

Inleiding Kunstmatige Intelligentie

Leerwijzer voor eerstejaars Bachelorstudenten
Kunstmatige Intelligentie



UNIVERSITEIT VAN AMSTERDAM

FNWI
2020/2021
Laatste aanpassing: 25 november 2020

5082INKI6Y

Cursuscoördinator:

Dr. Arnoud Visser

Werkcollege Docenten:

Wessel de Jong, Msc.

Max van den Broek, Msc.

Voor vragen over de colleges kun je mailen naar:
a.visser@uva.nl

...

Voor vragen over de werkcolleges kun je mailen naar:
w.dejong@uva.nl

...

Inhoud van het vak

In deze cursus wordt een overzicht gegeven van het vakgebied van de Kunstmatige Intelligentie. We behandelen daarbij de geschiedenis van het vakgebied, de ontwikkeling in de laatste decennia, en de moderne deelgebieden van de Kunstmatige Intelligentie. Daarbij gaan we vooral in op de theoretische ontwikkelingen die belangrijk zijn voor het onderzoek en de toepassingen van vandaag. De beschouwing over de toekomst wordt door jullie zelf vormgegeven in een uitvoerige eindopdracht. Na het volgen van het vak kunnen je de volgende vragen op een academisch niveau beantwoorden:

- Wat zijn de vragen die onderzoek naar Kunstmatige Intelligentie probeert te beantwoorden?
- Welke ontdekkingen zijn daaruit voortgekomen?
- Welke toepassingen heeft dat opgeleverd?
- Welke problemen zijn nog onopgelost?
- Wat zijn de grote uitdagingen voor de nabije en verre toekomst?

Opzet van het vak

Het vak bestaat uit hoorcolleges en werkcolleges, waarbij de hoorcolleges vooral bestaan uit gastcolleges van onderzoekers die een onderzoeksveld binnen Kunstmatige Intelligentie (KI) bespreken. De werkcolleges dienen ter verdieping van de hoorcolleges; hierin worden de corresponderende artikelen besproken. In het eerste gedeelte van het vak (week 1 t/m 3) zijn de student-assistenten die richting geven aan de interpretatie en bespreking, in het tweede gedeelte (week 5 t/m 7) zijn jullie zelf leidend. Verder dienen de werkcolleges ter ondersteuning van de huiswerkopdrachten en schrijfopdracht (zie 'opdrachten').

Perusall

De artikelen die bij de hoor- en werkcolleges horen worden gezamenlijk gelezen en geïnterpreteerd met behulp van PerusAll.

In PerusAll plaatsen jullie annotaties; dit kunnen vragen zijn, antwoorden op anderen hun vragen of gewoon interessante opmerkingen. Bij iedere tekst moeten jullie minimaal 7 annotaties plaatsen. Jullie annotaties worden automatisch becijferd door PerusAll op basis van uitgebreidheid, structuur en distributie over de tekst (zodat studenten de hele tekst moeten lezen). Hoe PerusAll het cijfer precies bepaalt is geheim, als jullie een goed cijfer willen hebben moeten jullie gewoon focussen op hele goede annotaties plaatsen.

Onderzoek tijdens het vak

Tijdens het vak wordt onderzoek gedaan naar de ontwikkeling van een dashboard met cijferinformatie. In dat kader krijg je een vragenlijst aan het begin en eind van het vak voorgelegd. Ook zijn er enkele korte vragen aansluitend bij opdrachten. Het beantwoorden van deze vragen is een verplicht onderdeel van het vak. Deelname aan het onderzoek staat daar los van. Aan het begin van het vak zal je door de onderzoekers worden gevraagd of jouw geanonimiseerde data ook gebruikt kan worden voor onderzoek.

Opdrachten

Quiz week 1: Korte quiz (multiple choice) over de stof en artikelen van week 1.

Huiswerk 1: Korte vragen beantwoorden over Russell & Norvig, Turing, Newell & Simon en Brooks (individueel & schriftelijk).

Huiswerk 2: Korte vragen beantwoorden over machine learning, causal inference & symbolische vs subsymbolische AI (individueel & schriftelijk).

Korte presentatie & bespreking: Presentatie (+/- 5 min.) over het paper dat hoort bij het gastcollege. Daarna het paper bespreken met de werkgroep middels een zelfbedachte werkvorm (b.v.: debat, quiz, tekstvragen maken & bespreken, andere opdracht, ...). In totaal ± 30 minuten. Groepsopdracht.

Schrijfoopdracht: Kies een paper behorend bij een gastcollege uit deel II (en niet het paper dat je al hebt gepresenteerd). De belangrijkste punten uit dit paper vat je samen, en vervolgens relateer je het paper aan drie benaderingen/concepten uit deel I van het vak. Individuele opdracht.

Posteropdracht - PAV: Zie het corresponderende gedeelte op Canvas.

Inhoudsopgave

Week 1	4
Werkgroep 1: Introductie Kunstmatige Intelligentie	5
Werkgroep 2: De Turing Test	7
Week 2	9
Werkgroep 3: Symbolische KI	10
Werkgroep 4: Machine Learning	12
Week 3	14
Werkgroep 5: Causal Reasoning	15
Werkgroep 6: Symbolisme en Connectionisme	18
Week 4	20
Deeltentamen	20
Week 5	21
Werkgroep 7 - I: Computer Vision	22
Werkgroep 7 - II: Information Retrieval	24
Week 6	27
Werkgroep 8 - I: Natural Language Processing	28
Werkgroep 8 - II: Data Driven Decision Making	30
Week 7	32
Werkgroep 9 - I: Social Robotics	33
Werkgroep 9 - II: Big Data and Implications on Society	35
Posterpresentaties	38

Week 1

Literatuur

- Russell, S.J. and P. Norvig (2010), [Introduction](#). In: Artificial Intelligence, A Modern Approach (3rd edition). Pearson - Prentice Hall, Upper Saddle River, New Jersey.
- Turing, A. (1950), '[Computing Machinery and Intelligence](#)', Mind, 59 (236): 433-460.

Docent



Dr. [Arnoud Visser](#) is docent Robotica bij de Universiteit van Amsterdam (UvA). Hij behaalde zijn doctoraal Experimentele Natuurkunde bij de Rijkuniversiteit Leiden, bij Prof. van der Waals.

Hij is de oprichter van het [Intelligent Robotics Lab](#) van de UvA.

Hoorcollege 0: de opzet van het vak

- [Opname op Canvas](#)

Hoorcollege 1: de grondslagen van KI

- [Opname op Canvas](#).

Hoorcollege 2: de geschiedenis van KI

- [Opname van webcollege](#)

Werkgroep 1: Introductie Kunstmatige Intelligentie

Instructies

Gebruik deze leerwijzer voor het voorbereiden van de college-dag. Lees de opgegeven artikelen en gebruik tijdens het lezen de conceptenlijst en eventueel bijgevoegde afbeeldingen om een overzicht te krijgen van de artikelen.

Formuleer tijdens het lezen vragen over de stof en stel deze vragen in PerusAll. Markeer het stuk tekst waar je vraag over gaat en stel je vraag zo duidelijk en uitgebreid mogelijk, zodat je TA en medestudenten antwoord kunnen geven.

Topics

- What is AI?
- Thinking Humanly
- Thinking Rationally
- Acting Humanly
- Acting Rationally
- History of AI

Literatuur

Russell, S.J. and P. Norvig (2010), **Introduction** In: Artificial Intelligence, A Modern Approach (3rd edition). Pearson - Prentice Hall, Upper Saddle River, New Jersey.

Kernpunten

Hoofdstuk 1 van het boek Russell en Norvig geeft een introductie van de wetenschap Kunstmatige Intelligentie. Focus je bij het lezen vooral op het onderscheid tussen de vier verschillende benaderingen van AI (Thinking Humanly, Thinking Rationally, Acting Humanly and Acting Rationally). Zorg ervoor dat je goed weet waar deze begrippen voor staan en dat je het onderscheid tussen de benaderingen uit kunt leggen. Probeer tijdens het lezen de verschillende voorbeelden van onderzoek binnen de Kunstmatige Intelligentie te plaatsen binnen een van de benaderingen. Welke benadering kiezen de auteurs?

