Andreas Holzinger

185.A83 Machine Learning for Health Informatics 2019S, VU, 2.0 h, 3.0 ECTS Lecture 02 - Dienstag, 19.03.2019



From Clinical Decision Support to Causal Reasoning and explainable AI

andreas.holzinger AT tuwien.ac.at

https://human-centered.ai/machine-learning-for-health-informatics-class-2019





ntered.ai (Holzinger Group)

2019 Machine Learning for Health 02

Decision support system (DSS)

■ MYCIN - Rule Based Expert System

GAMUTS in Radiology

Keywords

Reasoning under uncertainty

Example: Radiotherapy planning

Example: Case-Based Reasoning

Explainable Artificial intelligence

■ Re-trace > Understand > Explain

Transparency > Trust > Acceptance

Fairness > Transparency > Accountability

Causality > Causability

(Some) Methods of Explainable AI

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

🛄 Advance Organizer (1/2)

Causality = fundamental relationship between cause and effect

Causability = similar to the concept of usability the property of a human explanation Case-based reasoning (CBR) = process of solving new problems based on the solutions of similar past

Certainty factor model (CF) = a method for managing uncertainty in rule-based systems;

CLARION = Connectionist Learning with Adaptive Rule Induction ON-line (CLARION) is a cognitive architecture that incorporates the distinction between implicit and explicit processes and focuses on capturing the interaction between these two types of processes. By focusing on this distinction, CLARION has been used to simulate several tasks in cognitive psychology and social psychology, CLARION has also been used to implement intelligent systems in artificial intelligence applications.

Clinical decision support (CDS) = process for enhancing health-related decisions and actions with pertinent, organized clinical knowledge and patient information to improve health delivery

Clinical Decision Support System (CDSS) = expert system that provides support to certain reasoning tasks, in the context of a clinical decision:

Collective Intelligence = shared group (symbolic) intelligence, emerging from cooperation/competition of many individuals, e.g. for consensus decision making;

Counterfactual = relating to or expressing what has not happened or is not the case

Crowdsourcing = a combination of "crowd" and "outsourcing" coined by Jeff Howe (2006), and describes a distributed problem-solving model; example for crowdsourcing is a public software beta-test;

Decision Making = central cognitive process in every medical activity, resulting in the selection of a final choice of action out of several alternatives;

Decision Support System (DSS) = is an IS including knowledge based systems to interactively support decision-making activities, i.e. making data useful;

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group

Advance Organizer (2/2)



DXplain = a DSS from the Harvard Medical School, to assist making a diagnosis (clinical consultation), and also as an instructional instrument (education); provides a description of diseases, etiology, pathology, prognosis and up to 10 references for each

Etiology = in medicine (many) factors coming together to cause an illness (see

 Explainable AI = Explainability = upcoming fundamental topic within recent AI; answering e.g. why a decision has been made

Expert-System = emulates the decision making processes of a human expert to solve

GAMUTS in Radiology = Computer-Supported list of common/uncommon differential

 ILIAD = medical expert system, developed by the University of Utah, used as a teaching and testing tool for medical students in problem solving. Fields include Pediatrics, Internal Medicine, Oncology, Infectious Diseases, Gynecology, Pulmonology etc.

 Interpretability = there is no formal technical definition yet, but it is considered as a prerequisite for trust

MYCIN = one of the early medical expert systems (Shortliffe (1970), Stanford) to identify bacteria causing severe infections, such as bacteremia and meningitis, and to recommend antibiotics, with the dosage adjusted for patient's body weight;

Reasoning = cognitive (thought) processes involved in making medical decisions (clinical reasoning, medical problem solving, diagnostic reasoning;

Transparency = opposite of opacity of black-box approaches, and connotes the ability to understand how a model works (that does not mean that it should always be understood, but that - in the case of necessity - it can be re-enacted

ntered.ai (Holzinger Group)

Agenda



HCAI &

HCAI 1

• 00 Reflection – follow-up from last lecture

01 Decision Support Systems (DSS)

02 History of DSS = History of AI

• 03 Example: Towards Personalized Medicine

• 04 Example: Case Based Reasoning (CBR)

05 Causal Reasoning

🞹 How do you explain this ... Carrinoma: 135 images

06 Explainability – Causability

• 07 (Some) Methods of Explainable AI

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)



HCAI &



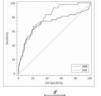
human-centered.ai (Holzinger Group

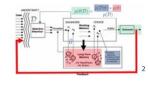
2019 Machine Learning for Health 02

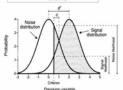
2019 Machine Learning for Health 02

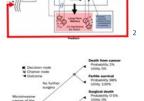
Reflection from last lecture



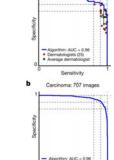


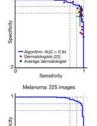




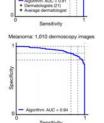








Melanoma: 130 images



Melanoma: 111 dermoscopy images

Andre Esteva Brett Kunnel Roberto A Novoa Justin Ko Susan M Swetter Helen M Blau & Schastian Thrun 201 Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542, (7639), 115-118 human-centered.ai (Holzinger Group)

Key Challenges



 Remember: Medicine is an complex application domain – dealing most of the time with probable information!

Some challenges include:

• (a) defining hospital system architectures in terms of generic tasks such as diagnosis, therapy planning and monitoring to be executed for (b) medical reasoning in (a);

(c) patient information management with (d) minimum uncertainty.

 Other challenges include: (e) knowledge acquisition and encoding, (f) human-ai interface and ai-interaction; and (g) system integration into existing clinical legacy and proprietary environments, e.g. the enterprise hospital information system; to mention only a few.

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

01 Decision Support Systems

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

human-centered.ai (Holzinger Group)

2019 Machine Learning for Health 02



human-centered.ai (Holzinger Group)

2019 Machine Learning for Health 02

Decision Making is central in any (medical) work

HCAI &



human-centered.ai (Holzinger Group)

of Medical Informatics, 80, 4,

Example: Clinical Guidelines

2019 Machine Learning for Health 02

HCAI &

The Medical Domain and Decision Making



 400 BC Hippocrates (460-370 BC), father of western medicine:

Remembe

Medical Action =

Decision Making

Search Task in ${\mathcal H}$

Problem: Time (t)

- A medical record should accurately reflect the course of
- A medical record should indicate the probable cause of
- **1890** William Osler (1849-1919), father of modern western medicine
 - Medicine is a science of uncertainty and an art of probabilistic decision making
- Todav
 - Prediction models are based on data features, patient health status is modelled as high-dimensional feature vectors ...

human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02

Digression: Clinical Guidelines as DSS & Quality Measure HCAI *

- Clinical guidelines are systematically developed documents to assist doctors and patient decisions about appropriate care;
- In order to build DS, based on a guideline, it is **formalized** (transformed from natural language to a logical algorithm), and
- implemented (using the algorithm to program a DSS);
- To increase the quality of care, they must be linked to a process of care, for example:
 - "80% of diabetic patients should have an HbA1c below 7.0" could be linked to processes such as:
 - "All diabetic patients should have an annual HbA1c test" and
 - "Patients with values over 7.0 should be rechecked within 2 months."
- Condition-action rules specify one or a few conditions which are linked to a specific action, in contrast to narrative guidelines which describe a series of branching or iterative decisions unfolding over time.
- Narrative guidelines and clinical rules are two ends of a continuum of clinical care standards.

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

Medlock, S., Opondo, D., Eslami, S., Askari, M., Wierenga, P., de Rooij, S. E. & Abu-Hanna, A. (2011) LERM (Logical Elements Rule Method): A method for assessing and formalizing clinical rules for decision support. International Journal



and Gamut with patients' clinical and lab findings to arrive at the most likely diagnosis

Reeder, M. M. & Felson, B. 2003

radiology: comprehensive lists of

roentgen differential diagnosis, New

Reeder and Felson's gamuts in

York, Springer Verlag.



