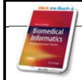


Andreas Holzinger
VO 709.049 Medical Informatics
25.11.2015 11:15-12:45

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

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
- 12. Methodology for Info Systems: System Design, Usability & Evaluation

Keywords of 7th Lecture

- Bayes theorem
- Case based reasoning
- Differential diagnosis
- Human decision making
- Hypothetico-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

Advance Organizer (1/2)

- **Brute Force** = a trivial very general problem-solving technique that consists of systematically enumerating all possible candidates for the solution and checking whether each candidate satisfies the problem's statement;
- **Cognition** = mental processes of gaining knowledge, comprehension, including thinking, attention, remembering, language understanding, decision making and problem-solving;
- **Cognitive load** = According to Sweller (1996) a measure of complexity and difficulty of a task, related to the executive control of the short-term memory, correlating with factors including (human) performance; based on the chunk-theory of Miller (1956);
- **Cognitive Science** = interdisciplinary study of human information processing, including perception, language, memory, reasoning, and emotion;
- **Confounding Variable** = an unforeseen, unwanted variable that jeopardizes reliability and validity of a study outcome.
- **Correlation coefficient** = measures the relationship between pairs of interval variables in a sample, from $r = -1.00$ to 0 (no correlation) to $r = +1.00$
- **Decision Making** = a central cognitive process in every medical activity, resulting in the selection of a final choice of action out of alternatives; according to Shortliffe (2011) DM is still the key topic in medical informatics;
- **Diagnosis** = classification of a patient's condition into separate and distinct categories that allow medical decisions about treatment and prognostic;
- **Differential 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 negligible levels;
- **Evidence-based medicine (EBM)** = aiming at the best available evidence gained from the scientific method to clinical decision making. It seeks to assess the strength of evidence of the risks and benefits of treatments (including lack of treatment) and diagnostic tests. Evidence quality can range from meta-analyses and systematic reviews of double-blind, placebo-controlled clinical trials at the top end, down to conventional wisdom at the bottom;

Advance Organizer (2/2)

- **External Validity** = the extent to which the results of a study are generalizable or transferable;
- **Hypothetico-Deductive Model (HDM)** = formulating a hypothesis in a form that could conceivably be falsified by a test on observable data, e.g. a test which shows results contrary to the prediction of the hypothesis is the falsification, a test that could but is not contrary to the hypothesis corroborates the theory – then you need to compare the explanatory value of competing hypotheses by testing how strong they are supported by their predictions;
- **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);
- **PDCA** = Plan-Do-Check-Act, The so called PDCA-cycle or Deming-wheel can be used to coordinate a systematic and continuous improvement. Every improvement starts with a goal and with a plan on how to achieve that goal, followed by action, measurement and comparison of the gained output.
- **Perception** = sensory experience of the world, involving the recognition of environmental stimuli and actions in response to these stimuli;
- **Qualitative Research** = empirical research exploring relationships using textual, rather than quantitative data, e.g. case study, observation, ethnography; Results are not considered generalizable, but sometimes at least transferable.
- **Quantitative Research** = empirical research exploring relationships using numeric data, e.g. surveys, quasi-experiments, experiments. Results should be generalized, although it is not always possible.
- **Reasoning** = cognitive (thought) processes involved in making medical decisions (clinical reasoning, medical problem solving, diagnostic reasoning, behind every action;
- **Receiver-operating characteristic (ROC)** = in signal detection theory this 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;
- **Symbolic reasoning** = logical deduction
- **Triage** = process of judging the priority of patients' treatments based on the severity of their condition;

Glossary

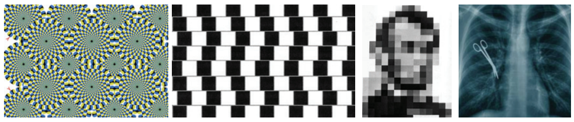
- CES = Central Executive System
- DDx = Differential Diagnosis
- DM = Decision Making
- DSS = Decision Support System
- EBM = Evidence-based medicine
- fMRI = functional Magnetic Resonance Image
- HDM = Hypothetico-Deductive Model
- IOM = Institute of Medicine
- LTS = Long Term Storage
- ME = Medical Error
- PDCA = Plan-Do-Check-Act
- QM = Quality Management
- ROC = Receiver Operating Characteristic
- ROC = Receiver-operating characteristic
- RST = Rough Set Theory
- STS = Short Term Storage
- USTS = Ultra Short Term Storage

Learning Goals: At the end of this 7th lecture you …

- … 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



[1] Gigerenzer, G. 2008. *Gut Feelings: Short Cuts to Better Decision Making* London, Penguin.
[2] Gigerenzer, G. & Gaissmaier, W. 2011. Heuristic Decision Making. In: Fiske, S. T., Schacter, D. L. & Taylor, S. E. (eds.) Annual Review of Psychology, Vol 62. pp. 451-482.
[3] Bialek, W. 1987. Physical Limits to Sensation and Perception. Annual Review of Biophysics and Biophysical Chemistry, 16, 455-478.

Medical Diagnosis - Decision Making

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

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

Slide 7-3: Reasoning Foundations of Medical Diagnosis

3 July 1959, Volume 130, Number 3366 SCIENCE

Reasoning Foundations of Medical Diagnosis

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

Robert S. Ledley and Lee B. Lusted

The purpose of this article is to analyze the complicated reasoning processes inherent in medical diagnosis. The importance of this problem has received recent emphasis by the increasing interest in the use of electronic computers as an aid to medical diagnostic processes

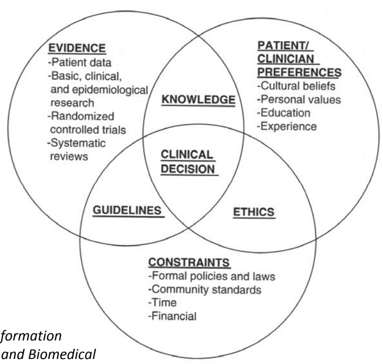
fitted into a definite disease category, or that it may be one of several possible diseases, or else that its exact nature cannot be determined." This, obviously, is a greatly simplified explanation of the process of diagnosis, for the physician might also comment that after seeing a

ance are the ones who do remember and consider the most possibilities."

