Andreas Holzinger  
VO 709.049 Medical Informatics  
25.11.2015 11:15-12:45  
Lecture 07  
Knowledge, Decision, Uncertainty, Bayesian Statistics, Probabilistic Modelling  
a.holzinger@tugraz.at  
Tutor: markus.plass@student.tugraz.at  
http://hci-kdd.org/biomedical-informatics-big-data

**Keywords of 7th Lecture**

- Bayes theorem  
- Case based reasoning  
- Differential diagnosis  
- Human decision making  
- Hypothetic-deductive method  
- Incomplete data  
- Model of human information processing  
- Modeling patient health  
- PDCA-Deming wheel  
- Receiver operating characteristics  
- Rough set theory  
- Selected attention  
- Signal detection theory  
- Triage

**Schedule**

1. Intro: Computer Science meets Life Sciences, challenges, future directions  
2. Back to the future: Fundamentals of Data, Information and Knowledge  
3. Structured Data: Coding, Classification (ICD, SNOMED, MeSH, UMLS)  
4. Biomedical Databases: Acquisition, Storage, Information Retrieval and Use  
5. Semi structured and weakly structured data (structural homologies)  
6. Multimedia Data Mining and Knowledge Discovery  
7. Knowledge, Decision, Cognition, Probability, Uncertainty, Bayes & Co  
8. Biomedical Decision Making: Reasoning and Decision Support  
9. Intelligent Information Visualization and Visual Analytics  
10. Biomedical Information Systems and Medical Knowledge Management  
11. Biomedical Data: Privacy, Safety and Security  

**Advance Organizer (1/2)**

- **Brake Force** = typical very general problem-solving technique that consists of systematically encompassing all possible candidate solutions and checking whether each candidate satisfies the problem's statement.  
- **Cognition** = mental processes of gaining knowledge, comprehending, including thinking, attention, remembering, language understanding, decision-making and problem solving.  
- **Cognitive load** = According to Sweller (1994) a measure of complexity and difficulty of a task, related to the executive control of the short-term memory, computing with factors including (human) processing power.  
- **Cognitive Science** = interdisciplinary study of human information processing, including perception, language, memory, reasoning, and emotion.  
- **Confounding variable** = an unobserved variable that Jarque normality and reliability of a study variable.  
- **Correlation coefficient** = represents the relationship between pairs of interval variables in a sample, from r = 1.000 (perfect correlation) to r = -1.000.  
- **Decision Making** = central cognitive process in every medical activity resulting in the selection of a final course of action out of alternatives, according to Shoemaker (2011) Del is still the key topic in medical research.  
- **Diagnostic** = classification of a patient's condition into separate and distinct categories that allow medical decisions about treatment and prognosis.  
- **Differential Diagnosis (Diagnosis) (DDx)** = a systematic method to identify the presence of an entity where multiple alternatives are possible, and the process of elimination, or interpretation of the probabilities of conditions to make levels.  
- **Evidence-based medicine (EBM)** = aiming for the best available evidence gained from the scientific methods to clinical decision making. It works to assess the strength of evidence of the risks and benefits of treatments (including lack of evidence) and diagnostic tests. Evidence quality can range from strong (level 1, systematic reviews) to weak (level 6, expert opinion).  
- **ICD-10** = World Health Organization's 10th revision of the International statistical classification of diseases and related health problems.  
- **IO9** = Institute of Medicine  
- **LTS** = Long Term Storage  
- **ME** = Medical Error  
- **PDCA** = Plan-Do-Check-Act  
- **QM** = Quality Management  
- **ROC** = Receiver Operating Characteristic  
- **RM2** = Receiver-operating characteristic  
- **RST** = Rough Set Theory  
- **STS** = Short Term Storage  
- **USTS** = Ultra Short Term Storage