Gamut F-137 PHRENIC NERVE PARALYSIS OR DYSFUNCTION

COMMON

- 1. Iatrogenic (eg, surgical injury; chest tube; therapeu-
- tic avulsion or injection; subclavian vein puncture) Infection (eg, tuberculosis; fungus disease; abscess)
- Neoplastic invasion or compression (esp. carcinoma

UNCOMMON

- Aneurysm_g, aortic or other
 Birth trauma (Erb's palsy)
- Herpes zoster
 Neuritis, peripheral (eg. diabetic neuropathy)
- Neurologic disease_g (eg. hemiplegia; encephalitis; polio; Guillain-Barré S.)
- 7. Trauma

Example - Gamuts in Radiology



- COMMON

 1. Acute gastritis (eg. alcohol abuse)
- 2. Crohn's disease II II 3. Drugs (eg. aspirin 🖽 🛍 NSAID 🛍 steroids

- 6. (Normal areae gastricae 🖽
- 7 Pentic ulcer hyperacidity

UNCOMMON

- Corrosive gastritis 🖽

- 3. [Lymphorna]
 4. Opporturistic infection (eg. candidiasis (moniliasis) III. harpes simplex: cytomegalovirus)
- 5. Postoperative gastritis
- 6. Radiation therapy 7. Zollinger-Elison S. 11 multiple endocrine neoplasia (MEN) S.
- * Superficial erosions or aphthoid ulcerations seen especially with double contrast technique.

[] This condition does not actually cause the gamuted imaging finding, but can produce imaging changes that simulate it.

http://rfs.acr.org/gamuts/data/G-25.htm

Reeder, M. M. & Felson, B. (2003) Reeder

differential diagnosis. New York, Springer

and Felson's gamuts in radiology:

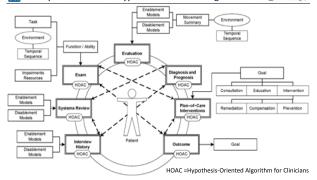
comprehensive lists of roentgen

human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02 2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

Iserson, K. V. & Moskop, J. C. 2007. Triage in Medicine, Part I: Concept, History, and Types. Annals of Emergency Medicine, 49, (3), 275-281

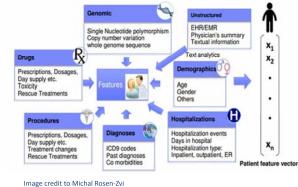
human-centered.ai (Holzinger Group)

19 Image Source: http://store.gomed.itech.com02



Schenkman, M., Deutsch, J. E. & Gill-Body, K. M. (2006) An Integrated Framework for Decision Making in Neurologic Physical Therapist Practice. *Physical Therapy*, 86, 12, 1681-1702.

human-centered.ai (Holzinger Group) 20 2019 Machine Learning for Health 02



mage or care to minima moon an

human-centered.ai (Holzinger Group) 21 2019 Machine Learning for Health 02

Example: Rheumatology





Chao, J., Parker, B. A. & Zvaifler, N. J. (2009) Accelerated Cutaneous Nodulosis Associated with Aromatase Inhibitor Therapy in a Patient with Rheumatoid Arthritis. *The Journal of Rheumatology, 36, 5, 1087-1088*.

human-centered.ai (Holzinger Group) 22 2019 Machine Learning for Health 02

Bone Changes ...



HCAI 1



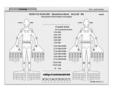
Ikari, K. & Momohara, S. (2005) Bone Changes in Rheumatoid Arthritis. New England Journal of Medicine, 353, 15, e13.

human-centered.ai (Holzinger Group) 23 2019 Machine Learning for Health 02

100+ clinical and functional parameter per Patient



- 50+ Patients per day ~
 5000 data points per day ...
- Aggregated with specific scores (Disease Activity Score, DAS)
- Current patient status is related to previous data
- = convolution over time



■ ⇒ time-series data

Simonic, K. M., Holzinger, A., Bloice, M. & Hermann, J. (2011). Optimizing Long-Term Treatment of Rheumatoid Arthritis with Systematic Documentation. Pervasive Health - 5th International Conference on Pervasive Computing Technologies for Healthcare, Dublin, IEEE, 550-554.

What Description of the Computing Technologies for Healthcare, Dublin, IEEE, 550-554.

III Gaining out Knowledge of time-series data



Simonic, K. M., Holzinger, A., Bloice, M. & Hermann, J. (2011). Optimizing Long-Term Treatment of Rheumatoid Arthritis with Systematic Documentation. Pervasive Health - 5th International Conference on Pervasive Computing Technologies for Healthcare, Dublin, IEEE, 550-552.

2019 Machine Learning for Health 02.
2019 Machine Learning for Health 02.

Can Computers help doctors to make better decisions?

For reading and discussion: Michael Duerr-Specht, Randy Goebel & Andreas Holzinger 2015. Medicine and Health Care as a Data Problem: Will Computers become better medical doctors? In: Holzinger, Andreas, Roecker, Carsten & Ziefle, Martina (eds.) Smart Health, State-of-the-Art SOTA Lecture Notes in Computer Science LNCS 8700. Heidelberg, Berlin, New York: Springer, pp. 21-40, doi:10.1007/978-3-319-16226-3_2.

!!! Computers help human doctors to make better decisions? **№ HCAI** ★

Reasoning Process	Human	Computer
Abductive Hypothesis generation	Uniquely capable of complex pattern recognition and creative thought. "the whole is greater than the sum of its parts"	Matches multiple individual correlations from extensive data banks based on preconceived algorithms. Secondary construction of relationships. "the whole equals the sum of its parts"
Inductive Symptom → Disease	Limited database. Subject to biases - Anchoring bias - Confirmation bias - Premature closure	Extensive database. Probability based on Bayesian statistics, no significant bias. Limitation based on available data.
Deductive Disease → Symptoms, Treatment	Limited database. Personal intuition and experience affect decision making.	Extensive database. Application of rules of evidence based medicine with potential biases.

Michael Duerr-Specht, Randy Goebel & Andreas Holzinger 2015. Medicine and Health Care as a Data Problem: Will Computers become better medical doctors? In: Holzinger, Andreas, Roecker, Carsten & Ziefle, Martina (eds.) Smart Health, State-of-the-Art SOTA Lecture Notes in Computer Science LNCS 8700. Heidelberg, Berlin, New York: Springer, pp. 21-40, doi:10.1007/978-3-319-16226-3_2.

for augmenting human

intellect

DR. D. C. ENGELBART a research center

Engelbart on a computer-based, interactive, multiconside diapital system which is being developed at Stanford Re-search Institute under the sponsorship of ARPA, NASA and RADC. The system is being used as an experimental sub-oratory for investigating principles by which interactive computer aids can augment intellectual capability. The sochriques which are being described will, themselves.

Superhuman AI for heads-up no-limit poker: Libratus beats top professional

Heimo Müller, Robert Reihs & Kurt Zatloukal 2017, Towards the Noam Brown & Tuomas Sandholm 2018, Superhuma Al for heads-up no-limit poker: Libratus beats top professionals. Science, 359, (6374), 418-424, doi:10.1126/science.aao1733

Source: https://web.stanford.edu/dept/SUL/library/extra4/sloan/mousesite/dce1968conferenceannouncement.jpg

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

W Two types of decisions (Diagnosis vs. Therapy)

HCAI 1

- Type 1 Decisions: related to the diagnosis, i.e. AI/ML is used to assist in diagnosing a disease on the basis of the individual patient data. Questions include:
 - What is the probability that this patient has a myocardial infarction on the basis of given data (patient history, ECG, ...)?
 - What is the probability that this patient has acute appendices, given the signs and symptoms concerning abdominal pain?
- Type 2 Decisions: related to therapy, i.e. AI/ML is used to select the best therapy on the basis of clinical evidence, e.g.:
 - What is the best therapy for patients of age x and risks y, if an obstruction of more than z % is seen in the left coronary artery?
 - What amount of insulin should be prescribed for a patient during the next 5 days, given the blood sugar levels and the amount of insulin taken during the recent weeks?