Computers are especially suited to help the physician collect and process clinical information and remind him of diagnoses which he may have overlooked. In many cases computers may be as simple as a set of hand-sorted cards, whereas in other cases the use of a large-scale digital electronic computer may be indicated. There are other ways in which computers may serve the physician, and some of these are suggested in this paper. For example, medical students might find the computer an important aid in learning the methods of differential diagnosis. But to use the computer thus we must understand how the physician makes a medical diagnosis. This, then, brings us to the subject of our investigation: the reasoning foundations of medical diagnosis and treatment.

Medical diagnosis involves processes that can be systematically analyzed, as well as those characterized as "intangible." For instance, the reasoning foundations of medical diagnostic procedures

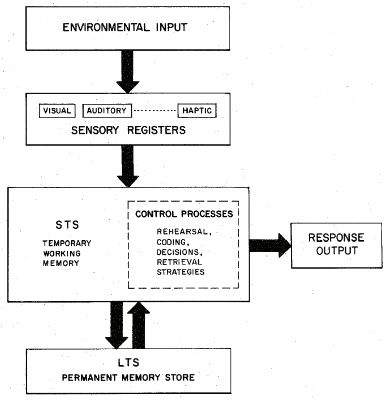
Slide 7-4 Decision Making is central in Medicine!



Hersh, W. (2010) *Information Retrieval: A Health and Biomedical Perspective*. New York, Springer.

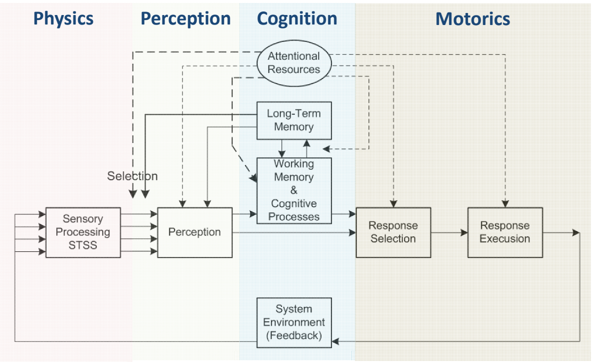
Human Information Processing

Slide 7-5 Human Information Processing Model (A&S)



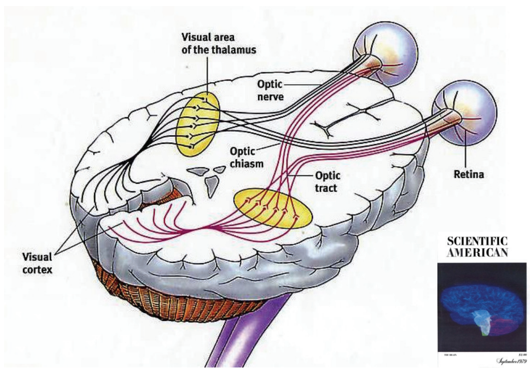
Atkinson, R. C. & Shiffrin, R. M. (1971) *The control processes of short-term memory (Technical Report 173, April 19, 1971). Stanford, Institute for Mathematical Studies in the Social Sciences, Stanford University.*

Slide 7-6 General Model of Human Information Processing



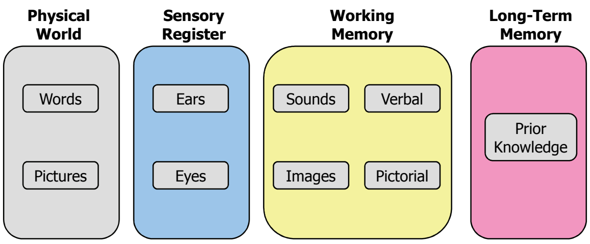
Wickens, C., Lee, J., Liu, Y. & Gordon-Becker, S. (2004) *Introduction to Human Factors Engineering: Second Edition. Upper Saddle River (NJ), Prentice-Hall.*

Slide 7-7 Example: Visual Information Processing



Source: Department of Neuroscience, The Mount Sinai School of Medicine (2004)

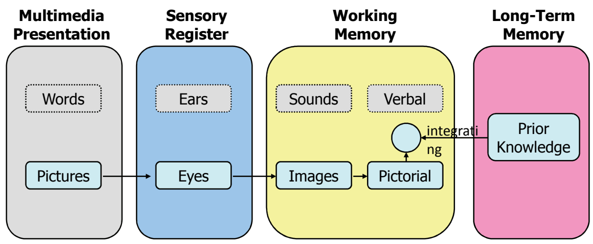
Slide 7-8 Schematic Information Processing Chain



cf. with Paivio (1973), Mayer & Moreno (1998), Holzinger (2000), Schnotz & Bannert (2002)

Slide 7-9 Information processing of images/pictures

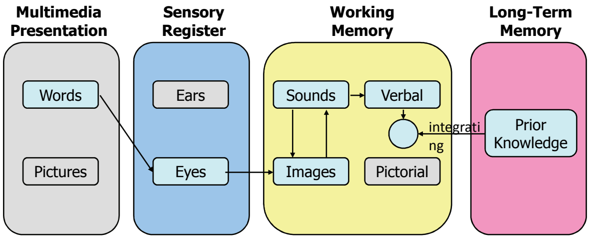
a) Processing of visual information (PICTURES)



cf. with Paivio (1973), Mayer & Moreno (1998), Holzinger (2000), Schnotz & Bannert (2002)

Slide 7-10 Information processing of words/pictures

b) Processing of visual information (PRINTED WORDS)