Lees daarbij ook aandachtig de paragraaf (1.3) over de geschiedenis van de Kunstmatige Intelligentie. Deze paragraaf geeft je een goed beeld van de grondslagen van het wetenschapsveld dat je de komende jaren zult bestuderen.

Concepten:

Artificial Intelligence (AI)

Many definitions of this scientific field have been given, most generally it can be stated that AI is the scientific field that tries to understand intelligence and tries to build intelligent entities.

Rationality

A system or entity is rational if it does the ‘right thing’, given what it knows.

Acting Humanly

The approach to AI that aims to build intelligent entities capable of showing human behavior. This approach is mainly concerned with the *behavior* of intelligent entities. In this approach we have successfully build and intelligent entity if it shows behavior equal to human beings.

Thinking Humanly

The approach to AI that aims to build intelligent entities capable of thinking like humans. This approach is mainly concerned with the *thought processes* and *reasoning* of intelligent entities. Instead of looking at behavior this approach considers an entity intelligent if it *thinks* like a human.

Thinking Rationally

The approach to AI that aims to build intelligent entities capable of thinking rationally. This approach also focuses on the *thought processes* and *reasoning* of intelligent entities. An entity is considered intelligent when it can reason according to specific logical rules.

Acting Rationally

The approach to AI that aims to build intelligent entities capable of showing rational behavior. In this approach the focus is again on *behavior* and an entity is considered intelligent when it shows behavior that can be considered rational.

Werkgroep 2: De Turing Test

Topics

- Can Machines Think?
- The Imitation Game/Turing Test
- Objections to the Turing Test

Literatuur

Turing, A. (1950), 'Computing Machinery and Intelligence', *Mind*, 59 (236): 433-460.

Kernpunten

In dit originele artikel Turing wordt de Turing test geïntroduceerd. Het is in eerste instantie belangrijk dat je kunt uitleggen hoe de Turing test werkt en welke vraag deze test probeert te beantwoorden. Richt je daarna op de verschillende voor- en nadelen van de test die Turing in paragraaf 2 bespreekt. Van de paragrafen die volgen, waarin Turing de machines bespreekt die worden gebruikt in de Turing test, is het vooral belangrijk dat je begrijpt wat een *discrete state machine* is en hoe deze globaal werkt. Vervolgens bespreekt Turing meerdere tegenwerpingen tegen zijn test. Het is belangrijk dat je de tegenwerpingen kent en weet hoe Turing die weer weerlegt.

Concepten:

The Imitation Game

The Imitation Game

Also called the Turing Test. Test proposed by Alan Turing (1950) that is supposed to provide a satisfactory definition of intelligence; if an entity/computer/program is capable of passing the test, it should be considered intelligent. A computer passes the test if a human interrogator, after posing some written questions, cannot tell whether the written responses come from a person or from a computer.

Total Turing Test

Form of the Turing test in which physical capabilities of a computer, such as perception and movement, are also tested.

Discrete-state Machine

Machines that operate by moving from one discrete state to another.

Digital Computer

See Turing (1950) paragraph 4. As Turing outlines: a digital computer is a discrete-state machine intended to carry out any operations which could be done by a human computer. The machine uses fixed rules. It usually consists out of the parts store, executive unit and control.

Objections to the Imitation Game

The Theological Objection

Machines cannot think since thinking is a function of man's immortal soul and God has only given such a soul to humans.

The "Heads in the Sand" Objection

Since the consequences of machines being able to think are too severe, it is best to believe that machines actually are not capable of thinking.

The Mathematical Objection

It has mathematically been proven that discrete-state machines are too limited to be capable of thinking. For example, it has been shown that such a machine will always be unable to answer a specific set of questions and it can therefore never pass the Turing Test (since asking those questions will always reveal the machine to be a machine).

The Argument from Consciousness

Machines are not capable of thinking because they lack consciousness.

Solipsism

The belief that only one's own mind is sure to exist. It is impossible to know if another person has a mind/consciousness, because the only way to know this is by being this other person.

Arguments from Various Disabilities

Machines will never be capable of thinking/passing the Turing Test because there will always be a set of abilities a machine cannot perform. For example, a machine will never be able to be kind, beautiful, friendly, fall in love or have a sense of humour.

Lady Lovelace's Objection

A machine is not capable of thinking since it will never be capable of doing something that it is not ordered to do. It will never surprise us or do something we would not have expected.

Argument from Continuity in the Nervous System

Machines are not capable of thinking because our nervous system certainly is not a discrete-state machine. Machines can therefore never mimic the behaviour of the nervous system.

The Argument from Informality of Behavior

This argument states that it is impossible to form a set of rules that captures human behavior in any conceivable situation. Therefore machines can never show the same behavior as humans and can never pass the Turing Test.

The Argument from Extrasensory Perception

If something like extrasensory perception exists, the Turing Test could not work since such powers could interfere with the test.

Week 2

Literatuur

- Newell, A., H.A. Simon (1976), 'Computer science as empirical inquiry: symbols and search', Communications of the ACM, 19(3): 113-126
- Brooks, R.A. (1991), 'Intelligence without representation', Artificial Intelligence, 47(1-3): 139-159
- Jordan, M.I., T.M. Mitchell (2015) 'Machine learning: Trends, perspectives, and prospects', Science, 349 (6245): 255-260.

Hoorcollege 3: de wetenschappelijke basis van KI

- [Opname van webcollege](#)

Hoorcollege 4: Machine Learning

Docent



Prof. dr. Max Welling heeft zowel een onderzoekspositie bij de Universiteit van Amsterdam (UvA), University of California Irvine (UCI) en de Canadian Institute for Advanced Research (CIFAR), Verder is hij mede-oprichter van het bedrijf Scyfer BV.

Van oorsprong is Prof. Welling van theoretisch natuurkundige, hij schreef zijn proefschrift bij de Nobel prijswinnaar Prof. 't Hooft.

Hij is de mede-oprichter van het Nederlandse Innovation Center for AI en het European Lab for Learning and Intelligent Systems (ELLIS). Hij is ook één van de initiatiefnemers van de Nationale AI Cursus en bekend van optredens bij de DWDD University en Science & Cocktails @ Paridiso.

- [Opname van webcollege](#)

Werkgroep 3: Symbolische KI

Topics

- Artificial Intelligence as an Empirical Science
- Physical Symbol System
- Physical Symbol System Hypothesis

Paper

Newell, A., H.A. Simon (1976), 'Computer science as empirical inquiry: symbols and search', Communications of the ACM, 19(3): 113-126

Brooks, R.A. (1991), 'Intelligence without representation', Artificial Intelligence, 47(1-3): 139-159

Kernpunten

Newell en Simon proberen in hun artikel Newell en Simon aan te tonen dat Kunstmatige Intelligentie een empirische wetenschap is. Zij doen dit door twee voorbeelden te bespreken die volgens hen van groot belang zijn voor de totstandkoming van intelligentie. Wij richten ons vooral op het eerste voorbeeld over *physical symbol systems*.

Volgens Newell en Simon bestuderen we in de Kunstmatige Intelligentie *physical symbol systems*. Probeer goed te begrijpen wat zulke *physical symbol systems* precies zijn en wat wordt bedoeld met de symbolen die deze systemen manipuleren. Vraag jezelf af waarom Newell en Simon specifiek over 'fysieke' systemen spreken en welke rol symbolen in deze systemen spelen.

