

An algorithm deduced the sexuality of people on a dating site with up to 91% accuracy, raising tricky ethical questions

40
45
50
55
60
65
70
75
80
85
90
95



© An illustrated depiction of facial analysis technology similar to that used in the experiment. Illustration: Alamy

Artificial intelligence can accurately guess whether people are gay or straight based on photos of their faces, according to new research that suggests machines can have significantly better "gaydar" than humans.

The study from Stanford University - which found that a computer algorithm could correctly distinguish between gay and straight men 81% of the time, and 74% for women - has raised questions about the biological origins of sexual orientation, the ethics of facial-detection technology, and the potential for this kind of software to violate people's privacy or be abused for anti-LGBT purposes.

The machine intelligence tested in the research, which was published in the Journal of Personality and Social Psychology and first reported in the Economist, was based on a sample of more than 35,000 facial images that men and women

f t c ...

This article is 2 months old

< 59,261

Sam Levin in San Francisco

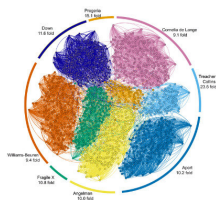
@SamLevin

email

Friday 8 September 2017 00.46 BST

²Deep neural networks are more accurate than humans at detecting sexual orientation from facial images, 2017. <https://osf.io/zn79k/>

Exempel: Ovanliga sjukdomar³



³Diagnostically relevant facial gestalt information from ordinary photos, eLife, Journal of Personality and Social Psychology, 2017.

<https://dx.doi.org/10.7554/eLife.02020>

Exempel: Självkörande bilar



JACK STEWART, TRANSPORTATION 03.30.18 10:34 PM

TESLA'S AUTOPILOT WAS INVOLVED IN ANOTHER DEADLY CAR CRASH



TESLA

TESLA NOW HAS another fatality to hang on its semi-autonomous driving system. The company just revealed that its Autopilot feature was turned on when a Model X SUV slammed into a concrete highway lane divider and burst into flames on the morning of Friday, March 23. The driver, Wei Huang, died shortly afterwards at the hospital.

Moraliska dilemman

Definition på *moraliskt dilemma*: När två eller flera etiska aspekter kommer i konflikt med varandra

Två vanliga aspekter som kommer i konflikt

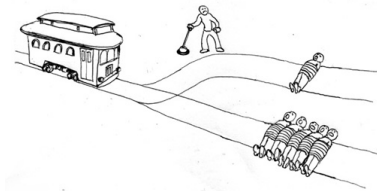
- Den totala nyttan för kollektivet (Utilitarism)
- Individens frihet, plikter och rättigheter (Deontologisk etik)

Moraliska dilemman

Definition på *moraliskt dilemma*: När två eller flera etiska aspekter kommer i konflikt med varandra

Två vanliga aspekter som kommer i konflikt

- Den totala nyttan för kollektivet (Utilitarism)
- Individens frihet, plikter och rättigheter (Deontologisk etik)



Diskussion: Är detta moraliska dilemman?

- 1 Ansiktsigenkänning: Klassificera homosexualitet
- 2 Klassificera ovanliga sjukdomar

Diskutera i grupper om 2 personer i 2 minuter: Finns några etiska aspekter i dessa exempel som skulle kunna komma i konflikt med varandra?

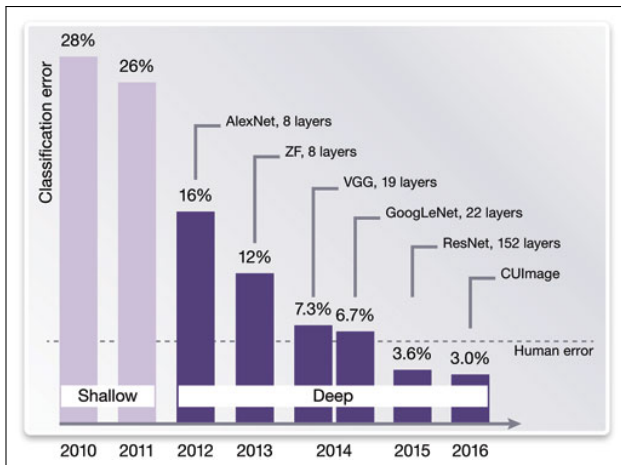
Några områden för maskininlärning

Några teknikområden där maskininlärning i allt ökande grad används:

- 1 Bildanalys och objektklassificering
- 2 Röst- och taligenkänning
- 3 Textanalys (*vec: word2vec, doc2vec, ...)
- 4 Självkörande bilar
- 5 Spel som schack och Go
- 6 ...