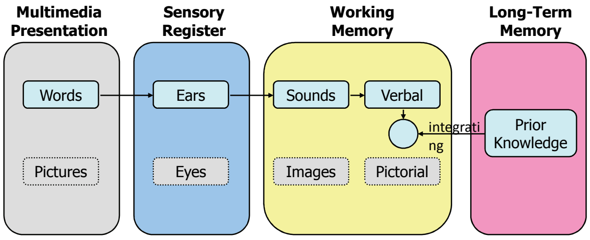


cf. with Paivio (1973), Mayer & Moreno (1998), Holzinger (2000), Schnotz & Bannert (2002)

A. Holzinger 709.049 19/79 Med Informatics L07

Slide 7-11 Information processing of words/sounds

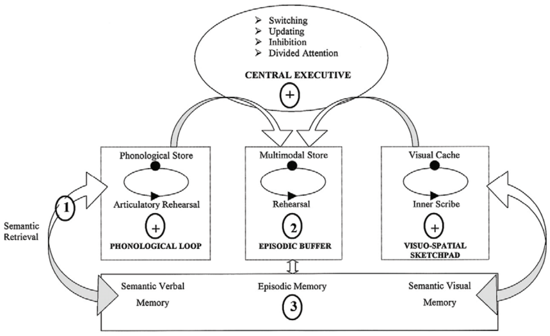
c) Processing of audio information (SPOKEN WORDS)



cf. with Paivio (1973), Mayer & Moreno (1998), Holzinger (2000), Schnotz & Bannert (2002)

A. Holzinger 709.049 20/79 Med Informatics L07

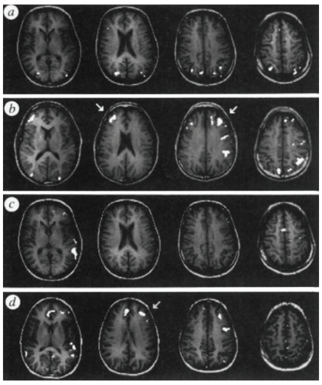
Slide 7-12 Alternative Model: Baddeley - Central Executive



Quinette, P., Guillery, B., Desgranges, B., de la Sayette, V., Viader, F. & Eustache, F. (2003) Working memory and executive functions in transient global amnesia. *Brain, 126, 9, 1917-1934.*

A. Holzinger 709.049 21/79 Med Informatics L07

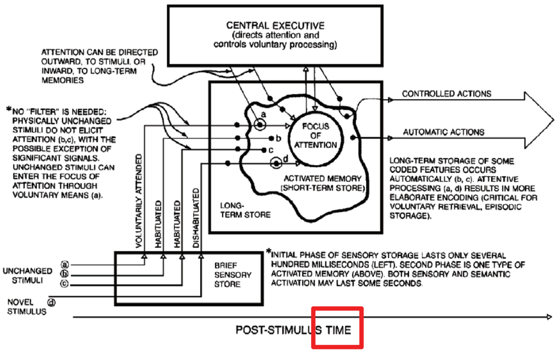
Slide 7-13 Neural Basis for the "Central Executive System"



D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S. & Grossman, M. (1995) The neural basis of the central executive system of working memory. *Nature, 378, 6554, 279-281.*

A. Holzinger 709.049 22/79 Med Informatics L07

Slide 7-14 Central Executive – Selected Attention



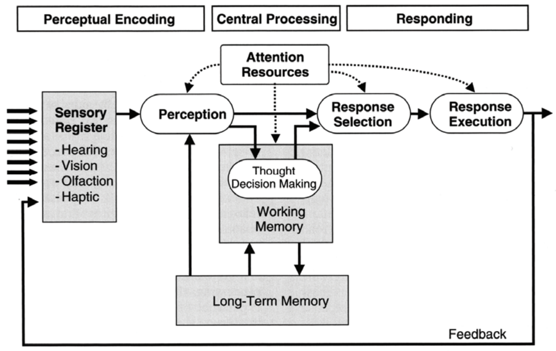
Cowan, N. (1988) Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104, 2, 163.*

A. Holzinger 709.049 23/79 Med Informatics L07

Simons, D. J. & Chabris, C. F. 1999. Gorillas in our midst: sustained inattentional blindness for dynamic events. Perception, 28, (9), 1059-1074.

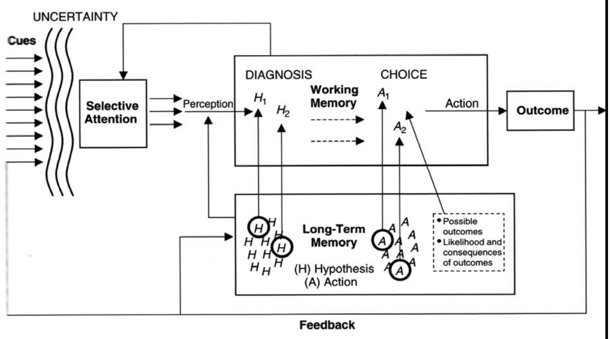
A. Holzinger 709.049 24/79 Med Informatics L07

Slide 7-15 Human Attention is central for problem solving



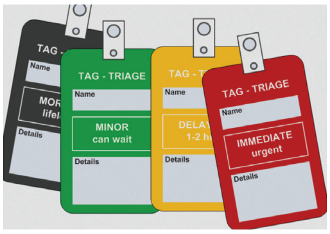
Wickens, C. D. (1984) *Engineering psychology and human performance. Columbus (OH), Charles Merrill.*

Slide 7-16 Human Decision Making



Wickens, C. D. (1984) *Engineering psychology and human performance. Columbus (OH), Charles Merrill.*

Start with the most simplest decision support system

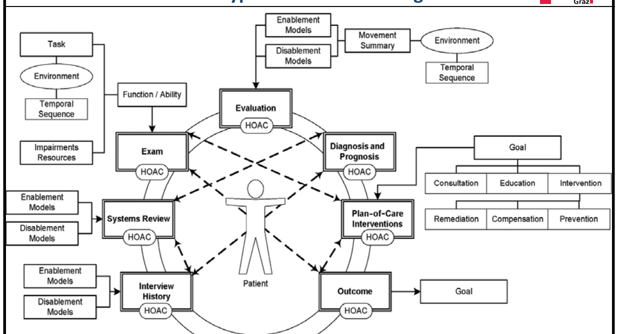


Iserson, K. V. & Moskop, J. C. 2007. Triage in Medicine, Part I: Concept, History, and Types. Annals of Emergency Medicine, 49, (3), 275-281.