**Advance Organizer (2/2)**

- **External Validity** = the extent to which the results of a study are generalizable or transferable;  
- **Hypothesis-Driven Medicine (HDM)** = formulating a hypothesis as the form that would counteract be falsified by a test on observable data, e.g. a test which shows results contrary to the prediction of the hypothesis is the basis for a test that could be but is contrary to the hypothesis (counterfactual theory) – then you need to compare the hypothetical outcome by treating how strong they are supported by their patients;  
- **Internal Validity** = the rigor with which a study was conducted (e.g., the design, the care taken to conduct measurements, and decisions concerning what was and was not measured);  
- **Learning Process** = the process of acquiring knowledge and skills, resulting in the change of knowledge or skills;  
- **Validity** = systematic and continuous improvement. Every improvement starts with a goal and a plan on how to achieve this goal, followed by action, measurement and comparison of the gained output.  
- **Perception** = sensory experience of the world, involving the recognition of environmental stimuli and action in a negative feedback cycle.  
- **Qualitative Research** = empirical research exploring relationships using textual, rather than quantitative data, e.g. case study, observation, ethnography, results are not considered generalizable for, sometimes qualitative research is qualitative or quantitative research.  
- **Quantitative Research** = empirical research exploring relationships using numeric data, e.g. surveys, case studies, experiments. Results should be generalizable, although it is not always possible;  
- **Reasoning** = cognitive (thought) process involved in making medical decisions (clinical reasoning, medical problem solving, diagnostic reasoning, before every action).  
- **Receiver operating characteristic (ROC)** = is signal detection theory is a graphical plot of the sensitivity or true positive rate vs. false positive rate (1-specificity, or 1 - true negative rate), for a binary classifier system as its discrimination threshold is varied;  
- **Symmetry reasoning** = logical deduction  
- **Triage** = process of judging the priority of patients' treatments based on the severity of their condition;
Learning Goals: At the end of this 7th lecture you will...

- are familiar with some principles and elements of human information processing;
- can discriminate between perception, cognition, thinking, reasoning & problem solving;
- have got insight into some basics of human decision making processes;
- got an overview of the Hypothetico-Deductive Method HDM versus PCDA Deming approach;
- have acquired some basics on modeling patient health, differential diagnosis, case-based reasoning and medical errors;

Slide 7-1 Key Challenges

- Time to make a decision = “5 Minutes” [1], [2]
- Limited perceptual, attentive and cognitive human resources [3], and Human error
- Noisy, missing, probabilistic, uncertain data


Medical Diagnosis - Decision Making

Slide 7-2: Decision Making is central in Biomedical Informatics

Source: Cisco (2008). Cisco HealthPresence Trial at Aberdeen Royal Infirmary in Scotland

Slide 7-3: Reasoning Foundations of Medical Diagnosis

Reasoning Foundations of Medical Diagnosis

- Symbolic logic, probability, and value theory and our understanding of how physicians reason.

Robert S. Sieffry and Lee R. Leaser

Slide 7-4 Decision Making is central in Medicine!


WS 2015 2
Human Information Processing

Slide 7-6 General Model of Human Information Processing

Slide 7-7 Example: Visual Information Processing

Slide 7-8 Schematic Information Processing Chain

Slide 7-9 Information processing of images/pictures
b) Processing of visual information (PRINTED WORDS)

- Multimedia Presentation
- Sensory Register
- Working Memory
- Long-Term Memory

<table>
<thead>
<tr>
<th>Words</th>
<th>Ears</th>
<th>Sounds</th>
<th>Verbal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pictures</td>
<td>Eyes</td>
<td>Images</td>
<td>Pictorial</td>
</tr>
</tbody>
</table>

Prior Knowledge


---

c) Processing of audio information (SPOKEN WORDS)

- Multimedia Presentation
- Sensory Register
- Working Memory
- Long-Term Memory

<table>
<thead>
<tr>
<th>Words</th>
<th>Ears</th>
<th>Sounds</th>
<th>Verbal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pictures</td>
<td>Eyes</td>
<td>Images</td>
<td>Pictorial</td>
</tr>
</tbody>
</table>