Na enige uitleg over *physical symbol systems* presenteren Newell en Simon de *Physical Symbol System Hypothesis* (PSSH). Dit is een belangrijke hypothese die we nog vaak zullen tegenkomen, zorg ervoor dat je hem goed kent. Vervolgens wordt beargumenteerd dat de PSSH een empirische hypothese is door het ontstaan van de PSSH te bespreken. Hier is het belangrijk de lijn van deze argumentatie te kunnen reproduceren. Ten slotte presenteren Newell en Simon bewijs voor de PSSH. Probeer goed het onderscheid te kennen tussen het bewijs voor het noodzakelijke (necessary) deel van de hypothese en het voldoende (sufficient) deel van de hypothese.

Brooks kijkt in het artikel Brooks heel anders tegen de totstandkoming van intelligentie aan. Volgens Brooks is representatie (in de vorm van symbolen) niet noodzakelijk voor het creëren van intelligente systemen. Hij stelt een andere benadering voor waarin we ons richten op het incrementeel opbouwen van systemen zonder interne representatie ('use the world as its own model'). Bij het opbouwen van zulke intelligente systemen moeten we ervoor zorgen dat we bij elke stap een compleet systeem bouwen, in tegenstelling tot het bouwen van losse aspecten van intelligentie (e.g. perceptie, taal etc.) die later worden samengevoegd. Daarbij moeten telkens complete systemen als een geheel getest worden in de echte wereld, ipv een abstractie daarvan.

In sectie 2 en 3 probeert Brooks te beargumenteren wat er mis is met de ontwikkeling van intelligente systemen zoals onder andere voorgesteld door Newell

en Simon. Zorg ervoor dat je Brooks' argumentatie over wat er mis is met de wijze waarop abstractie binnen KI gebruikt wordt goed begrijpt. Vervolgens richt Brooks zich in sectie 4, 5 en 6 op het presenteren en onderbouwen van zijn alternatief. Het is belangrijk goed te begrijpen wat hij bedoelt met zijn *Creatures* (sectie 4) en hoe deze ontwikkeld moeten worden (sectie 6.1). Zorg er daarbij ook voor dat je de voordelen van deze creatures, zoals benoemd in sectie 5, kunt reproduceren. Zorg er ten slotte voor dat je goed begrijpt hoe Brooks' voorstel zich onderscheidt van andere benaderingen (vooral secties 7.1 en 7.2) en wat eventuele beperkingen zijn (sectie 8).

Concepten:

Newell & Simon:

Empiricism

The philosophical theory that states that all knowledge originates from sensory experience.

Empirical Science

A science in which knowledge is mainly acquired through sensory experience, for example by observing behavior within an experiment.

Symbol

A physical pattern that represents an object in the world (e.g a table or a chair, but also an idea or a word). Multiple symbols can be combined to form (complex) expressions.

Physical Symbol System

A system that consists out of a set of symbols which it can manipulate to form many kinds of symbol structures.

Physical Symbol System Hypothesis

A physical symbol system has the necessary and sufficient means for general intelligent action. This means that every entity that shows general intelligent action has to be a physical symbol system (it is necessary) and that every physical symbol system has the means to show general intelligent action (it is sufficient).

Brooks:

Abstraction

A simplified representation of the world by only representing the pertinent and (according to humans) meaningful facts. A well-known example within AI is the 'Block world' in which many experiments have been run.

Creature

Autonomous mobile agents which are seen by humans as intelligent beings and are capable of co-existing with humans.

Layer

An activity producing sub part of an creature (or intelligent agent) that connects sensing to direct action. Within a creature multiple layers function in parallel.

Werkgroep 4: Machine Learning

Topics

- Introduction to Machine Learning

Literatuur

Jordan, M.I., T.M. Mitchell (2015) 'Machine learning: Trends, perspectives, and prospects', Science, 349 (6245): 255-260.

Kernpunten

In dit artikel geeft Jordan en Mitchell een overzicht van het veld Machine Learning in 2015. Ondertussen is er in de Machine Learning alweer veel vooruitgang geboekt, het artikel is dus niet helemaal up-to-date. Toch wordt er een mooi overzicht gegeven van de basisconcepten van Machine Learning. Het is bij dit artikel dan ook vooral belangrijk deze concepten goed te kennen en te begrijpen hoe deze zich tot elkaar verhouden. Bestudeer daarvoor ook goed de onderstaande begrippen lijst.

Concepten

Machine Learning:

Machine Learning

Subfield within Artificial Intelligence that focuses on making computers capable of learning through experience.

Learning Problem

The problem of improving some measure of performance when executing some task, through some type of training experience.

Training

Process in which a machine learning algorithm learns through experience the optimal values for a set of tunable parameters.

Testing

Process in which the performance of a machine learning algorithm is tested on before unseen data, using the parameter values that resulted from training the algorithm.

Supervised Learning

Form of learning in which a learning algorithm is presented with labeled data, usually in the form of (x, y) pairs. The goal is to produce an accurate prediction y' based on input x' . Predictions are generally formed by using a mapping $f(x)$, which is learned during training and produces an output y for every input x .

Artificial Neural Network (ANN)

A machine learning algorithm which is loosely based on human neural networks. An ANN consists out of at least three layers; the input data enters the network through an *input layer* (1) that consists out of multiple input nodes. The data is then passed on to a *hidden layer*(2). This layer applies specific weights to the inputs it receives and passes them on to a next hidden layer or to the *output layer* (3). The *output layer* presents the prediction the network makes based on the input data. ANN's with many hidden layers are called *Deep Neural Networks*.

Unsupervised Learning

Form of learning in which a learning algorithm is presented with unlabeled data. Here the goal is to find, under specific assumptions, patterns in the input data.

Clustering

An example of an unsupervised learning problem. This is the problem of finding a partition (a division into separate groups) within the input data in the absence of any labels that indicate a desired partition.

Reinforcement Learning

A third form of machine learning in which the training data is somewhere in between supervised and unsupervised learning. In contract to supervised learning, the training examples do not indicate the correct output (y') for a given input (x'). Instead the training data only give an indication as to whether an action is correct or not. Through such indications, also called rewards, the correct output is learned.

Environment

This refers to the environment in which a machine learning algorithm operates. Among others, this environment consists out of the computing architecture (e.g. run on only 1 machine or several processors), the source of the data and other machine learning systems or agents. All these aspects of the environment present a machine learning algorithm with resources but also place constraints on these resources (e.g privacy constraints on the data source).

Week 3

Literatuur

- Cuellar, A. (2017), 'Causal reasoning and data analysis: problems with the abusive head trauma diagnosis', Law, Probability and Risk, 16.4 (okt 2017), p. 223-239, issn: 1470-8396.
- Pearl J. (2018), 'Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution', arXiv:1801.04016v1.
- Besold, R. et al. (2017), 'Neural-Symbolic Learning and Reasoning: A Survey and Interpretation', p. 1-9, arXiv:1711.03902.
- Jieshu Wang (2017), 'Symbolism vs. Connectionism: A Closing Gap in Artificial Intelligence', blog published online

Hoorcollege 5: Causal Inference



Prof. Dr. Joris Mooij is hoogleraar bij het Korteweg – De Vries Instituut van de UvA. Zijn proefschrift heeft hij geschreven op de Radboud Universiteit Nijmegen, over inductie in grafische modellen.

- [Opname van webcollege](#)

Hoorcollege 6: Symbolic vs Sub-symbolic AI



Prof. Dr. Ing. Robert A.M. van Rooij is hoogleraar bij het Institute for Logic, Language and Computation van de UvA. Zijn ingenieurgraad behaalde hij in Bedrijfskunde in 's Hertogenbosch, daarna behaalde hij zowel een Master in Logica & Filosofie in Nijmegen en een Master in Linguistics in Tilburg. Zijn promotie was in Computational Linguistics in Stuttgart.

- [Opname van webcollege](#)

Werkgroep 5: Causal Reasoning

Topics

- Causation en correlation
- Forecasting, backcasting en attribution
- The Three Layer Causal Hierarchy

Literatuur

Cuellar, A. (2017), 'Causal reasoning and data analysis: problems with the abusive head trauma diagnosis', *Law, Probability and Risk*, 16.4 (okt 2017), p. 223-239, issn: 1470-8396.