Slide 7-17 Example: Triage Tags - International Triage Tags

Source: http://store.gomed-tech.com

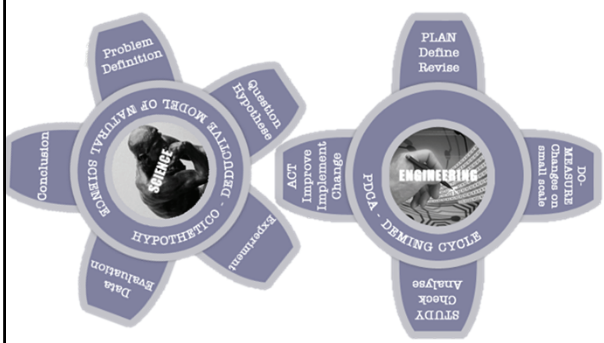
Slide 7-18 Clinical DM: Hypothesis-Oriented Algorithm



HOAC =Hypothesis-Oriented Algorithm for Clinicians

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.*

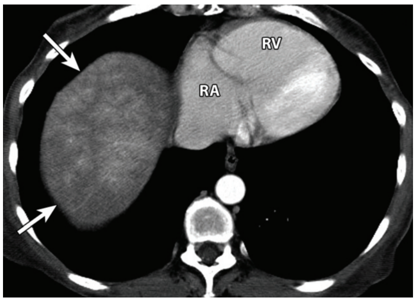
Slide 7-19 Hypothetico-Deductive vs PDCA Deming wheel



Holzinger, A. (2010) *Process Guide for Students for Interdisciplinary Work in Computer Science/Informatics. Second Edition. Norderstedt, BoD.* Online: http://www.hci4all.at

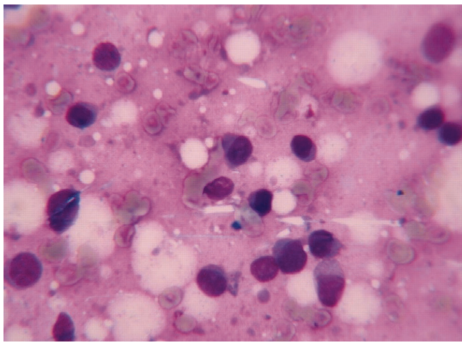
Example 1/4: carcinoid heart disease (chd)

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



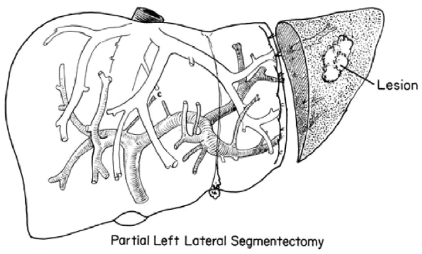
Shanbhogue, A. K. P., Shanbhogue, D. K. P., Prasad, S. R., Surabhi, V. R., Fasih, N. & Menias, C. O. (2010) Clinical Syndromes Associated with Ovarian Neoplasms: A Review. *Radiographics, 30, 4, 903-919.*

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



Prasad, M., Maitra, A., Sethiya, N., Bharadwaj, V. K., Chowdhury, V., Valecha, J. & Biswas, R. (2009) Acute renal failure followed by low back ache. *BMJ Case Reports, 2009.*

Example 3/4: partial liver resection (plr)

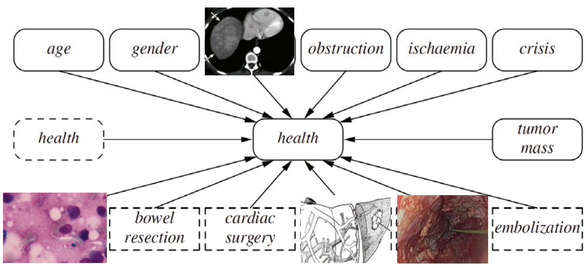


Zollinger, R. M. & Ellison, C. (2010) *Zollinger's Atlas of Surgical Operations (9th Edition). New York, McGraw Hill.*

Example 4/4: radiofrequency ablation (rfa)

van Vilsteren, F. G. I. et al. (2011) Stepwise radical endoscopic resection versus radiofrequency ablation for Barrett's oesophagus with high-grade dysplasia or early cancer: a multicentre randomised trial. *GUT.*

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



chd = carcinoid heart disease; bmd = bone-marrow depression; plr = partial liver resection; rfa = radiofrequency ablation; dashed … past states; square objects … treatments

van Gerven, M. A. J., Taal, B. G. & Lucas, P. J. F. (2008) Dynamic Bayesian networks as prognostic models for clinical patient management. *Journal of Biomedical Informatics, 41, 4, 515-529.*

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

Let $\mathbf{U} \subseteq \mathbf{X}$ denote this risk factors and
Let $\mathbf{V} = \mathbf{X} \setminus \mathbf{U}$ denote the complement.