Prior Knowledge


---

Alternative Model: Baddeley - Central Executive


Neural Basis for the “Central Executive System”


Central Executive – Selected Attention


Selective Attention

Human Attention is central for problem solving


Human Decision Making


Start with the most simplest decision support system


Clinical DM: Hypothesis-Oriented Algorithm


Hypothetico-Deductive vs PDCA Deming wheel

Example 1/4: carcinoid heart disease (chd)

Hepatic venous congestion and carcinoid heart disease secondary to an ovarian carcinoid tumour in a 56-year-old woman with elevated liver enzyme levels and right upper quadrant pain.

Example 2/4: bone-marrow depression (bmd)

Example 3/4: partial liver resection (plr)

Example 4/4: radiofrequency ablation (rfa)

Slide 7-20 Modeling Patient health (1/2)

Slide 7-21 Modeling Patient Health (2/2)

Let \( U \subseteq X \) denote this risk factors and let \( V = X \setminus U \) denote the complement. The risk of immediate death \( p(\text{health}(t) = \text{death}|X) \) can be expressed by calculation of the following product:

\[
p(U) = \prod_{i \in U} p(U)
\]

Further, we obtain

\[
p(\text{health}(t) = \text{death}|X) = p(h|P) \prod_{i \in U} p(\text{health}(t) \neq \text{death}|U, \text{health}(t - 1))\]

for \( h = \text{death} \)
What for?

Slide 7-23 Information Acquisition and criteria - bias

- Information acquisition: in the CT data, e.g. healthy lungs have a characteristic shape, the presence of a tumor might distort that shape (= anomaly).
- Tumors have different image characteristics: brighter or darker, different texture, etc.
- With proper training a doctor learns what kinds of things to look for, so with more practice/training they will be able to acquire more (and more reliable) information.
- Running another test (e.g., MRI) can be used to acquire more (relevant) information.
- The effect of information is to increase the likelihood of getting either a hit or a correct rejection, while reducing the likelihood of an outcome in the two error boxes (slide 33).
- Criterion: Additionally to relying on technology/testing, the medical profession allows doctors to use their own judgment.
- Different doctors may feel that the different types of errors are not equal.
- For example: a doctor may feel that missing an opportunity for early diagnosis may mean the difference between life and death.
- A false alarm, on the other hand, may result only in a routine biopsy operation. They may choose to err toward “yes” (tumor present) decisions.
- Other doctors, however, may feel that unnecessary surgeries (even routine ones) are very bad (expensive, stress, etc.).
- They may choose to be more conservative and say “no” (no tumor) more often. They will miss more tumors, but they will be doing their part to reduce unnecessary surgeries. And they may feel that a tumor, if there really is one, will be picked up at the next check-up.


Slide 7-24: Decision Making Process vs. Data Mining process

Slide 7-25 Decision Making Process - Signal Detection

Remember: Two doctors, with equally good training, looking at the same CT scan data, will have the same information ... but they may gain different knowledge due to bias/criteria.


http://www.psych.stanford.edu/~lera/psych115s/notes/signal

Slide 7-26 Receiver Operating Characteristics (ROC curve)

http://www.psych.stanford.edu/~lera/psych115s/notes/signal/
Bayes’ Rule (1763)

**Probable Information P(x)**

- **Sum Rule**: \( P(x) = \sum_{x \in X} P(x, y) \)
- **Product Rule**: \( P(x, y) = P(y|x)P(x) \)

Bayes’ Rule is a corollary of Sum Rule and Product Rule:

\[
P(x|y) = \frac{P(y|x)P(x)}{\sum_{x \in X} P(x, y)P(x)}
\]

Bayes’ Rule in words:

\[
P(h|d) = \frac{P(d|h)P(h)}{P(d)}
\]

*P(h|d):* posterior (probability of hypothesis \( h \) after having seen the data \( d \))

*P(d|h):* likelihood (probability of the data \( d \) if the hypothesis \( h \) is true)

*P(d):* data evidence (marginal probability of the data)

*P(h):* prior belief (probability of hypothesis \( h \) before seeing any data)

The inverse probability allows to infer unknowns, learn from data and make predictions = machine learning!