Jan H. Van Bemmel & Mark A. Musen 1997. Handbook of Medical Informatics, Heidelberg, Springer

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

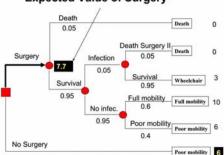
Instructors: Professor Lucila Ohno-Machado and Professor Staal Vinterbo

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

Helps to make rational decisions (risks vs. success)

HCAI 1

Expected Value of Surgery



nologist: Challenges of Explainable-Ai in Digital Patholog

Augmenting Human Doctors with Artificial Intelligence PHCAI &

Augmented Pathologist: Challenges of Explainable-Al in Digital

W Example: Knee Surgery of a Soccer Player

HCAI 1

2019 Machine Learning for Health 02





- Example of a Decision Problem
- Soccer player considering knee surgery
- Uncertainties:

human-centered.ai (Holzinger Group)

- Success: recovering full mobility
- Risks: infection in surgery (if so, needs another surgery and may loose more mobility)
- Survival chances of surgery

Harvard-MIT Division of Health Sciences and Technology HST.951J: Medical Decision Support, Fall 2005

Remember: Expected Utility Theory E(U|d)

HCAI 1

For a single decision variable an agent can select D = d for any $d \in dom(D)$.

The expected utility of decision D = d is

$$E(U \mid d) = \sum_{x_1,...,x_n} P(x_1,...,x_n \mid d) U(x_1,...,x_n,d)$$

An optimal single decision is the decision D = dmaxwhose expected utility is maximal:

$$d_{\max} = \arg \max_{d \in \text{dom}(D)} E(U \mid d)$$

John Von Neumann & Oskar Morgenstern 1944, Theory of games and economic behavior. Princeton, Princeton university press.

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group

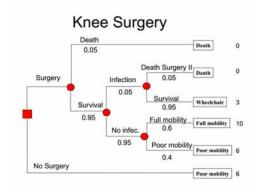
Decision Tree (this is known since Hippocrates!)

Pathologist level interpretable whole-slide diagnosis

HCAI :

Impuls - Metach Country

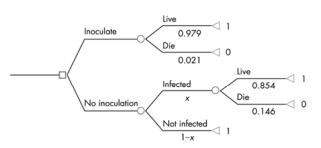
HCAI 1



2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

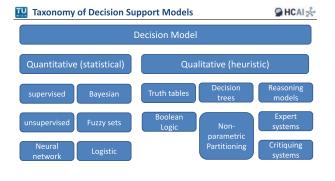
Clinical Decision Tree (CDT) is still state-of-the-art

HCAI 1



Ferrando, A., Pagano, E., Scaglione, L., Petrinco, M., Gregori, D. & Ciccone, G. (2009) A decisiontree model to estimate the impact on cost-effectiveness of a venous thromboembolism prophylaxis guideline. Quality and Safety in Health Care, 18, 4, 309-313.

2019 Machine Learning for Health 02 2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group) human-centered.ai (Holzinger Group)



Informatics. Heidelberg, Springer.

human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02

Extended by A. Holzinger after: Bemmel, J. H. v. & Musen, M. A. (1997) Handbook of Medical

Need for robust algorithms

 Need for trustworthy, fair and accountable algorithms

What makes decision support in health different?

- Augmenting the doctor not replacing them, but let "Chimpanzee"-Work do by algorithms
- Focus of the doctors to cognitively high-end demanding, challenging work
- Double-Check ("look at this corner, maybe there is something relevant)"
- Many of the questions of medical doctors need causal explanations "the why"!!

human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02

Learning Algorithm (Conf = 0.96) Himabindu Lakkaraju, Ece Kamar, Rich Caruana & Eric Horvitz, Identifying unknown unknowns in the open world:

Big chance for medicine Identifying Unknown Unknowns Phone HCAI

Representations and policies for guided exploration. Thirty-First AAAI Conference on Artificial Intelligence, 2017.

Example medical data sets openly available

HCAI &





HCAI 1

CAMELYON17

MYCIN

02 History of DSS = **History of Al**

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

Ultrashort history of Early AI

human-centered.ai (Holzinger Group



2019 Machine Learning for Health 02

- 1943 McCulloch, W.S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biology, 5, (4), 115-133, doi:10.1007/BF02459570.
- 1950 Turing, A.M. Computing machinery and intelligence. Mind, 59, (236), 433-460.
- 1958 John McCarthy Advice Taker: programs with common sense
- 1959 Samuel, A.L. Some studies in machine learning using the game of checkers. IBM Journal of research and development, 3, (3), 210-229, doi:10.1147/rd.33.0210.
- inexact reasoning in medicine. Mathematical biosciences, 23, (3-4), 351-379, doi:10.1016/0025-5564(75)90047-4.

human-centered.ai (Holzinger Group

DATA BASE

Knowledge of

the problem

Evolution of Decision Support Systems (Expert Systems)

BAOBAB

1970'5

human-centered.ai (Holzinger Group

Shortliffe, E. H. &

Rule-based expert

systems: the MYCIN

experiments of the

Stanford Heuristic Programming Project.

Addison-Wesley.

Buchanan, B. G. (1984)

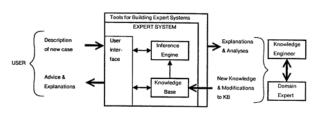
2019 Machine Learning for Health 02

TEIRESIAS - EMYCIN

Early Knowledge Based System Architecture

human-centered.ai (Holzinger Group)

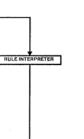
HCAI 1



Shortliffe, T. & Davis, R. (1975) Some considerations for the implementation of knowledge-based

about domain

CAPABILITY



HCAI 1

expert systems ACM SIGART Bulletin, 55, 9-12.

human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02

GUIDON

2019 Machine Learning for Health 02

2019 Machine Learning for Health 02

• 1975 Shortliffe, E.H. & Buchanan, B.G. 1975. A model of

■ 1978 Bellman, R. Can Computers Think? Automation of

Thinking, problem solving, decision-making ...

2019 Machine Learning for Health 02

🕎 Static Knowledge versus dynamic knowledge

entered by use USER made by system Shortliffe & Buchanan (1984 human-centered.ai (Holzinger Group

HCAI 1

MYCIN – rule based system - certainty factors

HCAI 1

Original Example from MYCIN

HCAI 1

- The information available to humans is often imperfect – imprecise - uncertain.
- This is especially in the medical domain the case.
- An human agent can cope with deficiencies.
- Classical logic permits only exact reasoning:
- IF A is true THEN A is non-false and IF B is false THEN B is non-true
- Most real-world problems do not provide this exact information, mostly it is inexact, incomplete, uncertain and/or un-measurable!

human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02 MYCIN is a rule-based Expert System, which is used for therapy planning for patients with bacterial infections

Goal oriented strategy ("Rückwärtsverkettung")

- To every rule and every entry a certainty factor (CF) is assigned, which is between 0 und 1
- Two measures are derived:
- MB: measure of belief
- MD: measure of disbelief
- Certainty factor CF of an element is calculated by: CF[h] = MB[h] - MD[h]
- CF is positive, if more evidence is given for a hypothesis, otherwise CF is negative
- CF[h] = +1 -> h is 100 % true
- CF[h] = -1 -> h is 100% false

human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02 h₁ = The identity of ORGANISM-1 is streptococcus h₂ = PATIENT-1 is febrile

h₃ = The name of PATIENT-1 is John Jones

CF[h₁,E] = .8 : There is strongly suggestive evidence (.8) that the identity of ORGANISM-1 is streptococcus