Bildanalys och objektklassificering

Resultat på *ImageNet competition*:



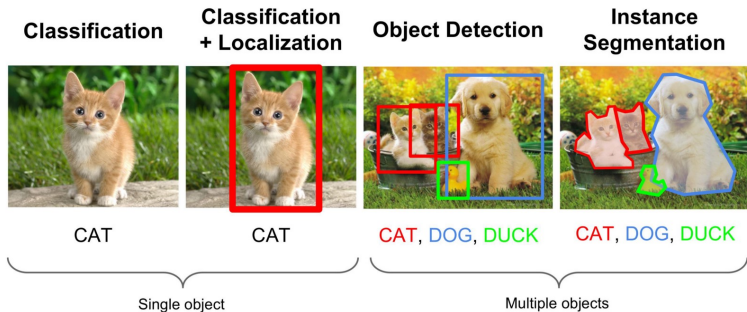
Vad är ImageNet?

Mycket av framgången beror på stora databaser med uppmärkt data, som ImageNet

- Uppmärkt data, 14 miljoner bilder relaterade till 1000 klasser
- Djupa neurala nätverk har slagit alla tidigare metoder
- I många områden har metoderna bättre prestanda än människor

Djupa neurala nätverk fungerar ofta bra med stora mängder data, men är inte alltid bäst

Olika delar av bildanalys



De stora förbättringarna kan förstås som en framgång för djupa neurala nätverk

Vad är Deep Learning?

En modell inspirerad av hjärnan

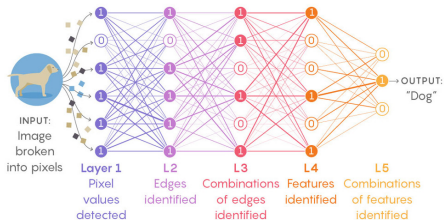


Vad är Deep Learning?

En modell inspirerad av hjärnan



Komplicerad representation: $y_i = f(x_i)$

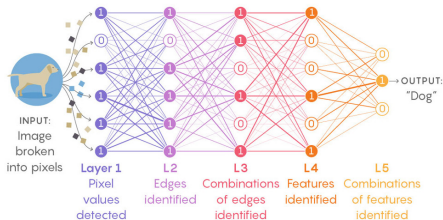


Vad är Deep Learning?

En modell inspirerad av hjärnan



Komplicerad representation: $y_i = f(x_i)$



Modellparametrar:

- Hur de olika lagren är ihopkopplade med vikter och tröskelvärden
- De neurala nätverk som fungerar bra har ofta tiotals/hundratals lager, med miljontals parametrar
- Väljs ut med en optimeringsmetod kallad *back propagation*⁴

⁴www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec3.pdf

Begränsningar med Deep Learning

- En begränsning i deep learning är svårigheten att sammanfatta och förklara resultaten på ett enkelt sätt

Begränsningar med Deep Learning

- En begränsning i deep learning är svårigheten att sammanfatta och förklara resultaten på ett enkelt sätt
- Bör vi lita på metoderna utan några garantier, i säkerhetskritiska system?

Begränsningar med Deep Learning

- En begränsning i deep learning är svårigheten att sammanfatta och förklara resultaten på ett enkelt sätt
- Bör vi lita på metoderna utan några garantier, i säkerhetskritiska system?
- Fooling Image Recognition with Adversarial Examples:

https://www.youtube.com/watch?v=piYnd_wYlT8

Transparens i maskininlärningsmodeller

Deep learning visar även att: Bra modeller kan vara svårbegripliga

Transparens i maskininlärningsmodeller

Deep learning visar även att: Bra modeller kan vara svårbegripliga

Tre skäl till det kan vara svårt att få insyn i modeller

- 1 Tillgång till data: Integritetsmässigt komplicerat med privatpersoners data, eller begränsad tillgång av andra skäl
- 2 Tillgång till modellen: Proprietära system med företagshemligheter
- 3 Komplicerad statistisk modell: Många modeller är tillräckligt komplicerade att programmerarna/matematikerna själva inte förstår varför resultatet fungerar så bra

Storskalig effekt av maskininlärning?