The risk of immediate death $p(health(t) = death|X)$ can be expressed by calculation of the following product:

$$\prod_{U \in \mathbf{U}} p(U)$$

Further, we obtain

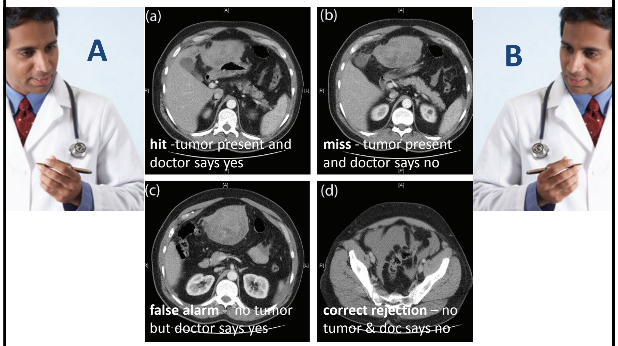
$$p(health(t) = h|\mathbf{X}) = p(h|\mathbf{V}) \prod_{U \in \mathbf{U}} p(health(t) \neq death|U, health(t-1))$$

for $h \neq death$

van Gerven, M. A. J., Taal, B. G. & Lucas, P. J. F. (2008) Dynamic Bayesian networks as prognostic models for clinical patient management. *Journal of Biomedical Informatics, 41, 4, 515-529.*

What for?

Slide 7-22 Signal Detection Theory on the MDM process



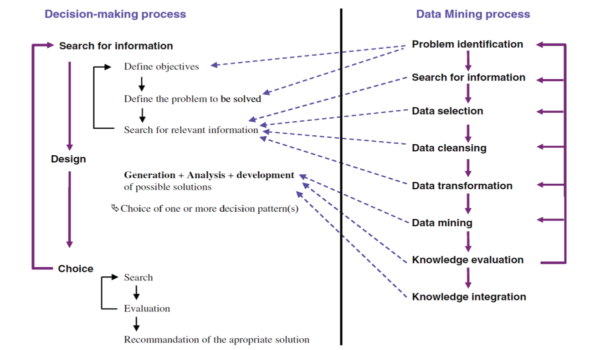
Two doctors, with equally good training, looking at the same CT scan, will have the same information … but they may have a different bias/criteria!

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 chose 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 chose to be more conservative and say ``no'' (no turmor) 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.

Mohamed, A. et al. (2010) Traumatic rupture of a gastrointestinal stromal tumour with intraperitoneal bleeding and haematoma formation. *BMJ Case Reports, 2010.*

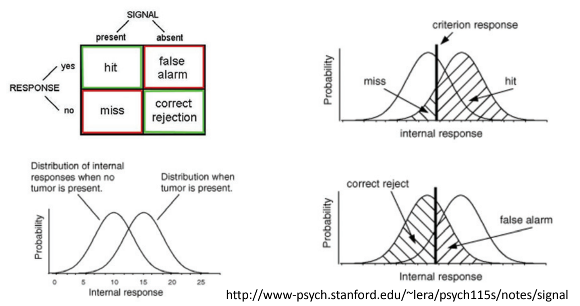
Slide 7-24: Decision Making Process vs. Data Mining process



Ayed, B. M., Ltifi, H., Kolski, C. & Alimi, A. (2010) A user-centered approach for the design & implementation of KDD-based DSS: A case study in the healthcare domain. *Decision Support Systems, 50, 64-78.*

Slide 7-25 Decision Making Process - Signal Detection

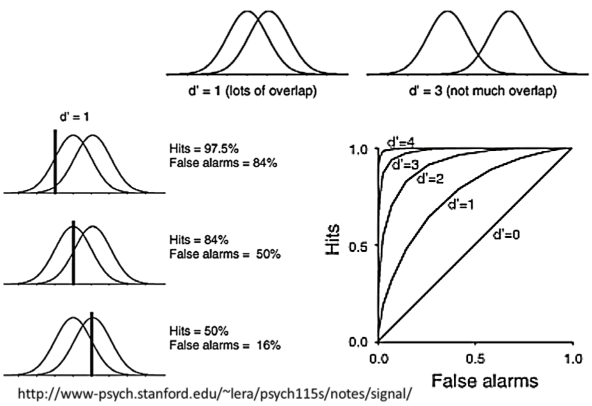
Remember: Two doctors, with equally good training, looking at the same CT scan data, will have the same information … but they may gain different knowledge due to *bias/criteria*.



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

For an example see: Braga & Oliveira (2003) Diagnostic analysis based on ROC curves: theory and applications in medicine. *Int. Journal of Health Care Quality Assurance, 16, 4, 191-198.*

Slide 7-26 Receiver Operating Characteristics (ROC curve)



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

Slide 7-27a Bayes' Rule (1763)

Probable Information P(x)

Sum Rule $\sum$

$$P(x) = \sum_{x \in X} P(x, y)$$

Thomas Bayes
1701 - 1761

Product Rule $\prod$

$$P(x, y) = P(y|x)P(x)$$

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

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

Slide 7-27b Bayes Law of Total Probability = data modelling

Bayes' Rule in words
d … data; h … hypothesis

$$P(h|d) = \frac{P(d|h)P(h)}{P(d)}$$

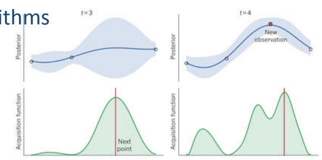
$P(h)$: prior belief (probability of hypothesis *h* before seeing any data)
$P(d\,|\,h)$: likelihood (probability of the data if the hypothesis *h* is true)
$P(d) = \sum_h P(d\,|\,h)P(h)$: data evidence (marginal probability of the data)
$P(h\,|\,d)$: posterior (probability of hypothesis *h* after having seen the data *d*)

$$posterior = \frac{likelyhood * prior}{evidence}$$

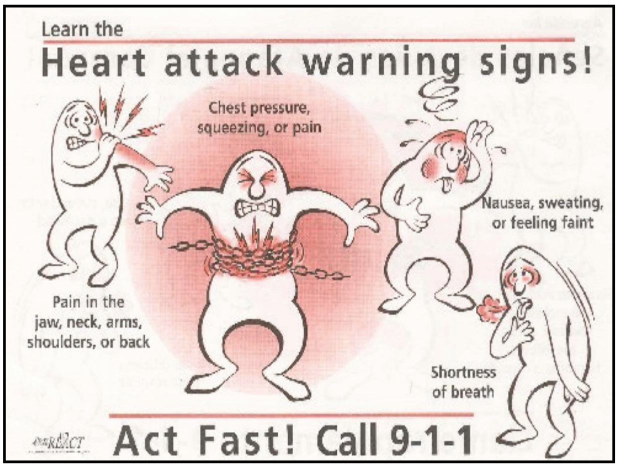
The inverse probability allows to infer unknowns, learn from data and make predictions = machine learning!