Pearl J. (2018), 'Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution', arXiv:1801.04016v1.

Kernpunten

Aan de hand van een juridisch voorbeeld probeert Cuellar het belang van *attribution* aan te tonen. Daarvoor zet Cuellar eerst uiteen wat er precies allemaal mis is met de data gebruikt in het voorbeeld. Zorg ervoor dat je alle punten die worden gemaakt goed begrijpt en onthoudt. Vervolgens worden de concepten *forecasting*, *backcasting* en *attribution* geïntroduceerd. Hier is het vooral belangrijk goed te begrijpen wat deze concepten betekenen, hoe deze zich tot elkaar verhouden en waarom *attribution* zo belangrijk is. Leer ten slotte hoe deze concepten gevangen kunnen worden in formele kansuitdrukkingen.

Het artikel van Pearl gaat over causale modellen. Hij biedt een hiërarchie aan met drie levels van causaliteit die een model kan hebben, en wat een model kan op dat level. Deze tabel is belangrijk omdat die illustreert wat causale modellen kunnen dat niet-causale modellen niet kunnen. In de rest van zijn artikel geeft Pearl een aantal concrete voorbeelden van belangrijke mijlpalen in AI die behaald zijn door het gebruik van causale modellen.

Concepten

Cuellar (2017):

Correlation

A correlation is a measure from statistics that represents the extent to which two variables fluctuate together. If there exists a strong positive correlation between two variables, an increase in one variable will also lead to an increase in the other variable.

Causation

Refers to the phenomenon of one event causing another event. When there is a causal relation between variables, changes in one variable cause a change in the other variable.

Logistic Regression

A type of regression analysis which is used to predict the outcome of

a binary variable (e.g. spam or no spam) based on a set of predictor variables.

Covariates

A variable that is related to the dependent variable (variable about which we would like to make predictions) in a regression analysis. A covariate can be an independent variable or it can be an unwanted (and often unobserved) third variable that is responsible for the correlation between two other variables.

Omitted Variable Bias

Bias that results from omitting one or more relevant variables in a statistical analysis.

Sample Selection Bias

Bias that results from selecting a non-random sample for a statistical analysis.

External Validity

The validity of a study to apply its conclusions outside of the context of the study. For example, a study has low external validity if the used sample is not representative of the entire population we like to make predictions about.

Attribution

Type of probabilistic inference concerning the probability that one event was a necessary cause of another, such that the effect would not be observed but for that cause.

Forecasting

Type of probabilistic inference concerning the probability that an outcome occurs given that an exposure occurred.

Backcasting

Type of probabilistic inference concerning the probability that an exposure occurred given that an outcome occurred.

Pearl (2018):

Causal reasoning

Causal reasoning is the kind of reasoning in which we try to identify causal relations of the form 'A causes B'.

Counterfactual

A conditional statement of a 'what if' form that states what did not happen but could have happened. For example, 'What would have happened to the world if the COVID-19 pandemic never happened?'

Association, intervention & counterfactuals

The three different layers of Pearl's Causal Hierarchy. Within each level different kinds of questions can be answered. See Figure 1.

Structural Causal Models

A mathematical framework that enables the formation of mathematical equations that capture causal relations. See figure 2.

Transparency

An characteristic of assumptions encoded in a model that enables

analysts to discern whether the assumptions are plausible, or whether additional assumptions are warranted.

Testability

An characteristic of assumptions encoded in a model that enables analysts to determine whether the assumptions are compatible with the available data and, if not, identify those that need repair.

Do-calculus

Logical framework that predicts the effect of interventions if this is feasible given the available data and the assumptions made.

Mediation

Statistical phenomenon in which a third variable explains the relation between two other variables.

Werkgroep 6: Symbolisme en Connectionisme

Topics

- Symbolism and Connectionism
- Neural-Symbolic AI

Literatuur

Besold, R. et al. (2017), 'Neural-Symbolic Learning and Reasoning: A Survey and Interpretation', p. 19, arXiv:1711.03902.

Jieshu Wang (2017), 'Symbolism vs. Connectionism: A Closing Gap in Artificial Intelligence', blog published online

Kernpunten

In het eerste paper van Besold e.a. wordt *Neural-Symbolic AI* uitgelegd. Neural-symbolic AI wordt gepresenteerd als een samenvoeging van neurale aanpakken voor AI en symbolische aanpakken voor AI. Neurale aanpakken worden ook wel connectionistische aanpakken genoemd, en symbolische aanpakken zijn sterk verwant met formele logica. De focus van dit paper is om te laten zien hoe deze twee verschillende aanpakken samengevoegd kunnen worden, en te laten zien wat voor voordelen dit heeft.

Het tweede document van Wang gaat verder op het onderscheid tussen symbolische en connectionistische AI. Het geeft een overzicht van de ontwikkeling van beide aanpakken, en in dit overzicht komen ook veel van de papers terug die jullie eerder al hebben gelezen. Het kernargument van dit paper is dat connectionistische en symbolische AI steeds meer samengevoegd worden, en dat het onderscheid eigenlijk ook nooit echt heeft bestaan.

Concepten

Besold (2017):

Neural symbolic system

Een Neural Symbolic System voegt connectionistische elementen en symbolische / logische elementen samen in 1 systeem. Typisch voeren de connectionistische subsystemen taken uit op een lager niveau van abstractie (zoals objectherkenning uit visuele input) en de symbolische subsystemen taken op een hoger abstractieniveau (bijvoorbeeld redeneren op basis van welke objecten zijn herkend).

Connectionistische AI

Een connectionistische AI bestaat uit losse elementen met gewichten ertussen, ongeveer zoals neuronen en synapses in het brein. Een connectionistische AI leert door deze gewichten te veranderen. In een connectionistische AI is vaak niet duidelijk sprake van representatie. Neural networks zijn bijvoorbeeld connectionistische AI's.

Symbolische AI

Een symbolische / logische AI bestaat typisch uit symbolen die iets representeren, en vaak vaste regels hanteren over hoe die symbolen met elkaar interacteren. Symbolische AI's bevatten bijvoorbeeld vaak expliciete 'if ... then ...' regels.

Week 4

Deeltentamen

Het deeltentamen zal online worden afgenomen via [ans-delft](#). Hierbij een voorbeeld van drie vragen van het eerste deeltentamen van vorig jaar.

10p **1c** Welke vraag hoort volgens Russel & Norvig (2010) *niet* bij Filosofie als basis (*Foundation*) van de Kunstmatige Intelligentie?

- | | |
|---|--|
| ☐ How does the mind arise from a physical brain? | ☐ How do we reason with uncertain information? |
| ☐ Can formal rules be used to draw valid conclusions? | ☐ How does knowledge lead to action? |

10p **2e** Wat is Brooks' belangrijkste kritiek op eerdere AI experimenten (zoals SHRDLU van Winograd 1972)?

- | | |
|--|--|
| ☐ De reasoning algorithms waren niet complex genoeg. | ☐ De omgeving van de experimenten was niet complex genoeg. |
| ☐ Zowel de algoritmen als de experimenten waren niet complex genoeg. | ☐ Eerdere AI experimenten maakten geen gebruik van evolutie. |

30p **4f** Leg in je eigen woorden uit wat het middelste niveau van Pearl's *Three Layer Causal Hierarchy* inhoudt. Doe dit **zonder** gebruik te maken van een voorbeeld.

Let op: Alleen de laatste vraag is een goed voorbeeld van wat je dit jaar kan verwachten. Omdat het examen online wordt afgenomen zullen het dit jaar voornamelijk open vragen zijn.

Week 5

Literatuur

- Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton (2017), 'ImageNet classification with deep convolutional neural networks', Communications of the ACM, Volume 60 Issue 6, p. 84-90.
- K. Hofmann, S. Whiteson, A. Schuth, and M. de Rijke, 'Learning to rank for information retrieval from user interactions', SIGWEB Newsletter, 5(Spring):1-7, April 2014. ISSN 1931-1745.
- T. Ruotsalo, G. Jacucci, P. Myllymaki, and S. Kaski, 'Interactive intent modeling: Information discovery beyond search', Communications of the ACM, 58(1):86-92, 2015.