Representation of uncertainty

- Many aspects of intelligence and learning depend on probabilistic representation of uncertainty:
  - Forecasting
  - Decision support
  - Learning from noisy, missing, uncertain data ...
  - Knowledge discovery
  - Probabilistic programming (e.g. Stochastic Python, Julia)
  - Universal inference algorithms
  - Global optimization

Clinical Example:

- \( D \): acute heart attack
- \( U_4 \): instable chest pain
- \( p(D) \): 37 of 1000 = 0.037 (heart attack)
- \( p(D) \): 963 of 1000 = 0.963 (no heart attack)
- 40% of patients report on instable chest pain
- \( p(U_4|D) \) = 0.4
- Unfortunately this symptoms also occur in 5 % of the healthy population
- \( p(U_4|\neg D) \) = 0.05
- We find the probability for a heart attack during this symptoms therefore by using Bayes’ Rule:

\[
p(D|U_4) = \frac{p(U_4|D)p(D)}{p(U_4|D)p(D) + p(U_4|\neg D)p(\neg D)} = 0.235
\]

Example Clinical Case: Serotonin Syndrome

What can we do if we have not only probabilistic, but also incomplete data ...

- is an extension of the Classical Set Theory, for use when representing incomplete knowledge.
- RS are sets with fuzzy boundaries – sets that cannot be precisely characterized using the available set of attributes, exactly like it is in medical decision making; based on 2 ideas:
  1) a given concept can be approximated by partition-based knowledge as upper and lower approximation – which corresponds to the focusing mechanism of differential medical diagnosis; upper approximation as selection of candidates and lower approximation as concluding a final diagnosis.
  2) a concept or observation can be represented as partitions in a given data set, where rough sets provides a rule induction method from given data. Thus, this model can be used to extract rule-based knowledge from medical databases.

Rough Set Theory

https://www.calvin.edu/~pribeiro/othrlnks/Fuzzy/fuzzyeng.htm

Diagnostic Procedure (Differential Diagnostic)

Example Symptom: Headache

- Focusing Mechanism (Selection of Candidates)
- Characterization (Negative Rules)
- Differential Diagnosis
- Discrimination (Positive Rules)
- Detection of Complications
- Comlications

Let \( U \) denote a non-empty, finite set called the universe and \( A \) denote a non-empty, finite set of attributes:

\[
\begin{align*}
\text{1) Myocardial infarction} & \quad \text{N.A. = not applicable} \\
\text{2) Nitroglycerin} & \quad \text{T = T-waves with peaking or inversion} \\
\end{align*}
\]

- **a**: \( U \rightarrow V_a \) for \( a \in A \)
- where \( V_a \) is called the domain of \( a \)
- Then, the decision table is defined as an information system:

\[
A = (U, A \cup \{d\}).
\]

The table shows an example of an information system with:

- \( U = \{1, 2, 3, 4, 5, 6\} \)
- \( A = \{\text{age, location, nature, prodrome, nausea, MI}\} \)
- \( d = \text{class} \)

For \( \text{location} \in A \), \( V_{\text{location}} \) is defined as \( \{\text{ocular, lateral, whole}\} \)

---

Is there another possibility?