The image is a screenshot of a news article. On the left, there is a social media-style header for 'Cambridge Analytica' with a house icon, the text 'The Cambridge Analytica Files', and author information 'Owen Bowcott and Alex Hern' dated 'Tue 10 Apr 2018 16:45 BST'. Below this are social media sharing icons and a view count of '6,417'. The main article title is 'Facebook and Cambridge Analytica face class action lawsuit'. The sub-headline reads 'Lawyers in UK and US allege four firms misused personal data of more than 71 m people'. A sub-section title says 'Five things we learned from Mark Zuckerberg's Facebook hearing'. The main image shows a person in profile using a laptop and a smartphone, with a large yellow crescent moon in the background and a network diagram overlay. Below the image, the article text states: 'The lawsuit claims the firms obtained Facebook users' private data to develop "political propaganda campaigns" in the UK and the US. British and US lawyers have launched a joint class action against Facebook, Cambridge Analytica and two other companies for allegedly misusing the personal data of more than 71 million people. The lawsuit claims the firms obtained users' private information from the social media network to develop "political propaganda campaigns" in the UK and the US. Facebook, it is said, may initially have been misled, but failed to act responsibly to protect the data of 1 million British users and 70.6 million people in America.'

Cambridge Analytica
The Cambridge Analytica Files

Owen Bowcott and Alex Hern
Tue 10 Apr 2018 16:45 BST

f t e ...

< 6,417

Facebook and Cambridge Analytica face class action lawsuit

Lawyers in UK and US allege four firms misused personal data of more than 71 m people

Five things we learned from Mark Zuckerberg's Facebook hearing



▲ The lawsuit claims the firms obtained Facebook users' private data to develop "political propaganda campaigns" in the UK and the US. Photograph: Getty Images

British and US lawyers have launched a joint class action against Facebook, Cambridge Analytica and two other companies for allegedly misusing the personal data of more than 71 million people.

The lawsuit claims the firms obtained users' private information from the social media network to develop "political propaganda campaigns" in the UK and the US.

Facebook, it is said, may initially have been misled, but failed to act responsibly to protect the data of 1 million British users and 70.6 million people in America.

Denna fråga kan förstås utifrån *både* övervakat och oövervakat lärande ...
Men är Facebook verkligen en politiskt relevant plattform?

Facebook: Politisk maktfaktor?

- Facebook har ihop med forskare själva släppt vissa studier
- Tillgång till över en miljard personer: Vad som är släppt som forskningen är troligtvis toppen på isberget
- En bra fråga är: Skulle Facebook och Google kunna påverka politiken?

Studie 1: Påverka valdeltagande via vänner⁵

Forskningsfråga: Kan Facebook öka valdeltagandet?

- Valen i USA 2010 och 2012: Små förändringar i gränssnittet
Försöket nådde ut till över 61 miljoner amerikaner
- Använde ett verktyg de kallade “Voter megaphone”, där användare klickade i att de röstat
- Metod: Undersök gruppptryckets effekt på att gå och rösta (genom att tala om för andra att man röstat)

Forskarnas uppskattning är att valdeltagandet ökade med 340,000 personer

⁵Bond et al., *A 61-Million-Person Experiment in Social Influence and Political Mobilization*, Nature, 2012

Studie 2: Påverka valdeltagande med nyheter⁶

Forskningsfråga: Kan politiska nyheter från vänner ändra politiskt beteende?

- Valet 2012: Romney vs Obama
- Två miljoner personer fick fler nyhetelänkar jämfört med annat (så som bilder på hundar och katter)
- Metod: Be personer att självrapportera om de röstat eller inte

Forskarnas uppskattning är att valdeltagandet i denna grupp gick upp från 64 till 67 procent

⁶Bond et al., *A 61-Million-Person Experiment in Social Influence and Political Mobilization*, Nature, 2012

Studie 3: Påverkan via sökmotorers ranking⁷

Forskningsfråga: Kan en sökmotors konfiguration påverka hur personer röstar?