Hoorcollege 7: Computer Vision by Learning



Prof. Dr. Cees G.M. Snoek is gepromoveerd aan de Universiteit van Amsterdam. Daarna kreeg hij o.a. een *Fullbright Scholarship* in Berkeley. Cees Snoek is expert op het terrein van video- en beeldherkenning.

- Opname van webcollege

Hoorcollege 8: Information Retrieval as Interaction



Prof. Dr. Maarten de Rijke is Universiteits hoogleraar. Hij stond aan de wieg van de Amsterdam School of Data Science. Recent kwam Prof. de Rijke in het nieuws met de aankondiging van de nieuwe samenwerking met Huawei.

- Opname van webcollege

Werkgroep 7 - I: Computer Vision

Topics

- Computer Vision
- Convolutional Neural Networks

Literatuur

Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton (2017), '[ImageNet classification with deep convolutional neural networks](#)', Communications of the ACM, Volume 60 Issue 6, p. 84-90.

Kernpunten

In dit artikel presenteren Krizhevsky, Sutskever en Hinton (2017) AlexNet. AlexNet is een convolutional neural network dat in 2012 meedeed aan de *ImageNet Large Scale Visual Recognition Challenge*. Het netwerk bleek zeer goede resultaten te behalen in het herkennen van objecten in foto's. Inmiddels zijn er alweer betere resultaten bereikt met andere netwerkkarchitecturen, maar AlexNet vormt nog steeds een basis voor veel hedendaagse convolutional neural networks. Zo introduceerde dit paper een aantal concepten/technieken die ervoor zorgde dat AlexNet de concurrentie kon verslaan en die nu nog steeds worden toegepast in veel hedendaagse computer vision toepassingen. Net zoals bij het artikel over Machine Learning is het wederom belangrijk deze concepten goed te begrijpen. Bestudeer daarvoor ook onderstaande begrippenlijst.

Concepten

Object recognition

A computer vision technique used for recognizing object in images or videos.

Convolutional Neural Network (CNN)

Class of neural networks mostly used for processing images. CNN's often have less connections and parameters than standard feedforward neural networks. Their capacity can be controlled by varying their depth and breadth, and they also make strong and mostly correct assumptions about the nature of images. A short introduction to CNN's can be found [here](#).

Overfitting

Phenomenon in which a machine learning model performs well on the data it has been trained on but is not capable of generalizing well to unseen data.

Top-5 error

The fraction of images on which the model is tested for which the correct label is not among the five labels considered most probable by the model. The top-1 error corresponds to the fraction of test images for which the model is not capable of predicting the right label.

Gradient descent

An iterative optimization function that is used in learning to find a set of weights connecting the different layers within a neural network that minimizes the value of the loss function. Loss functions are used to determine the error, or 'loss', between the output of a network and the desired target values.

Activation function

A function that determines the output of a neuron within a neural network based on an input or set of inputs received from neurons in the previous layer of the network. Much used examples are the ReLU and tanh activation functions. With the ReLU activation function the output is equal to the input if the input is bigger than 0, else the output is equal to 0.

Saturation

Phenomenon in which neurons within a neural network almost always output values that are close to the minimum or maximum value of an activation function. For example, the tanh activation function can only output activation values between -1 and 1. If a neuron predominantly outputs values close to 1, this significantly increases training time. An unbounded activation function like ReLU can be used to mitigate this problem.

GPU parallelization

Training neural networks on multiple GPU's at the same time.

Hyper-parameters

The variables that determine the structure of the network and the variables that determine the way the network is trained.

Pooling layer

A layer within a CNN that is used to reduce the size of the previous layer. This helps to extract dominant features from the images and might significantly reduce the computational power needed for training the CNN (see link added to Convolutional Neural Network).

Data augmentation

A process of artificially enlarging the available dataset by applying label-preserving transformations to the images in the dataset.

Weight decay

Parameter that is used to slightly decrease the weights that connect different neurons during training of the network.

Learning rate

Parameter of the optimization algorithm that is used to find the set of weights that minimizes the loss function. The learning rate might significantly influence the time needed to train a network.

Werkgroep 7 - II: Information Retrieval

Topics

- Information Retrieval
- Online learning to rank
- Probabilistic Interleaving
- Interactive Intent Modeling

Literatuur

Hofmann, S. Whiteson, A. Schuth, and M. de Rijke, 'Learning to rank for information retrieval from user interactions', SIGWEB Newsletter, 5(Spring):1-7, April 2014. ISSN 1931-1745.

T. Ruotsalo, G. Jacucci, P. Myllymaki, and S. Kaski, 'Interactive intent modeling: Information discovery beyond search', Communications of the ACM, 58(1):86-92, 2015.

Kernpunten

Hofmann e.a. (2014) presenteren een *information retrieval* systeem dat zijn resultaten kan verbeteren door te leren van gebruikersinteracties. Door te leren van zulke interacties kan het systeem een steeds betere ranking van de opgehaalde informatie bepalen. De data die wordt gebruikt om een optimale ranking te leren bestaat grotendeels uit *clicks* en bevat veel ruis. Zulke *clicks* zijn daarom vaak geen accurate representatie van de voorkeuren van de gebruiker (welke informatie de gebruiker wilt vinden). Het leren van een optimale ranking functie op basis van *clicks* is dan ook een lastige taak. In dit artikel worden een aantal oplossingen gepresenteerd om de uitdagingen van het leren van *clicks* te boven te komen.

Het is belangrijk om de oplossingen die Hofmann et al (2014) aandragen goed te bestuderen. Zorg dat je goed begrijpt wat nieuw is aan de oplossingen die worden aangedragen en waarom deze het leerproces verbeteren. Bestudeer daarvoor ook goed de afbeeldingen in het artikel.

Ruotsalo e.a. (2015) presenteren een alternatief voor hedendaagse zoekmachines zoals de zoekmachine van Google. Zulke zoekmachines zijn namelijk niet goed in *exploratory search*. Het alternatief dat wordt gepresenteerd combineert *interactive intent modeling* en visuele gebruikersinterfaces om gebruikers stapsgewijs naar relevante informatie te sturen. Dit gebeurt door de huidige zoekopdracht van een gebruiker te visualiseren binnen de informatieruimte waarin de zoekopdracht zich bevindt. Door middel van interacties met deze visualisatie kan de gebruiker zijn/haar zoekopdracht steeds meer verfijnen om zo op zoek te gaan naar de beste zoekopdracht.

Bestudeer in dit artikel goed hoe de gepresenteerde technieken ervoor zorgen dat gebruikers van *search engines* beter relevante informatie kunnen vinden. Zorg ervoor dat je in eigen woorden kunt uitleggen hoe het systeem werkt, wat de achterliggende principes zijn en welke grote voordelen dit met zich meebrengt.

Concepten

- Hofmann et al. (2014):

Information retrieval

Subfield within Artificial Intelligence that focuses on creating algorithms for the retrieval of information from ever growing databases of mostly textual information. Examples of information retrieval systems are web search engines (e.g Google) and recommender systems.

Result ranking

The ranking of retrieved information documents based on a specific query. Think of the order in which pages appear after entering a query in the Google search engine. The ranking of retrieved information is determined by a ranking function.

Online learning

A form of learning in which an information retrieval system learns directly from natural interactions with users. Such user interactions usually consist out of clicks by the users.

Presentation bias

The bias that is introduced by the order in which information is presented to a user. Information that is presented on the top of a result page has a higher probability of being clicked on.

Interleaving

A method for comparing different rankers by using user interactions and combining the results of both rankers. This means that the retrieved information of both rankers is combined to one search result, after which both rankers are compared based on the user interactions. The article presents a new method of interleaving called Probabilistic Interleave.