Slide 7-27c 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



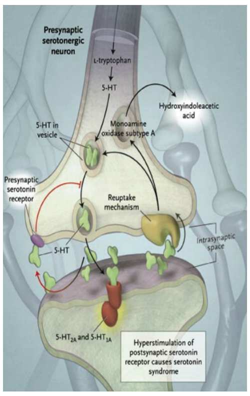
Ghahramani, Z. 2015. Probabilistic machine learning and artificial intelligence. Nature, 521, (7553), 452-459.

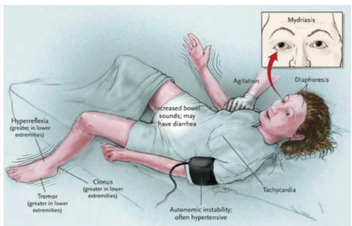


Slide 7-27d Bayes' in clinical practice

- Clinical Example:
- D … acute heart attack
- U_+… instable chest pain
- $p(D)$ … 37 of 1000 = 0,037 (heart attack)
- $p(\bar{D})$ … 963 of 1000 = 0,963 (no heart attack)
- 40% of patients report on instable chest pain
- $p(U_+|D)$ = 0,4
- Unfortunately this symptoms also occur in 5 % of the healthy population
- $p(U_+|\bar{D})$ = 0,05
- We find the probability for a heart attack during this symptoms therefore by using Bayes' Rule:
- $p(D|U_+) = \frac{p(U_+|D)*p(D)}{p(U_+|D)*p(D)+p(U_+|\bar{D})*p(\bar{D})}$ = 0,235

Slide 7-28: Example Clinical Case: Serotonin Syndrome

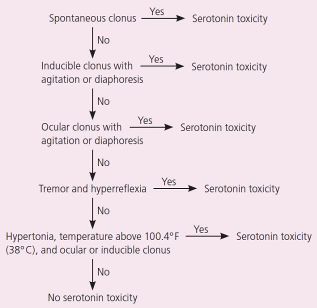




Boyer, E. W. & Shannon, M. (2005) The Serotonin Syndrome. *New England Journal of Medicine, 352, 11, 1112-1120.*

Slide 7-29 Differential Diagnosis on Serotonin Syndrome 1/2

Hunter's Decision Rules for Diagnosis of Serotonin Toxicity



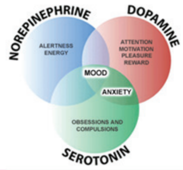
Signs & Symptoms of Serotonin Syndrome

Agitation (restlessness)*	Multi-organ failure†
Diaphoresis*	Myoclonus*
Diarrhea*	Ocular clonus
Disseminated intravascular coagulation†	Rhabdomyolysis†
Fever above 100.4°F (38°C)	Shivering*
Hyperreflexia*	Tonic-clonic seizures†
Incoordination (ataxia)*	Tremor*
Mental status changes	
Confusion*	
Hypomania*	

*—Sternbach's diagnostic criteria require three of 10 signs and symptoms.
†—Extremely severe cases.

Ables, A. Z. & Nagubilli, R. (2010) Prevention, recognition, and management of serotonin syndrome. *American family physician, 81, 9, 1139.*

Slide 7-30 Differential Diagnosis on Serotonin Syndrome 2/2

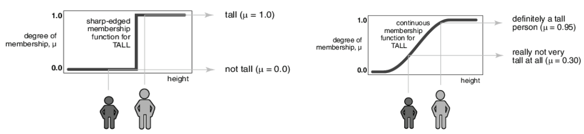


Clinical condition	*History*	*Vital signs*	*Clinical features*
Anticholinergic syndrome	Use of tricyclic antidepressants or other anticholinergic drugs	Tachycardia, tachypnea, hyperthermia (usually 102.2°F [39°C] or below)	Dry mouth, blurred vision, mydriasis, flushed skin, agitation/delirium, decreased bowel sounds
Malignant hyperthermia	Administration of halogenated inhalational anesthetics or depolarizing muscle relaxants	Hypertension, tachycardia, tachypnea, hyperthermia (up to 114.8°F [46°C])	Diaphoresis, mottled skin, agitation, decreased bowel sounds, muscular rigidity, hyporeflexia
Neuroleptic malignant syndrome	Ingestion of antipsychotic medications	Hypertension, tachycardia, tachypnea, hyperthermia (above 105.8°F [41°C])	Sialorrhea, diaphoresis, pallor, stupor, mutism, coma, normal or decreased bowel sounds, lead-pipe rigidity, bradyreflexia

Ables, A. Z. & Nagubilli, R. (2010) Prevention, recognition, and management of serotonin syndrome. *American family physician, 81, **9, 1139.***

What can we do if we have not only probabilistic, but also I c pl e data …

Rough Set Theory



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

Slide 7-31 Rough Set Theory for dealing with 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.

Slide 7-32 Diagnostic Procedure (Differential Diagnostic)

Example Symptom: Headache

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

Tsumoto, S. (2006) Pawlak Rough Set Model, Medical Reasoning and Rule Mining. In: Greco, S., Hata, Y., Hirano, S., Inuiguchi, M., Miyamoto, S., Nguyen, H. & Slowinski, R. (Eds.) *Rough Sets and Current Trends in Computing. Berlin, Heidelberg, Springer, 53-70.*

Slide 7-33 Rough Set Theory Example Symptom: Headache 1

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

- a : U → *Va* for a ∈ A
- where *Va* is called the domain of a
- Then, the decision table is defined as an information system:
- A = (U,A ∪ {d}).
- The table shows an example of an information system with
- U = {1, 2, 3, 4, 5, 6} and
- A = {age, location, nature, prodrome, nausea,M1} and
- d = class.
- For *location* ∈ A, $V_{location}$ is defined as {occular, lateral,whole}

No.	age	location	nature	prodrome	nausea	M1	class
1	50-59	occular	persistent	no	no	yes	m.c.h.
2	40-49	whole	persistent	no	no	yes	m.c.h.
3	40-49	lateral	throbbing	no	yes	no	migra
4	40-49	whole	throbbing	yes	yes	no	migra
5	40-49	whole	radiating	no	no	yes	m.c.h.
6	50-59	whole	persistent	no	yes	yes	psycho

Definitions. M1: tenderness of M1, m.c.h.: muscle contraction headache, migra: migraine, psycho: psychological pain.

