From Clinical Decision Support to Causal Reasoning and Explainable AI

Keywords
- Decision support system (DSS)
- MYCIN – Rule Based Expert System
- GAMUTS in Radiology
- 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

Agenda
- 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
- 06 Explainability – Causability
- 07 (Some) Methods of Explainable AI

Reflection from last lecture

How do you explain this...

00 Reflection

Key Challenges
- Remember: Medicine is a complex application domain – dealing most of the time with probabilistic 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–machine interface and interaction; and (g) system integration into existing clinical legacy and proprietary environments, e.g., the enterprise hospital information system; to mention only a few.
01 Decision Support Systems

Remember: Medical Action = Decision Making
Search Task in H
Problem: Time (t)

Decision Making is central in any (medical) work

The Medical Domain and Decision Making
- 400 BC Hippocrates (460-370 BC), father of western medicine:
  - A medical record should accurately reflect the course of a disease
  - A medical record should indicate the probable cause of a disease
- 1890 William Osler (1849-1919), father of modern western medicine
  - Medicine is a science of uncertainty and an art of probabilistic decision making

Today
Prediction models are based on data features, patient health status is modelled as high-dimensional feature vectors...

Example: Clinical Guidelines

Example: Triangulation to find diagnoses

Example - Gamuts in Radiology

Search in an arbitrarily high-dimensional space < 5 min!
Augmenting Human Capabilities: an old human dream?

Augmenting Human Doctors with Artificial Intelligence

Pathologist level interpretable whole-slide diagnosis

Two types of decisions (Diagnosis vs. Therapy)

- **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 risk y, if an obstruction of more than 2 % 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?

Example: Knee Surgery of a Soccer Player

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

Decision Tree (this is known since Hippocrates!)

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

Helps to make rational decisions (risks vs. success)

Expected Value of Surgery

For a single decision variable an agent can select

\[ D = d \text{ for any } d \in \text{dom}(D) \]

The expected utility of decision \( D = d \) is

\[ E(U | d) = \sum_{x_1, \ldots, x_n} P(x_1, \ldots, x_n | d)U(x_1, \ldots, x_n, d) \]

An optimal single decision is the decision \( D = d_{\text{max}} \) whose expected utility is maximal:

\[ d_{\text{max}} = \arg \max d E(U | d) \]
02 History of DSS = History of AI

- 1958 John McCarthy Advice Taker: programs with common sense.
- 1978 Bellman, R. Can Computers Think? Automation of Thinking, problem solving, decision-making ...
Dealing with uncertainty in the real world

- 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 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!

MYCN – rule based system - certainty factors

- MYCN 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 and 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

Original Example from MYCN

- \( h_1 = \text{The identity of ORGANISM-1 is streptococcus} \)
- \( h_2 = \text{PATIENT-1 is febrile} \)
- \( h_3 = \text{The name of PATIENT-1 is John Jones} \)
- \( CF[h_1,E] = 0.8 \) : There is strongly suggestive evidence (.8) that the identity of ORGANISM-1 is streptococcus
- \( CF[h_2,E] = -0.3 \) : 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


03 Example: P4-Medicine
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)**
Slide 8-30 Case Based Reasoning (CBR) Basic principle


Slide 8-31 The task-method decomposition of CBR

Aamodt & Plaza (1994)

Slide 8-32 CBR Example: Radiotherapy Planning 1/6

Source: http://www.terradigitalacademic.co.uk

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


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


Slide 8-35 CBR System Architecture 4/6


Slide 8-36 Membership function of fuzzy sets: Gleason score

Gleason score evaluates the grade of prostate cancer. Values: integer within the range of 2-12.

Slide 8-37 Case Based Reasoning 6/6

Petropoulos, S. (2011)

Dose plan suggested by Dempster-Shafer rule (60Gy+10Gy)

Dose received by 10% of patients is 56.02 Gy (maximum dose limit: 55 Gy)

Modification of dose plan:
New dose plan: 60Gy + 10Gy
Dose received by 10% of patients is 54.26 Gy (feasible dose plan)

05 Causal Reasoning
Humans can understand the context

“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 $Y \rightarrow X$


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.

Remember: Reasoning = Sensemaking

- Deductive Reasoning = Hypothesis > Observations > Logical Conclusions
  - DANGER: Hypothesis must be correct OR it defines whether the truth of a conclusion can be determined for that rule, based on the truth of premises $A \land B$, $B \land C$, therefore $A \land C$
  - Inductive reasoning = makes broad generalizations from specific observations
  - DANGER: allows a conclusion to be false if the premises are true
  - Generalize hypotheses and use DR for answering specific questions
- Abductive reasoning = inference to get the best explanation from an incomplete set of preconditions.
  - Given a true conclusion and a rule, it attempts to select some possible premises that, if true also, may support the conclusion, though not uniquely.
- Example: "When it rains, the grass gets wet. The grass is wet. Therefore, it might have rained." This kind of reasoning can be used to develop a hypothesis, which in turn can be tested by additional reasoning or data.