Pareto-dominates

One ranker Pareto-dominates another ranker when it ranks all documents clicked on by the user at least as high as the competing ranker.

Importance sampling

A statistical technique that is used by Probabilistic Interleave and which enables the reusing of historical user interaction data.

Reinforcement learning

See werkcollege on Machine Learning

Exploration-exploitation trade-off

A well-known problem within many reinforcement learning tasks that is concerned with finding the right balance between exploration, trying out new and before unseen options to gain new information, and exploitation, exploiting already learned information.

- Ruotsalo et al. (2015):

Exploratory search

Type of search in which users have difficulty expressing their information needs and new search intents may emerge and be discovered only as they learn by reflecting on the acquired information.

Vocabulary mismatch problem

The problem that in human communication often the humans writing the documents to be retrieved and the humans searching for them are likely to use very different vocabularies to encode and decode their intended meaning.

Week 6

Literatuur

- Repplinger, M. et al. (2018), 'Vector-space models of words and sentences', Nieuw Archief voor Wiskunde 5/19, 3: 167-174.
- Riederer et al. (2017), 'The Price of Fairness in Location Based Advertising', Proceedings of the FATREC-Workshop on Responsible Recommendation.

Hoorcollege 9: Natural Language Processing



Dr. Raquel Fernández is universitaire hoofddocent bij het *Institute for Logic, Language and Computation* van de UvA, waar ze leiding geeft aan de *Dialogue Modelling Group*. Ze deed een master Cognitive Science and Language in Barcelona en een PhD in Computational Linguistics bij het King's College London. Daarna was ze o.a. post-doc bij Stanford University.

- [Opname van webcollege](#)

Hoorcollege 10: Responsible Data-Driven Decision Making



Prof. Dr. Hinda Haned is buitengewoon hoogleraar aan de UvA. Zij heeft een bachelor in Mathematics & Computer Science, een master in Ecology, Evolution & Biometry, en een PhD in Applied Statistics van de University of Lyon. Voor haar positie als *lead data scientist* bij Ahold Delhaize heeft ze gewerkt als *research statistician* bij het Nederlands Forensisch Instituut.

- [Opname van webcollege](#)

Werkgroep 8 - I: Natural Language Processing

Topics

- Natural Language Processing
- Syntax and semantics
- Linguistic models

Literatuur

Repplinger, M. et al. (2018), 'Vector-space models of words and sentences', Nieuw Archief voor Wiskunde 5/19, 3: 167-174.

Kernpunten

Repplinger, Beinborn en Zuidema (2018) presenteren in dit artikel de verschillende stadia waar het wetenschapsveld *natural language processing* (NLP) doorheen is gegaan om te komen tot de huidige state-of-the-art deep learning modellen voor het verwerken van taal. Waar NLP in de vroege jaren 20 vooral gedomineerd werd door op logica gebaseerde modellen, zijn tegenwoordig *vector-space models* het populairst. Deze modellen maken gebruik van numerieke vector representaties en lineaire algebra om de betekenis van zinnen te achterhalen. Het artikel beschrijft in 4 stappen hoe NLP van logische modellen de transitie naar *vector-space models* heeft gemaakt. Probeer voor elke stap goed duidelijk te krijgen wat de karakteristieke elementen van die stap zijn, wat de grote voordelen daarvan zijn en wat de zwaktepunten.

Concepten

Natural language processing

Subfield within Artificial Intelligence focused on developing mathematical and computational models of language.

Vector-space models

Language models in which words are represented by numerical vectors and sentence meanings are computed by using a variety of operations from linear algebra on these vectors.

Montague semantics

Montague semantics entail the first successful attempt of formalizing the semantics of substantial fragments of natural language (e.g. a sentence). The logic-based system provides a systematic way to translate natural language to a logical language.

Semantics

The study of meaning. In the context of NLP this can be interpreted as the study of the meaning of words and sentences.

Syntax

Syntax is the study of the structure of sentences and the rules of grammar. The semantics of a sentence refers to its meaning and the syntax to its structure.

Principle of compositionality

Principle that states that the meaning of a complex expression (e.g. sentence) is determined by the meaning of its constituents (e.g. words) and the rules used to combine them.

Recursion

Recursion in NLP refers to the process in which the same grammatical rule can be used repeatedly to form a sentence. This allows us to form infinitely many complex expressions from a finite set of simple grammar rules.

Distributional semantics

A linguistic system in which the meaning of a word is modelled based on the words surrounding it. More specifically this is done by representing a word as a numerical vector which is filled by counting how often other specific words occur near to it in large databases of text. Central to this system is the distributional hypothesis that states that "you shall know a word by the company it keeps".

Semantic space

A multi-dimensional space in which words can be represented as vectors and in which the distance between different vectors is supposed to represent the semantic similarity between the words.

Neural word embeddings

This term refers to vector representations of words that are acquired by training neural networks to predict a word based on its context or the other way around.

Continuous Bag-of-Words (CBOW)

Algorithm used to predict the occurrence of a word, given a context of surrounding words.

Skip-Gram

Algorithm used for predicting a context (the surrounding words), given an individual word. The word2vec neural word embeddings are based on the CBOW and Skip-Gram algorithms.

Compositional distributional semantics

A linguistic system that is supposed to bring together the strengths of the symbolic tradition in NLP (e.g. Montague semantics) and the strengths of vector-space models from the distributional and neural tradition. More specifically, this approach tries to combine insights about the compositionality of language (symbolic tradition) with the vector representations (neural tradition) of words to model the meaning of sentences.

Werkgroep 8 - II: Data Driven Decision Making

Topics

- Fairness in algorithms
- Algorithmic bias
- Location based advertising

Literatuur

Riederer et al. (2017), 'The Price of Fairness in Location Based Advertising', Proceedings of the FATREC-Workshop on Responsible Recommendation.

Kernpunten

Riederer en Chaintreau (2017) onderzoeken wat het effect is van het mee laten wegen van *fairness* in adverteren op de omzet van een bedrijf. Zij doen dit aan de hand van een empirische analyse van een echt voorbeeld, namelijk op locatie gebaseerde personalisatie van advertenties voor Instagram gebruikers. Tegenwoordig beschikken veel bedrijven over locatie data van individuen en deze data wordt vaak gebruikt om gericht te kunnen adverteren. Hoewel zulk gebruik van locatie data veel voordelen heeft, bestaan er ook grote risico's. Zo kan bijvoorbeeld gemakkelijk discriminatie tegen kwetsbare bevolkingsgroepen worden geïntroduceerd. In dit artikel worden methoden onderzocht om dit soort risico's te vermijden en op een eerlijke manier advertenties te personaliseren op basis van locatie data.

Een interessante bijdrage die Riederer en Chaintreau leveren met dit onderzoek is een praktische defintie van *fairness*. Zorg ervoor dat je beide definties van *fairness* die worden gegeven goed begrijpt. Probeer daarbij een helder beeld te krijgen van het onderzoek dat zij hebben uitgevoerd. Wat voor soort data is gebruikt en hoe is de data verwerkt? Hoe zijn verschillende manieren van adverteren vergeleken? Wat laten de resultaten zien?

Concepten

Algorithmic bias

Systematic and repeated errors in the outcomes of an algorithm that cause unfair predictions.

Individual fairness

Fairness defined as 'similar people should see similar outcomes'. More specifically, a person A that shows the same traits as another person B should have a similar probability of a specific outcome as the similar person B. In case of an advertising system this means that similar users must be shown similar sets of ads.

Group fairness

Fairness defined as 'the difference in probability between two groups of receiving a specific outcome (e.g. seeing a particular ad) should be bounded by a value ϵ '.

Earth Mover's Distance

A metric used to measure the distance between two probability distributions.

Granularity

Granularity refers to the level of detail captured in our data. When working with location data for example, we could gather data on entire provinces (e.g. Noord Holland) or we could use a finer granularity and gather data on the level of cities (e.g. Amsterdam) or streets (e.g. Science Park).