- Två Google-forskare (Robert Epstein och Ronald E. Robertson) bad osäkra väljare i USA och Indien att använda en sökmotor för att bilda sig en uppfattning om valet
- Sökmotorerna var specialprogrammerade till att framhäva ett parti framför ett annat

Forskarnas uppskattning är att röstningspreferenserna kunde ändras i runt 20% av fallen

⁷Bakshy et al., *Exposure to Ideologically Diverse News and Opinion*, Science, 2015

Studie 3: Kritik

- Inget talar för att Google faktiskt försöker påverka valutgången
- 73% av amerikander trodde 2012 att sökresultaten är både korrekta och opartiska⁸

⁸Purcell et al., *Search Engine Use 2012*, Pew Research Center, 2012

Inga tecken på att företagen själva använder detta

Cathy O'Neil kallar inte dessa algoritmer för *WMDs*, men menar att potentialen finns för missbruk...

I wouldn't yet call Facebook or Google's algorithms political WMDs, because I have no evidence that the companies are using their networks to cause harm. Still, the potential for abuse is vast. The drama occurs in code and behind imposing firewalls. And as we'll see, these technologies can place each of us into our own cozy political nook.

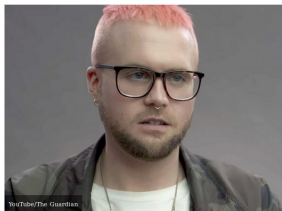
Utesluter inte att andra kan använda plattformen

Storskalig påverkan via Facebook⁹

Here's the personality test Cambridge Analytica had millions of Facebook users take

Follow @EthanGale 7.4k likes 2.4k followers

EthanGale 19 Mar 2018 9:01 PM 1256



- Reports published this weekend from The New York Times and The Observer revealed that personality-profiling company Cambridge Analytica harvested data from millions of users and potentially used it in the most recent US presidential election.
- The personality test that the company used gives users a score called their "OCEAN" score, referring to how it calculates their performance on a measure of Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism.
- The test is available for free online at the University of Cambridge Psychometrics Center and forms the basis of many existing psychological studies on happiness and longevity.

A personality-profiling company called Cambridge Analytica harvested personal data from millions of users and may have used that data to sway voters during the 2016 US presidential election, according to reports published this weekend from The New York Times and The Observer.

⁹<http://nordic.businessinsider.com/facebook-personality-test-cambridge-analytica-data-trump-election-2018-3?>

Psykografiska modeller

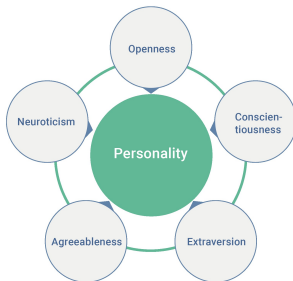
- Av det som framkommit har man använt något de kallat “psykografiska modeller”
- Har sitt ursprung i modeller av människors personlighet
- Upptäcktes ursprungligen med statistiska metoder som inom maskininlärning kan kallas för *oövervakat lärande*

Personlighet

- Personlighet: Stabila dispositioner över situation och tid

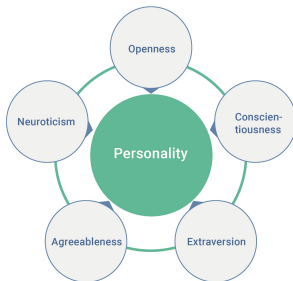
Personlighet

- Personlighet: Stabila dispositioner över situation och tid
- En vanlig modell: OCEAN, delar upp personer i $2^5 = 32$ typer



Personlighet

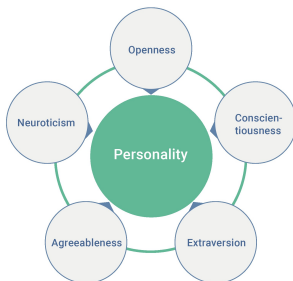
- Personlighet: Stabila dispositioner över situation och tid
- En vanlig modell: OCEAN, delar upp personer i $2^5 = 32$ typer



- Varför dessa faktorer? Kommer från *faktoranalys*, en form av övervakat lärande för att hitta struktur i data

Personlighet

- Personlighet: Stabila dispositioner över situation och tid
- En vanlig modell: OCEAN, delar upp personer i $2^5 = 32$ typer



- Varför dessa faktorer? Kommer från *faktoranalys*, en form av övervakat lärande för att hitta struktur i data
- Tillämpat på stora insamlade datamängder: Frågebatterier om egenskaper hos personer - vilka termer samvarierar?