 $CF[h_2, E] = -.3$: There is weakly suggestive evidence (.3) that

PATIENT-1 is not febrile

 $CF[h_3, E] = +1$: It is definite (1) that the name of PATIENT-1 is

John Jones

Shortliffe, E. H. & Buchanan, B. G. (1984) Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project. Addison-Wesley.

human-centered.ai (Holzinger Group 2019 Machine Learning for Health 02

MYCIN was no success in the clinical routine

HCAI %



2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

However, Al was extremely popular in the 1970ies







Image credit to Bernhard Schölkopf

human-centered.ai (Holzinger Group

!!! Cybernetics was praised as the solution for everything

HCAI 1



Image credit to Bernhard Schölkopf

2019 Machine Learning for Health 02

HCAI 1

The AI winter was bitter cold ...

human-centered.ai (Holzinger Group)

James Hendler 2008. Avoiding another Al winter. IEEE

HCAI 1

From Al summer to Al summer

https://www.computer.org/csl/ mags/ex/2003/03/x3018.html

human-centered.ai (Holzinger Group)

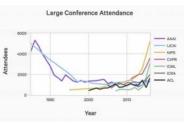


2019 Machine Learning for Health 02

https://aaai.org/Conferences/AAAI-20/ International Joint Conference on Artificial Intelligence:

https://ijcai20.org/

AAAI = AAAI Conference on Artificial Intelligence:



https://medium.com/machine-learning-in-practice/nips-accepted-papers-stats-26f124843aa0

03 Example: **P4-Medicine**

Intelligent Systems, 23, (2), 2-4, doi:10.1109/MIS.2008.20

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02

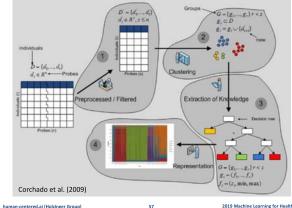
Kapur, K., Xing, Y., Ouyang, Z. & Wong, W. (2007) Exon arrays provide accurate assessments of gene expression. Genome Biology, 8, 5, R82.

human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02 2 - Exon array probe

Exon array structure. Probe design of exon arrays. (1) Exon—intron structure of a gene. Gray boxes represent introns, rest represent exons. Introns are not drawn to scale. (2) Probe design of exon arrays. Four probes target each putative exon. (3) Probe design of 30expression arrays. Probe target the 30end of mRNA sequence.

Corchado, J. M., De Paz, J. F., Rodriguez, S. & Bajo, J. (2009) Model of experts for decision support in the diagnosis of leukemia patients. Artificial Intelligence in Medicine, 46, 3, 179-200.

human-centered.ai (Holzinger Group)



2019 Machine Learning for Health 02

HCAI % ide 8-25 Computational leukemia cancer detection 3/6

A = acute, C = chronic, L = lymphocytic, M = myeloid

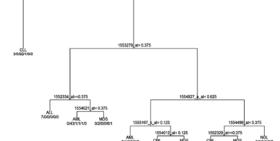
- . ALL = cancer of the blood AND bone marrow caused by an abnormal proliferation of lymphocytes.
- . AML = cancer in the bone marrow characterized by the proliferation of myeloblasts, red blood cells or abnormal platelets.
- . CLL = cancer characterized by a proliferation of lymphocytes in the bone marrow
- · CML = caused by a proliferation of white blood cells in the bone marrow.
- MDS (Myelodysplastic Syndromes) = a group of diseases of the blood and hone marrow in which the hone marrow does not produce a sufficient amount of healthy cells.
- · NOL (Normal) = No leukemias Corchado et al. (2009)

human-centered.ai (Holzinger Group)

2019 Machine Learning for Health 02

26 Computational leukemia cancer detection 4/6

Further Reading: Breiman, Friedman, Olshen, & Stone (1984). Classification and Regression Trees. Wadsworth, Belmont, CA.



2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

Corchado et al. (2009)

HCAI 1

HCAI 1

Computational leukemia cancer detection 6/6

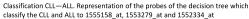
- The model of Corchado et al. (2009) combines:
- 1) methods to reduce the dimensionality of the original data set:
- 2) pre-processing and data filtering techniques;
- 3) a clustering method to classify patients; and
- 4) extraction of knowledge techniques
- The system reflects how human experts work in a lab, but
- 1) reduces the time for making predictions;
- 2) reduces the rate of human error; and
- 3) works with high-dimensional data from exon arrays

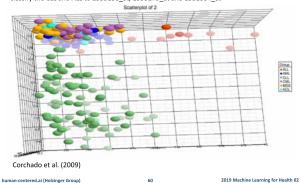
04 Example: **Case Based Reasoning** (CBR)

27 Computational leukemia cancer detection 5/6

HCAI %

HCAI 1





Slide 8-29 Thinking – Reasoning – Deciding – Acting

Critical Thinking,

Clinical Reasoning,

Clinical Judgment

Critical Thinking

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group

HCAI 1

reasoning: Foundational issues, methodological variations, and system approaches. AI Communications, 7, 1, 39-59.

human-centered.ai (Holzinger Group)

human-centered.ai (Holzinger Group)

Aamodt, A. & Plaza, E. (1994) Case-based

human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02

Aamodt & Plaza (1994)

human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02 Slide 8-32 CBR Example: Radiotherapy Planning 1/6

human-centered.ai (Holzinger Group)

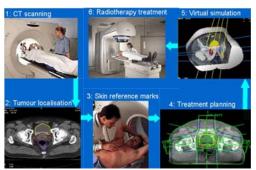
2019 Machine Learning for Health 02

IIIIIide 8-33 CBR Example: Radiotherapy Planning 2/6

HCAI %

2019 Machine Learning for Health 02

2019 Machine Learning for Health 02



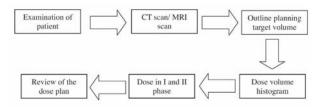
Source: Imaging Performance Assessment of CT Scanners Group, http://www.impactscan.org

Slide 8-34 CBR Example: Radiotherapy Planning 3/6

Slide 8-31 The task-method decomposition of CBR

HCAI 1

HCAI &



Measures:

- 1) Clinical Stage = a labelling system
- 2) Gleason Score = grade of prostate cancer = integer between 1 to 10; and
- Prostate Specific Antigen (PSA) value between 1 to 40
- 4) Dose Volume Histogram (DVH) = pot. risk to the rectum (66, 50, 25, 10 %)

Petrovic, S., Mishra, N. & Sundar, S. (2011) A novel case based reasoning approach to radiotherapy planning. Expert Systems With Applications, 38, 9, 10759-10769.

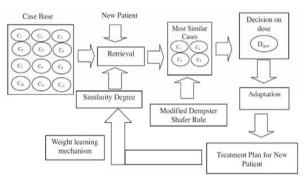
human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02

Slide 8-35 CBR System Architecture 4/6



HCAI 1

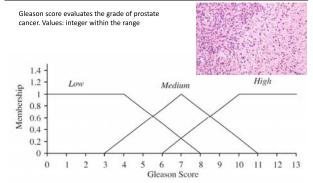
HCAI &



Petrovic, S., Mishra, N. & Sundar, S. (2011) A novel case based reasoning approach to radiotherapy planning. Expert Systems With Applications, 38, 9, 10759-10769.

human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02

III Slide 8-36 Membership funct. of fuzzy sets Gleason score 5 ← HCAI ★



Petrovic, S., Mishra, N. & Sundar, S. (2011) A novel case based reasoning approach to radiotherapy planning. Expert Systems With Applications, 38, 9, 10759-10769.