Week 7

Literatuur

- Breazeal et al. (2016), '[Social Robotics](#)', Chapter 72, In: Sicilian, B. and O. Khatib (eds.), Springer Handbook of Robotics (pp. 1935-1960), 2nd edition, Berlin: Springer.
- Andrej Zwitter (2014), '[Big Data ethics](#)', Big Data & Society, July-December 2014: 1-6.
- Claudia Aradau and Tobias Blanke (2015), '[The \(Big\) Data-security assemblage: Knowledge and critique](#)', Big Data & Society, July-December 2015: 1-12.

Hoorcollege 11: Social Robotics



Prof. Dr. Koen Hindriks is professor aan de Vrije Universiteit. Hij studeerde af in Groningen, behaalde zijn PhD in Utrecht, was *assistant professor* in Nijmegen en *associate professor* in Delft. Met zijn bedrijf Interactive Robotics heeft hij meegewerkt aan de documentaire Robo-Sapiens van Jelle Brandt Corstius.

- [Opname van webcollege](#)

Hoorcollege 12: Big Data and implications on society



[Prof. Dr. Tobias Blanke](#) is universiteits hoogleraar aan de UvA. Hij behaalde zijn master in Advanced Computing op Queen Mary, London. Hij heeft zowel een PhD in Political Philosophy als in Computer Science (resp. Berlin en Glasgow). Hij is o.a. Executive van het European Holocaust Research Infrastructure project.

- [Opname van webcollege](#)

Werkgroep 9 - I: Social Robotics

Topics

- Social Robots
- Human-robot Interaction

Literatuur

Breazeal et al. (2016), 'Social Robotics', Chapter 72, In: Sicilian, B. and O. Khatib (eds.), Springer Handbook of Robotics (pp. 1935 - 1960), 2nd edition, Berlin: Springer.

Kernpunten

Het hoofdstuk van Breazeal, Dautenhahn en Kanda presenteert een uitgebreid overzicht van onderzoekstrends in het vakgebied Sociale Robotica. Dit deelgebied van AI heeft tot doel sociale robots te creëren die competente en capabele partners zijn voor mensen. Het ontwikkelen van dergelijke robots is echter een flinke uitdaging. Zo is het bijvoorbeeld een lastige taak gebleken om een robot op natuurlijke wijze met ons zowel verbaal als non-verbaal te laten communiceren. Om robots van dergelijke kwaliteiten te kunnen voorzien, is een diep begrip van onze eigen intelligentie en gedrag nodig. In dit hoofdstuk zullen we meerdere voorbeelden zien van onderzoek dat is gedaan om robots te maken die op een natuurlijke of interpersoonlijke manier met mensen om kunnen gaan. Probeer tijdens het lezen een overzicht te vormen van de vaardigheden die een robot nodig zou hebben om op een natuurlijke manier met mensen om te gaan. Waar moet een robot bijvoorbeeld allemaal toe in staat zijn om op een natuurlijke manier met mensen te communiceren? Zijn er voorbeelden van robots die dat al kunnen en welk gedrag laten deze robots precies zien? Verder worden er een aantal theorieën benoemd waarop wij ons kunnen baseren bij het implementeren van sociale vaardigheden. Zorg ervoor dat je deze theorieën goed kent.

Concepten

Social Robots

Robots designed to interact with people in a natural, interpersonal manner - often to achieve positive outcomes in diverse applications such as education, health, quality of life, entertainment, communication, and tasks requiring collaborative teamwork.

Anthropomorphisme

The attribution of human traits to non-human entities. Humans often tend to anthropomorphize animals and social robots. For this reason, a lot of social robots are designed to mimic human appearances.

Uncanny valley

Term that refers to the phenomenon in which the appearance and movement of a robot resemble more of an animate corpse than a living human. This often elicits strong negative emotions in humans.

Appraisal theory

Theory of emotion that emphasizes a causal connection between cognition and emotion. According to this theory emotions are invoked by patterns of judgments that characterize the personal significance of events. A computational model based on this theory often lends itself to more symbolic AI, in contrast to models based on dimensional theories of emotion.

Theory of Mind

The ability to attribute mental states such as beliefs, intents, desires, feelings, among others, to oneself and to others. A social robot would need some kind of theory of mind to recognize, understand, and predict human behavior in terms of the underlying mental states.

Simulation Theory

Theory that states that certain parts of the brain have dual use; they are used to not only generate our own behavior and mental states, but also simulate the introceptive states of the other person.

Werkgroep 9 - II: Big Data and Implications on Society

Topics

- Big Data Ethics
- Moral Agency
- Big Data uses in Security Practices

Literatuur

Andrej Zwitter (2014), 'Big Data ethics', Big Data & Society, July-December 2014: 1-6.

Claudia Aradau and Tobias Blanke (2015), 'The (Big) Data-security assemblage: Knowledge and critique', Big Data & Society, July-December 2015: 1-12.

Kernpunten

Zwitter (2014) beargumenteert in zijn artikel dat de razendsnelle ontwikkelingen in Big Data ons noodzaken om te heroverwegen hoe wij over ethiek nadenken. Volgens hem passen traditionele ethische opvattingen niet meer in deze tijd waarin Big Data toepassingen alom zijn. Zulke traditionele opvattingen gaan vaak uit van individuele morele keuzevrijheid waarin het individu de verantwoordelijkheid draagt het moreel juiste te doen. Volgens Zwitter (2014) is er in ethische kwesties omtrent Big Data vaak helemaal geen sprake van zulke morele keuzevrijheid. Hij pleit dan ook voor een nieuwe conceptie van ethiek waarbij we onze morele overwegingen niet meer uitsluitend baseren op het idee van individuele keuzes met voorspelbare uitkomsten.

Probeer bij het lezen van dit artikel vooral goed te begrijpen hoe de besproken eigenschappen van Big Data zorgen voor een nieuw begrip van ethiek. Bestudeer daarbij ook goed de specifieke ethische uitdagingen omtrent het gebruik van Big Data besproken in de voorlaatste sectie.

Aradau en Blanke (2015) stellen dat de opkomst van Big Data en gebeurtenissen zoals de Snowden-onthullingen opnieuw het debat hebben aangewakkerd over de manier waarop beveiligingspraktijken worden ingezet in ons digitale tijdperk. Binnen dit debat heeft kennis uit de computer- en informatiewetenschappen volgens hen te weinig aandacht gekregen. Er zou dan ook meer aandacht moeten worden gegeven aan kritische kennis uit dit wetenschapsveld om beweringen van beveiligingsprofessionals over Big Data aan te vechten. Door samen te werken met computerwetenschappers kunnen sociale wetenschappers voorbij de complexiteit van algoritmische methoden komen. Daarmee zullen zij beter in staat zijn een kritische analyse te maken van de veelvoorkomende opvatting dat Big Data een noodzakelijke 'game changer' is. In dit artikel wordt dat gedaan door een aantal veel voorkomende opvattingen over Big Data te ontcrachten aan de hand van informatie uit de computer- en informatiewetenschappen.

In dit artikel is het vooral belangrijk om de drie verschillende opvattingen over Big Data en de ontcrachting daarvan goed te bestuderen. Hoe wordt bijvoorbeeld het onderscheid tussen inhoud en metadata door beveiligingsprofessionals gebruikt om het op grote schaal verzamelen van data te rechtvaardigen? En

hoe proberen de auteurs dit sterke onderscheid te ontcrachten met kennis uit de computer- en informatiewetenschappen?

Concepten

- Zwitter (2014):

Moral agency

The capability of an individual to make moral decisions based on a conception of good and evil and to take responsibility for one's actions.

Moral culpability

The extend to which an agent can be held morally responsible for its actions.

'Many hands' problem

A philosophical problem introduced by Big Data uses which refers to the occurrence of an undesirable effect to which many agents contributed and in which it is very hard, or impossible, to hold any individual agent responsible.

Concept of centrality

A concept relevant to definitions of power in which power relations between agents are modelled within a network. This concept states that the more connections an specific agent has within the power network, the more power the agent can exert.