1) Logic
2) Statistics/Probability
3) Heuristics

---

**Human Error**
Slide 7-44 Medical Errors: The IOM Study


Slide 7-45 Definitions of medical errors

- Medical error = any failure of a planned action;
- Serious ME = causes harm; includes preventable adverse events, intercepted serious errors, and non-intercepted serious errors. Does not include trivial errors with little or no potential for harm or non-preventable adverse events;
- Intercepted serious error = is caught before reaching patients;
- Non-intercepted serious error = reaches the patient but of good fortune or sufficient reserve to buffer the error, it did not cause harm;
- Adverse event = any injury (e.g. a rash caused by an antibiotic, deep vein thrombosis following omission to continue prophylactic subcutaneous heparin orders on transfer to the critical care unit, ventricular tachycardia due to placement of a central venous catheter tip in the right ventricle etc.);
- Non-preventable adverse event = Unavoidable injury due to inappropriate medical care.


Slide 7-46 Framework for understanding human error


Slide 7-47 Future Outlook

Sample Questions (1)

- What is the main and central topic in medical informatics?
- Please explain the information flow within the memory system according to Atkinson & Shiffrin!
- Explain the general model of human information processing following the model of Wickens?
- Explain the processing of visual (image, pictorial) information?
- What is so different in the alternative memory model according to Baddeley (1986)?
- Why is Attention of importance for medical informatics?
- Please explain the process of human decision making according to the model of Wickens (1984)?
- What is Triage?
- Please explain the hypothesis-oriented algorithm for Clinicians!
- What is the big difference between the Hypothetico-Deductive Method and the Plan-Do-Check-Act Deming Model?
- How can we model patient health – please provide an example!
Sample Questions (2)

- Please contrast the decision making process with the data mining process!
- Why is Signal Detection Theory important for us?
- Please provide an example for the application of Bayes’ Theorem!
- How does Differential Diagnosis work?
- How can we apply Rough Set Theory for differential diagnostics?
- What is Heuristic Decision Making?
- What is problematic when dealing with heuristic decision making from an informatics viewpoint?
- What is Case Based Reasoning (CBR)?
- How are medical errors defined?
- How does the framework for understanding human error work?

Appendix: NEJM Interactive Multimedia Cases

Appendix: Sample Why

- What are diagnostics?
- What is the Theorem?
- What can diagnostics provide from the problem?
- What is the decision framework an important decision making for medical human understanding?
Slide 7-35 Classification Accuracy and Coverage

Definition 1. Let \( R \) and \( D \) denote a formula \( \psi (\mathbf{D}, \mathbf{X}, \mathbf{V}) \) and a set of objects which belong to a decision \( d \). Classification accuracy and coverage are defined as:

\[
\text{Accuracy} (\mathcal{D}) = \frac{|R \cap \mathcal{D}|}{|\mathcal{D}|} \quad \text{and} \\
\text{Coverage} (\mathcal{D}) = \frac{|R \cap \mathcal{D}|}{|R|} 
\]

where \( |\cdot| \) denotes the cardinality of a set, \( \mathcal{D} \) denotes a training set, \( \mathcal{D} \) denotes a test set, and \( \mathcal{D} \) is the probability of \( \mathcal{D} \), respectively.

Slide 7-36 Probabilistic Rules – modus ponens

By the use of accuracy and coverage, a probabilistic rule is defined as:

\[
R \rightarrow D \quad \text{st.} \quad \nabla_f (D) = \frac{\sum_{R \in \mathcal{R}} \frac{P(R) \cdot f(D|R)}{P(D)}}{\sum_{R \in \mathcal{R}} \frac{P(R)}{P(D)}} > \delta_k
\]

where \( \delta_k \) is the minimum acceptable probability of \( \mathcal{D} \).

Slide 7-37 Positive Rules

A positive rule is defined as a rule supported by only positive examples. The classification accuracy of a positive rule is equal to 1.0. It is not the case that the set supporting a rule is a subset of the set of positive examples. The coverage of a positive rule is defined as:

\[
\text{Coverage} (\mathcal{D}) = \frac{|R \cap \mathcal{D}|}{|\mathcal{D}|} 
\]

Figure 1 shows the Venn diagram of a positive rule. As shown in this figure, the meaning of \( R \) is a subset of that of \( D \). This diagram is equivalent to the classic proposition \( R \rightarrow D \). In the above example, one positive rule of “m.c.h.” (mindset contraction hierarchy) is:

\[
D \rightarrow R \quad \text{st.} \quad \alpha = 3/3 = 1.0
\]