Slide 8-37 Case Based Reasoning 6/6 HCAI & Petrovic et al. (2011) Dose plan suggested by Dempster-Shafer rule (62Gy+10Gy) Dose received by 10% of rectum is 56.02 Gy (maximum dose limit =55 Gy) Feasible dose plan Modification Proposed dose plan Modification of dose plan: New dose plan: 62Gy +8 Gy Dose received by 10% of rectum is: 54.26 Gy (feasible dose plan)

05 Causal Reasoning

human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02



Decide if X → Y, or Y → X using only observed data

HCAI 1

Remember: Reasoning = "Sensemaking"

of premises: A=B, B=C, therefore A=C

from an incomplete set of preconditions.

Deductive Reasoning = Hypothesis > Observations > Logical

Inductive reasoning = makes broad generalizations from

DANGER: Hypothesis must be correct! DR defines whether the truth

• DANGER: allows a conclusion to be false if the premises are true

generate hypotheses and use DR for answering specific questions

possible premises that, if true also, may support the conclusion,

Therefore, it might have rained." This kind of reasoning can be used

to develop a hypothesis, which in turn can be tested by additional

• Abductive reasoning = inference = to get the best explanation

Given a true conclusion and a rule, it attempts to select some

• Example: "When it rains, the grass gets wet. The grass is wet.

High dimensionality (curse of dim., many factors contribute)

Little prior data (no mechanistic models of the data)

• *) = Def.: a sequence or collection of random variables is

Complexity (real-world is non-linear, non-stationary, non-IID *)

Sören Sonnenburg, Gunnar Rätsch, Christin Schaefer & Bernhard Schölkoof 2006, Large scale multiple kernel learning, Journal o

independent and identically distributed if each random variable has

the same probability distribution as the others and all are mutually

of a conclusion can be determined for that rule, based on the truth

HCAI 1

"How do humans generalize from few examples?"

- Learning relevant representations
- Disentangling the explanatory factors
- Finding the shared underlying explanatory factors, in particular between P(x) and P(Y|X), with a causal link between $Y \rightarrow X$

Bengio, Y., Courville, A. & Vincent, P. 2013. Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence, 35, {8}, 1798-1828, doi:10.1109/TPAMI.2013.50. Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. 2011. How to grow a mind: Statistics, structure, and abstraction. Science, 331, (6022), 1279-1285, doi:10.1126/science.119278

human-centered.ai (Holzinger Group)

2019 Machine Learning for Health 02

Learning Research,

17, (1), 1103-1204.

human-centered.ai (Holzinger Group)





 $\mathbb{P}_{Y \mid s} \neq \mathbb{P}_{Y \mid \text{do}(x), s} = \mathbb{P}_{Y \mid x, s}$ $\mathbb{P}_{X \mid s} \neq \mathbb{P}_{X \mid do(y),s} = \mathbb{P}_{X \mid y,s}$

human-centered.ai (Holzinger Group

Conclusions

specific observations

though not uniquely.

reasoning or data.

Remember: hard inference problems

Need of large top-quality data sets

independent

2019 Machine Learning for Health 02

2019 Machine Learning for Health 02

HCAI &

HCAI 1

Important Definition: Ground truth



- := information provided by direct observation (empirical evidence) in contrast to information provided by inference
 - Empirical evidence = information acquired by observation or by experimentation in order to verify the truth (fit to reality) or falsify (non-fit to reality).
 - Empirical inference = drawing conclusions from empirical data (observations, measurements)
 - Causal inference = drawing a conclusion about a causal connection based on the conditions of the occurrence of an effect.
 - Causal inference is an example of causal reasoning.

human-centered.ai (Holzinger Group)

2019 Machine Learning for Health 02



Gottfried W. Leibniz (1646-1716) Hermann Weyl (1885-1955) Vladimir Vapnik (1936-) Alexey Chervonenkis (1938-2014) Gregory Chaitin (1947-)

human-centered.ai (Holzinger Group)

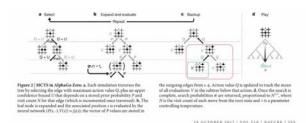


2019 Machine Learning for Health 02

HCAI &

Mastering the game of Go without human knowledge





 $(p, v) = f_{\theta}(s)$ and $l = (z - v)^2 - \pi^T \log p + c \|\theta\|^2$

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre George Van Den Driessche, Thore Graepel & Demis Hassabis 2017. Mastering the game of go without human knowledge, Nature, 550, (7676), 354-359, doi:doi:10.1038/nature24270

human-centered.ai (Holzinger Group)



David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel & Demis Hassabis 2016. Mastering the game of Go with deep neural networks and tree search. Nature, 529, (7587), 484-489, doi:10.1038/nature16961

06 Explainability

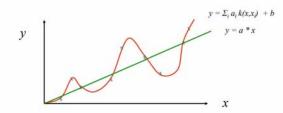
 $\mathbb{P}_X = \mathbb{P}_{X \mid \mathrm{do}(y)} \neq \mathbb{P}_{X \mid y}$ $\mathbb{P}_X \neq \mathbb{P}_{X \mid do(y)} = \mathbb{P}_{X \mid y}$ Joris M. Mooij, Jonas Peters, Dominik Janzing, Jakob $\mathbb{P}_{Y} \neq \mathbb{P}_{Y \mid do(x)} \neq \mathbb{P}_{Y \mid x}$ $\mathbb{P}_Y = \mathbb{P}_{Y \mid do(x)} = \mathbb{P}_{Y \mid x}$ Zscheischler & Bernhard Schölkopf $\mathbb{P}_X \neq \mathbb{P}_{X \mid do(y)} \neq \mathbb{P}_{X \mid y}$ $\mathbb{P}_X = \mathbb{P}_{X \mid do(y)} = \mathbb{P}_{X \mid y}$ 2016. Distinguishing cause from effect using observational data: methods and benchmarks. The Journal of Machine

 $\mathbb{P}_Y = \mathbb{P}_{Y \mid do(x)} \neq \mathbb{P}_{Y \mid x}$

 $\mathbb{P}_X = \mathbb{P}_{X \mid do(y)} \neq \mathbb{P}_{X \mid y}$

Empirical Inference Example

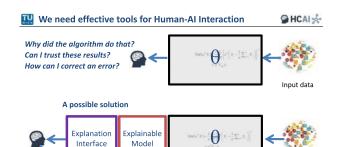




human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

2019 Machine Learning for Health 02

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group



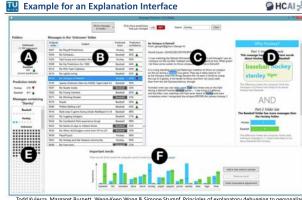
The domain expert can understand why ... The domain expert can learn and correct errors ... The domain expert can re-enact on demand ...

(3)

human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02

1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 0.533 0.894 0.620 0.699 0.629 0.546 0.540 1.000 0.526 1.000 0.522 0.483 0.471 1.000 0.522 0.576 0.658 0.639 0.615 0.748 0.639 0.911 0.796 0.647 0.614 0.529 0.553 0.588 0.651 0.644 0.585 0.433 0.606 0.588 0.467 0.313 0.363 0.349 0.415 0.578 0.512 0.305 0.274 0.256 0.569 0.661 0.486 0.605 0.448 0.494 0.705 0.730 0.579 0.532 0.526 0.623 0.518 0.387 0.310 0.338 0.466 0.378 0.559 0.479 0.444 0.430 0.494 0.405 0.232 0.248 0.23 0.590 (66) (346) (367) (344) (349) (347) (379) (379) (379) (332) (335) (335) (335) (335) (335) (336) (347) (335) (344) (345) (344) (345) (344) (345) (344) (345) (344) (345) (344) (345) (344) (345) (344) (345) (0.461.0503.0513.0.432.0.537.0.537.0.467.0.530.0.387.0.504.0.353.0.362.0.456.0.222.0.241.0.342.0.510.0.622.0.454.0.441.0.285.0.218.0.545.0.502.0.445.0.508.0.623.0.529.0.464.0.455.0.824.0.470.0.213.0.549.0.569.0.522.0.500.0.493.0.529.0.422.0.210.0.242.0.281.0.309.0.295.0.241.0.213.0.549.0.569.0.522.0.500.0.493.0.5290.383 0.458 0.482 0.370 0.384 0.361 0.400 0.391 0.320 0.319 0.425 0.377 0.433 0.528 0.497 0.285 0.247 0.198 0.226 0.410 0.570 0.597 0.576 0.588 0.531 0.493 0.546 0.459 0.476 0.391 0.431 0.563 0.321 0.364 0.382 0.365 0.368 0.405 0.287 0.263 0.509 0.666 0.569 0.509 0.554 0.551 0.591 0.622 0.647 0.612 0.648 0.594 0.537 0.546

human-centered.ai (Holzinger Group)