Infraethics

Agents that through power relations hinder or facilitate the capability of other agents to act morally. This introduces 'dependent agency', since the capability of one agent to act morally is dependent on other agents.

- Aradau and Blanke (2015):

Big Data-security assemblage

Term used to refer to the increasing use of Big Data techniques in security practices.

Metadata

Usually defined as data that provides information about other data. For example, data that captures the content of a phone call can be described by metadata that contains the timestamp the phone call started, the duration of the phone call, the phone number of the caller, etc.

Mass surveillance

The often intricate monitoring by intelligence agencies of an entire or substantial fraction of a population. In contrast to mass surveillance, targeted surveillance is directed to specific persons of interest.

Epistemology

Subfield within philosophy that studies the nature, origin and limits of knowledge. The epistemic transformation that took place after the ‘winter of AI’ refers to the transformation in the kind of knowledge that was the subject of study in AI and the way we tried to gather such knowledge (i.e. from logic based models to statistical models).

Epistemic capability of algorithms

The capabilities of algorithms to lead to new knowledge or to make new discoveries. A common believe between security professionals is that gathering more and more data enhances such capabilities of algorithms.

False positive

A classification error in which a model wrongly predicts the presence of a condition. When a predictive model predicts a healthy individual to be sick, this is considered a false positive.

Heavy-tailed distribution

Probability distribution over possible outcomes with a high amount of (extreme) outliers (e.g. conducting a terrorist attack) with only a small probability.

Posterpresentaties

De posterpresentaties vinden plaats op vrijdag 16 oktober. Elk groepje zal zijn poster presenteren voor groepjes uit andere werkgroepen. Dat betekent dus dat je gaat presenteren voor een groep medestudenten die jullie poster nog niet hebben gezien. Gebruik de generale repetitie in de PAV werkgroep in week 6 dan ook goed om nog wat laatste punten van feedback te verzamelen van de studenten uit jouw werkgroep.

Omdat jullie niet voor eigen werkgroep gaan presenteren is het niet handig om de gebruikelijke Discord groepen te gebruiken. De presentaties zullen daarom op Zoom zijn. Er zullen drie Zoom-sessies met presentaties tegelijkertijd bezig zijn. Het rooster vinden jullie op Canvas onder Modules → Opdracht Wetenschappelijke Poster, of klik [hier](#) (onderaan pagina). Zorg ervoor dat je de Zoom sessie niet eerder binnengaat dan 5 minuten voordat jullie presentatieronde begint. Let er ook op dat we tijdens het presenteren streng zullen zijn op de tijd die jullie hebben. Iedere presentatie duurt 10 minuten inclusief 2 à 3 minuten voor vragen uit het publiek. Oefen je presentatie daarom van tevoren alvast met een timer zodat je zeker weet dat het jullie lukt alle info in de gegeven tijd te presenteren.

Let op! Mocht je echt niet aanwezig kunnen zijn op het aangegeven tijdstip in het rooster, probeer dan eerst met jouw tutor en de rest van de PAV groep een oplossing te vinden (door bijvoorbeeld twee groepjes om te ruilen). Mocht het echt niet lukken, dan zal ik (Wessel) met de tutor een andere oplossing proberen te vinden.

Referenties

- [1] Claudia Aradau en Tobias Blanke. „The (Big) Data-security assemblage: Knowledge and critique”. In: *Big Data & Society* 2.2 (okt 2015). URL: <https://doi.org/10.1177/2053951715609066>.
- [2] Tarek R. Besold e.a. *Neural-Symbolic Learning and Reasoning: A Survey and Interpretation*. Nov 2017. arXiv: [1711.03902](https://arxiv.org/abs/1711.03902) [cs.AI]. URL: <https://arxiv.org/abs/1711.03902>.
- [3] Cynthia Breazeal, Kerstin Dautenhahn en Takayuki Kanda. „Social Robotics”. In: *Springer Handbook of Robotics*. Red. door Bruno Siciliano en Oussama Khatib. Springer International Publishing, 2016, p. 1935–1972. ISBN: 978-3-319-32552-1. URL: https://doi.org/10.1007/978-3-319-32552-1_72.
- [4] Rodney A. Brooks. „Intelligence without representation”. In: *Artificial Intelligence* 47.1 (jan 1991), p. 139–159. ISSN: 0004-3702. URL: <http://www.sciencedirect.com/science/article/pii/000437029190053M>.
- [5] Maria Cuellar. „Causal reasoning and data analysis: problems with the abusive head trauma diagnosis”. In: *Law, Probability and Risk* 16.4 (okt 2017), p. 223–239. ISSN: 1470-8396. URL: <https://doi.org/10.1093/lpr/mgx011>.
- [6] Katja Hofmann e.a. „Learning to Rank for Information Retrieval from User Interactions”. In: *SIGWEB Newsletter* Spring (apr 2014). ISSN: 1931-1745. URL: <https://doi.org/10.1145/2591453.2591458>.
- [7] Michael I. Jordan en Tom M. Mitchell. „Machine learning: Trends, perspectives, and prospects”. In: *Science* 349.6245 (jul 2015), p. 255–260. ISSN: 0036-8075. URL: <https://science.sciencemag.org/content/349/6245/255>.
- [8] Alex Krizhevsky, Ilya Sutskever en Geoffrey E. Hinton. „ImageNet Classification with Deep Convolutional Neural Networks”. In: *Communications of the ACM* 60.6 (mei 2017), p. 84–90. ISSN: 0001-0782. URL: <https://doi.org/10.1145/3065386>.
- [9] Allen Newell en Herbert A. Simon. „Computer Science as Empirical Inquiry: Symbols and Search”. In: *Commun. ACM* 19.3 (mrt 1976), p. 113–126. ISSN: 0001-0782. URL: <https://doi.org/10.1145/360018.360022>.
- [10] Judea Pearl. *Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution*. Jan 2018. arXiv: [1801.04016](https://arxiv.org/abs/1801.04016) [cs.LG]. URL: <https://arxiv.org/abs/1801.04016>.
- [11] Michael Repplinger, Lisa Beinborn en Willem Zuidema. „Vector-space models of words and sentences”. In: *Nieuw Archief voor Wiskunde* 19.3 (2018), p. 167–174. URL: <http://www.nieuwarchief.nl/serie5/toonnummer.php?deel=19&nummer=3&taal=0>.
- [12] Christopher Riederer en Augustin Chaintreau. „The Price of Fairness in Location Based Advertising”. In: *Proceedings of the FATREC-Workshop on Responsible Recommendation*. Aug 2017. URL: <https://scholarworks.boisestate.edu/fatrec/2017/1/5/>.

- [13] Tuukka Ruotsalo e.a. „Interactive Intent Modeling: Information Discovery beyond Search”. In: *Communications of the ACM* 58.1 (dec 2014), p. 86–92. ISSN: 0001-0782. URL: <https://doi.org/10.1145/2656334>.
- [14] Stuart J. Russell en Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall series in Artificial Intelligence. Prentice Hall, 2010. ISBN: 9780136042594. URL: <https://www.pearson.com/us/higher-education/product/Russell-Artificial-Intelligence-A-Modern-Approach-3rd-Edition/9780136042594.html>.
- [15] Alan M. Turing. „Computing Machinery and Intelligence”. In: *Mind* 59.236 (okt 1950), p. 433–460. ISSN: 0026-4423. URL: <https://doi.org/10.1093/mind/LIX.236.433>.
- [16] Jieshu Wang. *Symbolism vs. Connectionism: A Closing Gap in Artificial Intelligence*. Jieshu’s Blog. Dec 2017. URL: <http://wangjieshu.com/2017/12/23/symbol-vs-connectionism-a-closing-gap-in-artificial-intelligence/>.
- [17] Andrej Zwitter. „Big Data ethics”. In: *Big Data & Society* 1.2 (nov 2014). URL: <https://doi.org/10.1177/2053951714559253>.