Tillämpning av faktoranalys

Exempel: En mycket noggrann person kan beskrivas som både 'metodisk', 'fokuserad på detaljer', 'pedantisk', och så vidare...

- Målet med faktoranalys: Reducera dimensionen i en högdimensionell datamängd (många frågor)
- Hypotes: Det finns några få dolda faktorer som vi inte direkt kan observera (personlighet) som beskriver regelbundenheter i observerade data
- Metod: Sök ett mindre antal dolda faktorer som förklarar hur de observerade variablerna varierar/samvarierar

Sådan gruppering av data kan beskrivas som oövervakat lärande och inom fältet personlighetspsykologi har många studier gjorts för att söka efter effekten av dessa faktorer (t ex med personlighetstester)

Kopplingen till övervakat lärande

En möjlighet har varit att:

- Be en mindre population av personer att på olika sätt göra personlighetstester (och klassificera person i med y_i , i termer av deras OCEAN-värden), samla samtidigt in annan information som de har uppgett i sina profiler (x_i)

Kopplingen till övervakat lärande

En möjlighet har varit att:

- Be en mindre population av personer att på olika sätt göra personlighetstester (och klassificera person i med y_i , i termer av deras OCEAN-värden), samla samtidigt in annan information som de har uppgett i sina profiler (x_i)
- Använd övervakat lärande för att uppskatta $y_i = f(x_i)$

Kopplingen till övervakat lärande

En möjlighet har varit att:

- Be en mindre population av personer att på olika sätt göra personlighetstester (och klassificera person i med y_i , i termer av deras OCEAN-värden), samla samtidigt in annan information som de har uppgett i sina profiler (x_i)
- Använd övervakat lärande för att uppskatta $y_i = f(x_i)$
- Samla in likadana x för en mycket större population (miljoner personer, utan direkt vetskap)

Kopplingen till övervakat lärande

En möjlighet har varit att:

- Be en mindre population av personer att på olika sätt göra personlighetstester (och klassificera person i med y_i , i termer av deras OCEAN-värden), samla samtidigt in annan information som de har uppgett i sina profiler (x_i)
- Använd övervakat lärande för att uppskatta $y_i = f(x_i)$
- Samla in likadana x för en mycket större population (miljoner personer, utan direkt vetskap)
- Tillämpa $f(x)$ för varje individ, för att välja politisk information som har framgång hos en viss personlighetstyp...

Cathy O'Neil reflektarar

In *The Selling of the President*, which followed Richard Nixon's 1968 campaign, the journalist Joe McGinniss introduced readers to the political operatives working to market the presidential candidate like a consumer good. By using focus groups, Nixon's campaign was able to hone his pitch for different regions and demographics.

...

The convergence of Big Data and consumer marketing now provides politicians with far more powerful tools. They can target microgroups of citizens for both votes and money and appeal to each of them with a meticulously honed message, one that no one else is likely to see. It might be a banner on Facebook or a fund-raising email.

Cathy O'Neil reflekterar (forts.)

...

But each one allows candidates to quietly sell multiple versions of themselves and it's anyone's guess which version will show up for work after inauguration.

Slutsatser

- Vi lever redan nu i ett samhälle där maskininlärning kommer användas för att snabbt hantera stora datamängder: Ersätter delvis den traditionella programmaren för att hantera och upptäcka mönster
- Individer, företag, men också myndigheter ställs inför nya moraliska dilemman i takt med att metoder från maskininlärning blir allt kraftfullare
- Moraliska dilemman: Olika etiska mål kommer i konflikt
- Maskininlärning kommer spela en roll i många framtida moraliska dilemman, delvis på grund av att det kan vara svårt att förenkla vad som händer i modellen:
 - 1 Då modellen är hemlig/eller datan proprietär
 - 2 Då modellen är så komplicerad att programmarna/matematikerna själva inte förstår den

I många fall finns redan (i) och (ii) samtidigt

Vad bör ni ha lärt er?

- Vad maskininlärning är: Två former av maskininlärning
- Definition på moraliskt dilemma
- Känna till några nya moraliska dilemman relaterade till användandet av maskininlärning