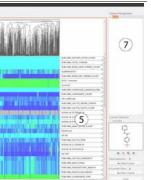
Todd Kulesza, Margaret Burnett, Weng-Keen Wong & Simone Stumpf. Principles of explanatory debugging to personalize interactive machine learning. Proceedings of the 20th International Conference on Intelligent User Interfaces (IUI 2015), 2015 Atlanta. ACM, 126-137, doi:10.1145/2678025.2701399.

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)







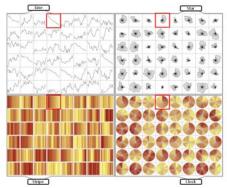


Werner Sturm, Till Schaefer, Tobias Schreck, Andeas Holzinger & Torsten Ullrich, Extending the Scaffold Hunter Visualization Toolkit with Interactive Heatmaps In: Borgo, Rita & Turkay, Cagatay, eds. EG UK Computer Graphics & Visual Computing CGVC 2015, 2015 University College London (UCL). Euro Graphics (EG), 77-84, doi:10.2312/cgvc.20151247. human-centered.ai (Holzinger Group) 85 2019 Machine Learning for Health 02

What is understandable, interpretable, intelligible?



HCAI &

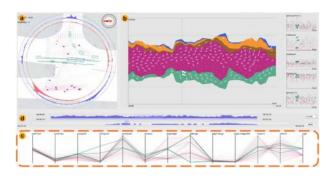


https://www.vis.uni-konstanz.de/en/members/fuchs/ 2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

Explainable AI is a huge challenge for visualization



HCAI &



2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group

Kandinsky Patterns: A possible IQ-Test for machines ...







Methods of ex-Al





- 1) Gradients
- 2) Sensitivity Analysis
- 3) Decomposition Relevance Propagation (Pixel-RP, Layer-RP, Deep Taylor Decomposition, ...)
- 4) Optimization (Local-IME model agnostic, BETA transparent approximation, ...)
- 5) Deconvolution and Guided Backpropagation
- 6) Model Understanding
 - Feature visualization, Inverting CNN
 - Qualitative Testing with Concept Activation Vectors TCAV
 - Network Dissection

07 Methods of **Explainable AI**

Heimo Müller & Andreas Holzinger 2019. Kandinsky Patterns. arXiv:1906.00657.

2019 Machine Learning for Health 02 2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group) human-centered.ai (Holzinger Group) human-centered.ai (Holzinger Group)

Andreas Holzinger LV 706.315 From explainable AI to Causability, 3 ECTS course at Graz University of Technology

https://human-centered.ai/explainable-ai-causability-2019 (course given since 2016)

2019 Machine Learning for Health 02

Causability:=

Explainability :=

a property of a system

a property of a person

("the Human explanation)

("the AI explanation)

HCAI %

Our goal is to provide interfaces for effective mapping

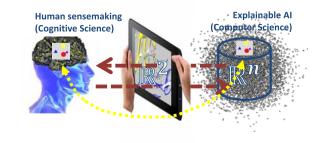
HCAI &

HCAI %

HCAI 1

Causability := a property of a person (Human)

Explainability := a property of a system (Computer)



human-centered.ai (Holzinger Group) 92 2019 Machine Learning for Health 02

Gradients > Sensitivity Analysis > Heatmapping

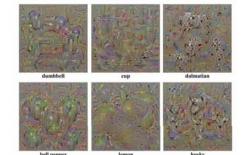
Cost City

| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost City
| Cost

Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh & Dhruv Batra. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. ICCV, 2017. 618-626. human-centered.al (Holzinger Group) 2019 Machine Learning for Health

Gradients

human-centered.ai (Holzinger Group)



Karen Simonyan, Andrea Vedaldi & Andrew Zisserman 2013. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv:1312.6034.

human-centered.ai (Holzinger Group) 95 2019 Machine Learning for Health 02

Gradients

human-centered.ai (Holzinger Group)

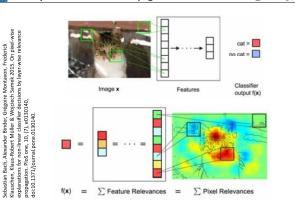
HCAI &

2019 Machine Learning for Health 02

LRP Layer-Wise Relevance Propagation

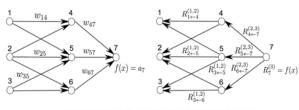
HCAI &

2019 Machine Learning for Health 02



https://www.khanacademy.org/math/multivariable-calculus/multivariable-derivatives/partial-derivative-and-gradient-articles/a/the-gradient

A NN-classifier during prediction time





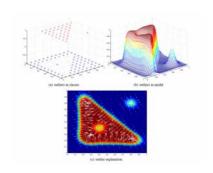
Sebastian Bach, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller & Wojciech Samek 2015. On piech-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PloS one, 10, (7), e0130140, 610:10.1371/journal.pone.0.130140.

Gradients

Gradients



HCAI %

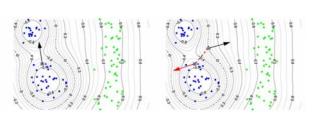


David Baehrens, Timon Schroeter, Stefan Harmeling, Motoaki Kawanabe, Katja Hansen & Klaus-Robert Mueller 2010. How to explain individual classification decisions. Journal of machine learning research (JMLR), 11, (6), 1803-1831.

human-centered.ai (Holzinger Group) 96 2019 Machine Learning for Health 02

Example Taylor Decomposition





Sebastian Bach, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller & Wojciech Samek 2015. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PloS one, 10, (7), e0130104, 0oi:10.1371/journal.pone.0130140.

human-centered.ai (Holzinger Group) 97 2019 Machine Learning for Health 02

human-centered.ai (Holzinger Group)

20

2019 Machine Learning for Health 02

relevance detected by the model:

heatmap are greater or equal to zero, that is:

 $\forall x: f(x) = \sum R_p(x).$

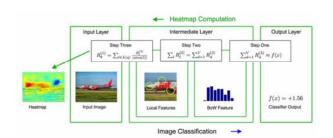
 $\forall x, p: R_p(x) \geq 0$

doi:10.1016/j.patcog.2016.11.008.

Relevance Redistribution

human-centered.ai (Holzinger Group)

and 2.