Tsumoto, S. (2006) Pawlak Rough Set Model, Medical Reasoning and Rule Mining. In: Greco, S., Hata, Y., Hirano, S., Inuiguchi, M., Miyamoto, S., Nguyen, H. & Slowinski, R. (Eds.) *Rough Sets and Current Trends in Computing. Berlin, Heidelberg, Springer, 53-70.*

Is there another possibility?

1) Logic
2) Statistics/Probability
3) Heuristics

Slide 7-41: Science Vol. 185, pp.1124-1131, Sept. 1974

Judgment under Uncertainty: Heuristics and Biases

Biases in judgments reveal some heuristics of thinking under uncertainty.

Amos Tversky and Daniel Kahneman

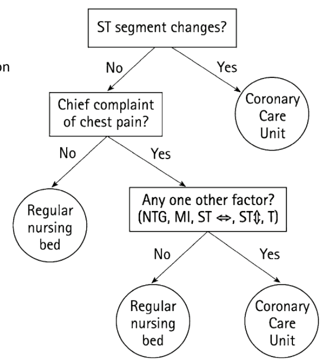
Many decisions are based on beliefs concerning the likelihood of uncertain events such as the outcome of an election, the guilt of a defendant, or the future value of the dollar. These beliefs are usually expressed in statements such as "I think that . . . ," "chances are . . . ," "it is unlikely that . . . ," and so forth. Occasionally, beliefs concerning uncertain events are expressed in numerical form as odds or subjective probabilities. What determines such beliefs? How do people assess the probmated when visibility is good because the objects are seen sharply. Thus, the reliance on clarity as an indication of distance leads to common biases. Such biases are also found in the intuitive judgment of probability. This article describes three heuristics that are employed to assess probabilities and to predict values. Biases to which these heuristics lead are enumerated, and the applied and theoretical implications of these observations are discussed.

occupation from a list of possibilities (for example, farmer, salesman, airline pilot, librarian, or physician)? How do people order these occupations from most to least likely? In the representativeness heuristic, the probability that Steve is a librarian, for example, is assessed by the degree to which he is representative of, or similar to, the stereotype of a librarian. Indeed, research with problems of this type has shown that people order the occupations by probability and by similarity in exactly the same way (*1*). This approach to the judgment of probability leads to serious errors, because similarity, or representativeness, is not influenced by several factors that should affect judgments of probability.

Insensitivity to prior probability of outcomes. One of the factors that have no effect on representativeness but should have a major effect on probability is the prior probability, or base-rate frequency, of the outcomes. In the case of Steve, for example, the fact that there are many more farmers than librarians in the population should enter into any reasonable estimate of the probability that Steve is a librarian rather than a farmer. Considerations of base-rate frequency, however, do not

Slide 7-42 Heuristic Decision Making

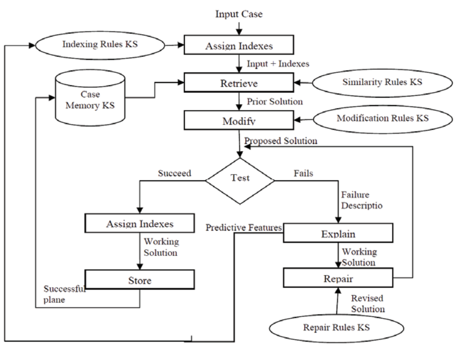
MI = myocardial infarction
N.A. = not applicable
NTG = nitroglycerin
T= T-waves with peaking or inversion



Gigerenzer, G. & Gaissmaier, W. (2011) Heuristic Decision Making. In: Fiske, S. T., Schacter, D. L. & Taylor, S. E. (Eds.) *Annual Review of Psychology, Vol 62. Palo Alto, Annual Reviews, 451-482.*

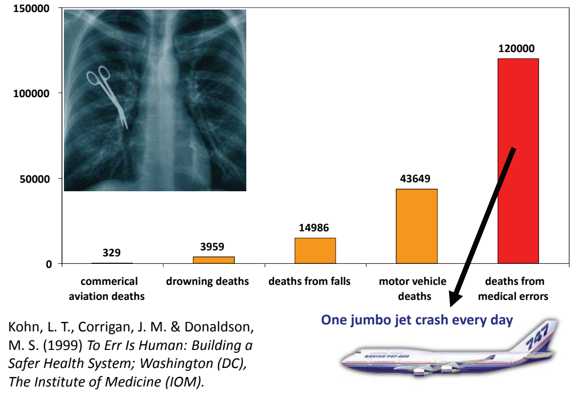
Slide 7-43 Case Based Reasoning (CBR)

Boxes = processes; ovals = knowledge structures (KS)



Salem, A. B. M. (2007) Case based reasoning technology for medical diagnosis. *Proc. World Academy of Science, Engineering and Technology, 25, 9-13.*

Slide 7-44 Medical Errors: The IOM Study



Kohn, L. T., Corrigan, J. M. & Donaldson, M. S. (1999) *To Err Is Human: Building a Safer Health System; Washington (DC), The Institute of Medicine (IOM).*

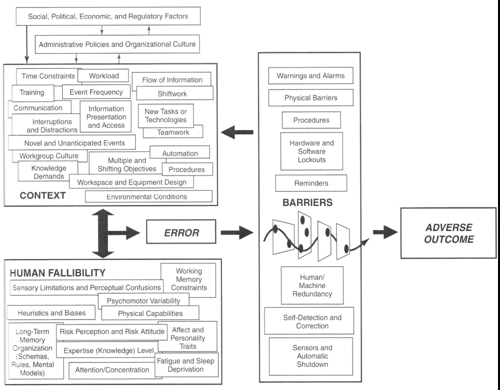
Slide 7-45 Definitions of medical errors

- 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 reserves 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 appropriate medical care.
- Preventable adverse event = Injury due to a non- intercepted serious error in medical care.