Remember: hard inference problems

- High dimensionality ( curse of dim., many factors contribute)
- Complexity (real-world is non-linear, non-stationary, non-IID *)
- Need of large top-quality data sets
- Little prior data (no mechanistic models of the data)
  - * = Def: a sequence or collection of random variables is independent and identically distributed if each random variable has the same probability distribution as the others and all are mutually independent

06 Explainability

Why did it make this decision???
We need effective tools for Human-AI Interaction

**Why did the algorithm do that?**
*Can I trust these results?*
*How can I correct an error?*

A possible solution

- Explanation Interface
- Explainable Model

Input data

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

---

Example for an Explanation Interface

- open work

---

Example for an Explanation Interface - open work

---

What is understandable, interpretable, intelligible?

---

Explorable AI is a huge challenge for visualization

---

Methods of ex-AI

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

---

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

---

07 Methods of Explainable AI

---

Andrea Holzinger (U 18.11.15) from Explorable-AI-Causality, EETS course at Graz University of Technology
https://www.see.oeaw.ac.at/explainable-ai-causality/
Explainability := a property of a system ("the AI explanation")
Causability := a property of a person ("the Human explanation")

Our goal is to provide interfaces for effective mapping

- Causability := a property of a person (Human)
- Explainability := a property of a system (Computer)

Human sensemaking (Cognitive Science)
- Explainable AI (Computer Science)
**Heatmap Computation**


**Pixel-wise decomposition for bag-of-words features**


**Deep Taylor Decomposition**

**Definition 1.** A heatmapping $R(x)$ is conservative if the sum of assigned relevances in the pixel space corresponds to the total relevance detected by the model:

\[ \forall x : \mathbf{f}(x) = \sum P \mathbf{R}_j(x). \]

**Definition 2.** A heatmapping $R(x)$ is positive if all values forming the heatmap are greater or equal to zero, that is:

\[ \forall x : p : R(x) \geq 0. \]

**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 and 2.

**Sensitivity Analysis vs. Decomposition**

**Relevance propagation**

**Relevance Redistribution**

**Example 1**

**Example 2 Histopathology**

**LIME – Local Interpretable Model Agnostic Explanations**

Framework for vision: AND-OR Graphs

- Algorithm for this framework
- Top-down/bottom-up computation
- Generalization of small sample
- Use Monte Carlo simulation to synthesis more configurations
- Fill semantic gap

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

- Terminal (leaf) node: $T(v)$
- And-Or node: $F'(pg), F''(pg)$
- Set of links: $E(pg)$
- Switch variable at Or-node: $w(t)$
- Attributes of primitives: $a(t)$

\[ p(pg; \theta, R) = \frac{1}{Z(\theta)} \exp(\langle \xi \rangle) \]
\[ = \sum_{w(t)} \sum_{a(t)} \lambda_w (w(t)) + \sum_{a(t)} \lambda_a (a(t)) \]
\[ + \sum_{a(t)} \lambda_{a}(a(t)) \]

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

TCAV Testing with Concept Activation Vectors

- Testing with Concept Activation Vectors (TCAV)

Input images

Output images
1. **Stochastic Model on AND-OR graph**: Zhaoyin Jia (2009)
   - Terminal (leaf) node: \( T(pg) \)
   - And-Or node: \( \nu^m(pg), \nu^m(pg) \)
   - Set of links: \( E(pg) \)
   - Switch variable at And node: \( w(t) \)
   - Attributes of primitives: \( \alpha(t) \)

   \[
   p(pg; \theta, R, \lambda) = \frac{1}{Z(\theta)} \exp(-\langle \xi(pg) \rangle)
   \]

   \[
   \xi(pg) = \sum_{n=0}^{\infty} \frac{\lambda_n}{n!} \xi_n + \sum_{n=0}^{\infty} \frac{\lambda_n}{n!} \xi_n
   \]

   Spatial and appearance between primitives (parts or objects)

2. **Stochastic graph grammar/com. object representation**

3. **Future Work**

4. **Seemingly trivial questions ... ?**
   - What is a good explanation?
   - (obviously if the other did understand it)
   - Experiments needed!
   - What is explainable/understandable/eligible?
   - When it is enough (Sättigungsgrad – you don’t need more explanations – enough is enough)
   - But how much is it ...

5. **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!

6. **Combination of Deep learning with Ontologies**

   **Explainable AI with Deep Tensor and Knowledge Graph**

7. **IBM is doing it now: teaching meaningful explanations**

   **Teaching Meaningful Explanations**

---

Conclusion

- Computational approaches can find in $R^n$ what no human is able to see
- However, still there are many hard problems where a human expert in $R^2$ can understand the context and bring in experience, expertise, knowledge, intuition, ...
- Black box approaches can not explain WHY a decision has been made ...

The underlying architecture: Multi-Agent System

- 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

This is compatible to interactive machine learning

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

The fist wave of AI [1943-1975]: Handcrafted Knowledge

- Myth 1a: Superintelligence by 2100 is inevitable!
- 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: AI can never control us humans
  Fact: Intelligence is an enabler for control: We control tigers by being smarter ...

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

The third wave of AI (?): Adaptive Context Understanding

Thank you!