Sebastian Bach, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller & Wojciech Samek 2015. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PloS one, 10, (7), e0130140, doi:10.1371/journal.pone.0130140.

human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02

 $f(x) = \max(0, x_1) + \max(0, x_2)$

Sebastian Bach, Alexander Binder, Grégoire Montayon, Frederick Klauschen, Klaus-Robert Müller & Woiciech Samek 2015. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PloS one, 10, (7), e0130140, doi:10.1371/journal.pone.0130140.

tered.ai (Holzinger Group) 2019 Machine Learning for Health 02

Definition 1. A heatmapping R(x) is conservative if the sum of

assigned relevances in the pixel space corresponds to the total

Definition 2. A heatmapping R(x) is *positive* if all values forming the

Definition 3. A heatmapping R(x) is *consistent* if it is conservative and positive. That is, it is consistent if it complies with Definitions 1

Gregoire Montavon, Sebastian Lapuschkin, Alexander Binder, Wojciech Samek & Klaus-Robert Müller 2017. Explaining

nonlinear classification decisions with deep taylor decomposition. Pattern Recognition, 65, 211-222,

Sensitivity Analysis vs. Decomposition

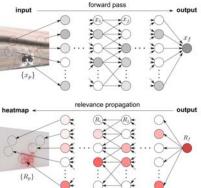


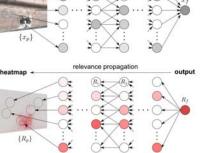




2019 Machine Learning for Health 02

HCAI &





2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

sensitivity analysis: $(\partial f/\partial x_1)^2 = 1_{x_1>0}$ $(\partial f/\partial x_2)^2 = 1_{x_2>0}$

 $R_1(x) = \max(0, x_1)$ $R_2(x) = \max(0, x_2)$

human-centered.ai (Holzinger Group)

human-centered.ai (Holzinger Group)

2019 Machine Learning for Health 02

2019 Machine Learning for Health 02

HCAI %

Example 1

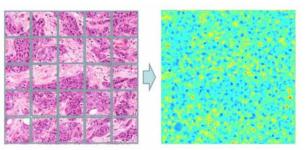




Example 2 Histopathology

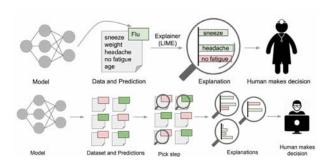
human-centered.ai (Holzinger Group)

human-centered.ai (Holzinger Group)



Sebastian Bach, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller & Wojciech Samek 2015. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PloS one, 10, (7), e0130140, doi:10.1371/journal.pone.0130140.

LIME – Local Interpretable Model Agnostic Explanations



Marco Tulio Ribeiro, Sameer Singh & Carlos Guestrin. Why should i trust you?: Explaining the predictions of any classifier. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

Deep Taylor (CaffeNet)

2019 Machine Learning for Health 02

2019 Machine Learning for Health 02



human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02

Classifiers Statistical Structural Naïve Bayes Rule based Distance based Neural Networks Decision Multi-Laver Functional Nearest Neighbor

https://stats.stackexchange.com/questions/271247/machine-learning-statistical-vs-structural-classifiers
nan-centered.ai (Holzinger Group) 110 2019 Machine Learning for Health 02

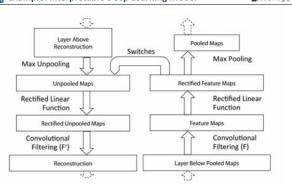
If Age <50 and Male =Yes: If Past-Depression =Yes and Insomnia =No and Melancholy =No, then Healthy If Past-Depression =Yes and Insomnia =Yes and Melancholy =Yes and Tiredness =Yes, then Depression If Age ≥ 50 and Male =No: If Family-Depression =Yes and Insomnia =No and Melancholy =Yes and Tiredness =Yes, then Depression If Family-Depression ≡No and Insomnia ≡No and Melancholy ≡No and Tiredness ≡No, then Healthy Default If Past-Depression =Yes and Tiredness =No and Exercise =No and Insomnia =Yes, then Depression If Past-Depression =No and Weight-Gain =Yes and Tiredness =Yes and Melancholy =Yes, then Depression If Family-Depression =Yes and Insomnia =Yes and Melancholy =Yes and Tiredness =Yes, then Depression

Himabindu Lakkaraju, Ece Kamar, Rich Caruana & Jure Leskovec 2017. Interpretable and Explorable Approximations of Black Box Models. arXiv:1707.01154.

human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02

Example: Interpretable Deep Learning Model





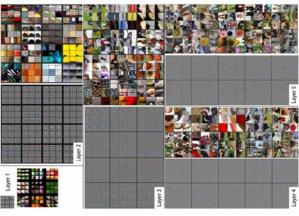
Matthew D. Zeiler & Rob Fergus 2013. Visualizing and Understanding Convolutional Networks. arXiv:1311.2901

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

HCAI & W Visualizing a Conv Net with a De-Conv Net image size 224 filter size 7 Input Image

Matthew D. Zeiler & Rob Fergus 2014. Visualizing and understanding convolutional networks. In: D., Fleet, T., Paidla, B., Schiele & T., Tuytelaars (eds.) ECCV, Lecture Notes in Computer Science LNCS 8689. Cham: Springer, pp. 818-833, doi:10.1007/978-3-319-10590-1_53.

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)



Matthew D. Zeiler & Rob Fergus 2013. Visualizing and Understanding Convolutional Networks. arXiv:1311.2901.

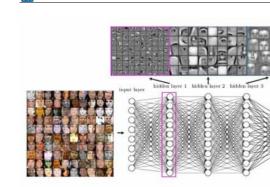
2019 Machine Learning for Health 02

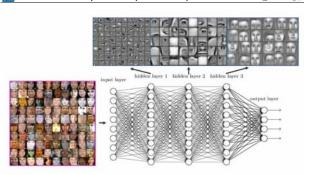
The world is compositional (Yann LeCun)



The world is compositional (Yann LeCun)



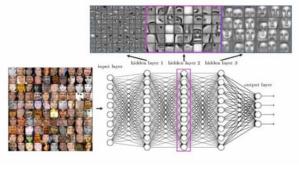




Matthew D. Zeiler & Rob Fergus 2013. Visualizing and Understanding Convolutional Networks. arXiv:1311.2901

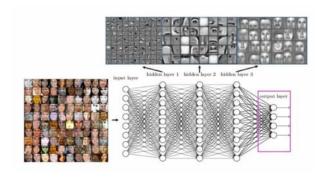
2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02 2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group) human-centered.ai (Holzinger Group)

HCAI &



2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)



2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

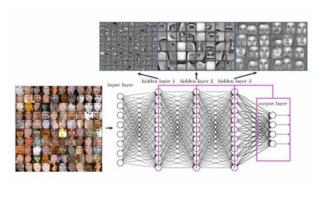
HCAI %

III TCAV Testing with Concept Activation Vectors

HCAI %

Stochastic AND-OR Templates for visual objects

HCAI %



2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

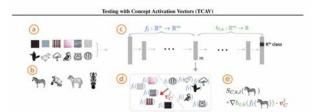
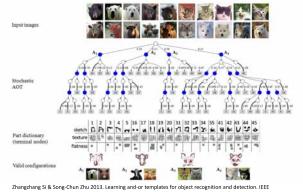


Figure 1. Testing with Concept Activation Vectors: Given a user-defined set of examples for a concept (e.g., 'striped'), and random examples (a), labeled training-data examples for the studied class (zebras) (b), and a trained network (c), TCAV can quantify the model's sensitivity to the concept for that class. CAVs are learned by training a linear classifier to distinguish between the activations produced by a concept's examples and examples in any layer (§). The CAV is the vector entilogenal to the classification boundary (e/c, red arrow). For the class of interest (zelevals), TCAV uses the directional derivative Sc_{xx}(x) to quantify conceptual sensitivity of

https://github.com/tensorflow/tcav

Been Kim, Martin Wattenberg, Justin Gilmer, Carrie Cai, James Wexler & Fernanda Viegas. Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (TCAV). International Conference on Machine Learning (ICML), 2018, 2673-2682.

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)



transactions on pattern analysis and machine intelligence, 35, (9), 2189-2205, doi:10.1109/TPAMI.2013.35.