Slide 7-38 Exclusive Rules

Before defining a negative rule, let us first introduce an exclusive rule. An exclusive rule is defined as a rule supported by all the positive examples, the coverage of which is equal to 1.0. That is, an exclusive rule expresses the minimum condition of a decision. It is not the case that the set supporting an exclusive rule is a subset of the set of positive examples. The coverage of an exclusive rule is defined as:

\[
\text{Coverage} (\mathcal{D}) = \frac{|R \cap \mathcal{D}|}{|\mathcal{D}|} 
\]

Figure 1 shows the Venn diagram of an exclusive rule. As shown in this figure, the meaning of \( R \) is a subset of that of \( D \). This diagram is equivalent to the classic proposition \( R \rightarrow D \). In the above example, the exclusive rule of “m.c.h.” is:

\[
D \rightarrow R \quad \text{st.} \quad \alpha = 1/1 = 1.0
\]

Slide 7-39 Negative Rule

A negative rule is defined as the complement of an exclusive rule. The meaning of \( R \) is an object satisfying any attribute value pairs which do not satisfy the condition of a positive rule. Therefore, we can construct a decision tree by constructing.

\[
\text{Negative} \quad \text{Rule} (R) = \neg R
\]

where \( D \) is a set of objects which belong to a class.

Slide 7-40 Example: Algorithms for Rule Induction

algorithm Exclusive and Negative Rules

\[
\begin{align*}
\text{procedure} & \quad \text{Exclusive and Negative Rules} \\
\text{input} & \quad \text{a list of elementary attribute-value pairs} \\
\text{begin} & \quad \text{begin} \\
\text{end} & \quad \text{end}
\end{align*}
\]

algorithm Partial Positive Rules

\[
\begin{align*}
\text{procedure} & \quad \text{Partial Positive Rules} \\
\text{input} & \quad \text{a list of elementary attribute-value pairs} \\
\text{begin} & \quad \text{begin} \\
\text{end} & \quad \text{end}
\end{align*}
\]

algorithm Construct Negative Rules

\[
\text{procedure} \quad \text{Construct Negative Rules} \\
\text{begin} \\
\text{end}
\]

Example: Algorithms for Rule Induction

procedure Construct Negative Rules:

\[
\begin{align*}
\text{procedure} & \quad \text{Construct Negative Rules} \\
\text{input} & \quad \text{a list of elementary attribute-value pairs} \\
\text{begin} & \quad \text{begin} \\
\text{end} & \quad \text{end}
\end{align*}
\]

procedure Construct Positive Rules:

\[
\begin{align*}
\text{procedure} & \quad \text{Construct Positive Rules} \\
\text{input} & \quad \text{a list of elementary attribute-value pairs} \\
\text{begin} & \quad \text{begin} \\
\text{end} & \quad \text{end}
\end{align*}
\]

example: Construct Positive Rules:

\[
\begin{align*}
\text{procedure} & \quad \text{Construct Positive Rules} \\
\text{input} & \quad \text{a list of elementary attribute-value pairs} \\
\text{begin} & \quad \text{begin} \\
\text{end} & \quad \text{end}
\end{align*}
\]
The Naïve Bayes Classifier

- What can we do if our data $d$ has several attributes?
- Naïve Bayes assumption: Attributes that describe data instances are conditionally independent given the classification hypothesis
  \[ P(d|h) = P(a_1, a_2 | h) = \prod P(a_i | h) \]
  - It is a simplifying assumption, obviously it may be violated in reality
  - In spite of that, it works well in practice
- The Bayesian classifier that uses the Naïve Bayes assumption and computes the MAP hypothesis is called Naïve Bayes classifier
- One of the most practical learning methods
- Successful applications:
  - Medical Diagnosis