123 2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

🔛 Stochastic Model on AND-OR graph: Zhaoyin Jia (2009) 🛮 😭 HCAI 🧩

Framework for vision: AND-OR Graphs

O and-node or-node

leaf node

human-centered.ai (Holzinger Group)





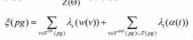
- Top-down/bottom-up computation
- Generalization of small sample
- Use Monte Carlos simulation to synthesis more configurations
- Fill semantic gap

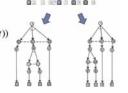
Images credit to Zhaoyin Jia (2009)

2019 Machine Learning for Health 02

HCAI & Stochastic Model on AND-OR graph: Zhaoyin Jia (2009)

- ▶ Terminal (leaf) node: T(pg)
- ▶ And-Or node: V^{or}(pg),V^{and}(pg)
- ▶ Set of links: *E*(*pg*)
- Switch variable at Or-node: w(t)



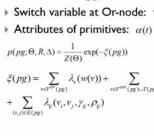


▶ Set of links: E(pg) Switch variable at Or-node: w(t) Attributes of primitives: α(t) $p(pg;\Theta,R,\Delta) = \frac{1}{Z(\Theta)} \exp(-\xi(pg))$

SCFG: weigh the frequency at the children of or-nodes

▶ Terminal (leaf) node: T(pg)

And-Or node: Vor (pg), Varid (pg)



human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group) 2019 Machine Learning for Health 02

And-Or node: V^{or}(pg), V^{and}(pg) ▶ Set of links: *E(pg)* Switch variable at Or-node: w(t) • Attributes of primitives: $\alpha(t)$ $p(pg;\Theta,R,\Delta) = \frac{1}{Z(\Theta)} \exp(-\xi(pg))$ 0--0 $\xi(pg) = \sum_{v \in V^{Or}(pg)} \lambda_v(w(v)) + \sum_{v \in V^{Ord}(pg) \cup I}$ $\lambda_t(\alpha(t))$

Weigh the local compatibility of primitives (geometric and appearance)

human-centered.ai (Holzinger Group)

 $+ \sum_{(i,j) \in E(pg)} \lambda_y(v_i,v_j,\gamma_y,\rho_y)$

2019 Machine Learning for Health 02

Stochastic Model on AND-OR graph: Zhaoyin Jia (2009)

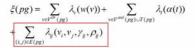
▶ Terminal (leaf) node: T(pg) And-Or node: Vor (pg), Vond (pg)

▶ Set of links: *E*(*pg*)

Switch variable at Or-node: w(t)

• Attributes of primitives: $\alpha(t)$

$$p(pg;\Theta,R,\Delta) = \frac{1}{Z(\Theta)} \exp(-\xi(pg))$$



Spatial and appearance between primitives (parts or objects)

human-centered.ai (Holzinger Group)

▶ Terminal (leaf) node: T(pg)

🔛 Stochastic Model on AND-OR graph: Zhaoyin Jia (2009) 🛮 😭 HCAI 🦟

And-Or node: Vor (pg), Vand (pg)

▶ Set of links: *E*(*pg*)

Switch variable at Or-node: w(t)

Attributes of primitives: α(t)

$$p(pg;\Theta,R,\Delta) = \frac{1}{Z(\Theta)} \exp(-\xi(pg))$$

$$\begin{split} \xi(pg) &= \sum_{v \in F^{ob}(pg)} \lambda_{v}(w(v)) + \sum_{v \in F^{obs}(pg) \cup T(pg)} \lambda_{t}(\alpha(t)) \\ &+ \sum_{(i,j) \in E(pg)} \lambda_{ij}(v_{i}, v_{j}, \gamma_{ij}, \rho_{ij}) \end{split}$$

human-centered.ai (Holzinger Group)

2019 Machine Learning for Health 02

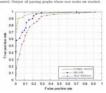
Stochastic graph grammar/comp. object representation Graph Grammar/comp.



HCAI 1

I Schedule the next node A to visit from the

- 2 Call Bottom-up(A) to update A's open list.



Liang Lin, Tianfu Wu, Jake Porway & Zijian Xu 2009, A stochastic graph grammar for compositional object representation and recognition. Pattern Recognition, 42, (7), 1297-1307, doi:10.1016/j.patcog.2008.10.033.

2019 Machine Learning for Health 02 entered.ai (Holzinger Group)

HCAI &

2019 Machine Learning for Health 02

HCAI &

Future Work

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

Explanations in Artificial Intelligence will be necessary

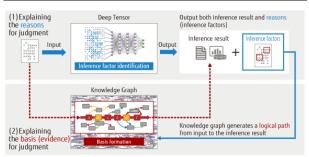


- Justification, Explanation and Causality
- Trust > scaffolded by justification of actions (why)
- Please always take into account the inherent uncertainty and incompleteness of medical data!

Alex John London 2019, Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability. Hastings Center Report, 49, (1), 15-21, doi:10.1002/hast.973.

Combination of Deep Learning with Ontologies





Explainable AI with Deep Tensor and Knowledge Graph

http://www.fujitsu.com/jp/Images/artificial-intelligence-en_tcm102-3781779.png

2019 Machine Learning for Health 02 human-centered.ai (Holzinger Group)

IBM is doing it now: teaching meaningful explanations

HCAI 1

Teaching Meaningful Explanations

Soel C. F. Codella,* Michael Hind,* Karthikeyar Murray Campbell, Amit Dhurandhar, Kush R. Aleksandra Mojsilović

What is a good explanation?

- (obviously if the other did understand it)
- Experiments needed!

Seemingly trivial questions ...?

- What is explainable/understandable/intelligible?
- When is it enough (Sättigungsgrad you don't need more explanations – enough is enough)
- But how much is it ...

Noel C.F. Codella, Michael Hind, Karthikeyan Natesan Ramanurty, Murray Campelle, Mnit Dhurandhar, Kush R. Varsihney, Dennis Wei & Aleksandra Mojsilovic 2018. Teaching Meaningful Explanations. arXiv:1805.11648.



human-centered.ai (Holzinger Group)

2019 Machine Learning for Health 02

2019 Machine Learning for Health 02

HCAI 1

2019 Machine Learning for Health 02

The second wave of AI (1975 –): Statistical Learning

HCAI %



Image credit to John Launchbury

- Engineers create learning models for specific tasks and train them with "big data" (e.g. Deep Learning)
- Advantage: works well for standard classification tasks and has prediction capabilities
- Disadvantage: No contextual capabilities and minimal reasoning abilities

human-centered.ai (Holzinger Group)

2019 Machine Learning for Health 02

HCAI 1

Thank you!



This is compatible to interactive machine learning

what no human is able to see

•Computational approaches can find in \mathbb{R}^n

the context and bring in experience,

■Black box approaches can not explain

expertise, knowledge, intuition, ...

WHY a decision has been made ...

However, still there are many hard problems

where a human expert in \mathbb{R}^2 can understand



Image credit to John Launchbury

- A contextual model can perceive, learn and understand and abstract and reason
- Advantage: can use transfer learning for adaptation on unknown unknowns
- Disadvantage: Superintelligence ...

human-centered.ai (Holzinger Group)

Three (selected) dangers and myths about AI generally AICAI

■ Myth 1a: Superintelligence by 2100 is inevitable!

🚻 The fist wave of AI (1943-1975): Handcrafted Knowledge 🛭 😭 HCAI 🦟

Engineers create a set of logical rules to represent

Advantage: works well in narrowly defined problems

Disadvantage: No adaptive learning behaviour and

knowledge (Rule based Expert Systems)

of well-defined domains

poor handling of p(x)

human-centered.ai (Holzinger Group

Perceiving

Learning Abstracting

Reasoning

Image credit to John Launchbury

- Myth 1b: Superintelligence by 2100 is impossible!
- Fact: We simply don't know it!
- Myth 2: Robots are our main concern Fact: Cyberthreats are the main concern: it needs no body - only an Internet connection
- Myth 3: Al can never control us humans Fact: Intelligence is an enabler for control: We control tigers by being smarter ...