Rothschild et al. (2005) The Critical Care Safety Study: The incidence and nature of adverse events and serious medical errors in intensive care. *Critical Care Medicine, 33, 8, 1694.*

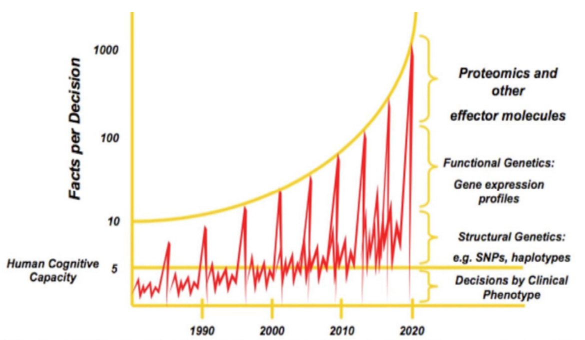
Slide 7-46 Framework for understanding human error

Sharit, J. (2006) Human Error. In: Salvendy, G. (Ed.) *Handbook of Human Factors and Ergonomics, Third Edition. Hoboken (NJ), Wiley, 708-760.*



Thank you!

Slide 7-47 Future Outlook



William Stead, IOM Meeting, 8 October 2007. Growth in facts affecting provider decisions versus human cognitive capacity.

Sample Questions (1)

- What is still considered 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)

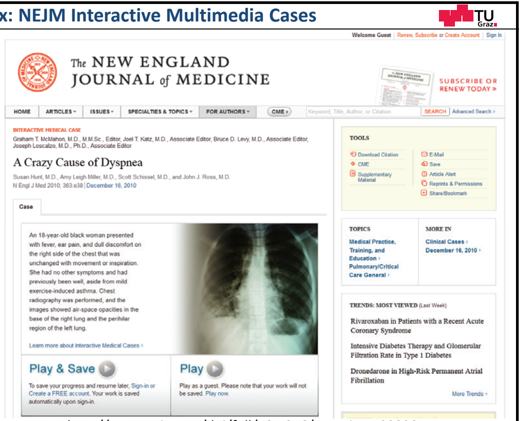
- Please contrast the decision making process with the data mining process!
- Why is Signal Detection Theory important for us?
- Please provide an Example for the application of Bayes' Theorem!
- How does Differential Diagnosis work?
- How can we apply Rough Set Theory for differential diagnostics?
- 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?

Some Useful Links

- http://www.anaesthetist.com/mnm/stats/roc
- http://sbml.org
- http://www.hci4all.at
- http://www.lcb.uu.se/tools/rosetta
- http://wise.cgu.edu/sdtmod/overview.asp (excellent Tutorial on SDT)
- http://www.iom.edu (Institute of Medicine)
- http://www.ahrq.gov/qual/patientsafetyix.htm (Agency for Health Care Research and Quality)
- http://www.fda.gov/drugs/drugsafety/medicationerrors/default.htm (Food and Drug Administration, FDA, medication errors)

Appendix: NEJM Interactive Multimedia Cases

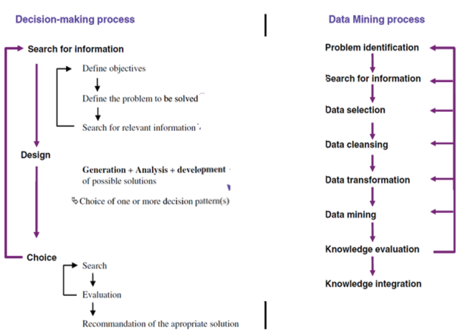
Hunt, S., Miller, A. L., Schissel, S. & Ross, J. J. (2010) A Crazy Cause of Dyspnea. Interactive Multimedia Case *New England Journal of Medicine, 363, 25, e38.*



http://www.nejm.org/doi/full/10.1056/NEJMimc1008281

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 and Decision: Cognitive Science & Human-Computer Interaction
- 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
- 12. Methodology for Info Systems: System Design, Usability & Evaluation



Slide 7-34 Rough Set Theory Example Symptom: Headache 2

- The atomic formula over
- $B \subseteq A \cup \{d\}$ and V are expressions of the form [a = v]
- called descriptors over B, where $a \in B$ and $v \in Va$.
- The set F(B, V) of formulas over B is the least set containing all atomic formulas over B and closed with respect to disjunction, conjunction and negation. For example, [location = occular] is a descriptor of B.
- For each $f \in F(B, V)$, fA denote the meaning of f in A, i.e., the set of all objects in U with property f, defined inductively as follows.
- 1. If f is of the form [a = v] then, fA = $\{s \in U | a(s) = v\}$
- 2. $(f \wedge g)A = fA \cap gA$; $(f \vee g)A = fA \vee gA$; $(\neg f)A = U - fa$
- For example, f = [location = whole] and fA = {2, 4, 5, 6}. As an example of a conjunctive formula, g = [location = whole] $\wedge$ [nausea = no] is a descriptor of U and fA is equal to glocation, nausea = {2, 5}.

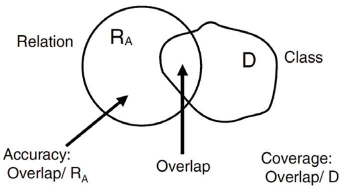
Slide 7-35 Classification Accuracy and Coverage

Definition 1. *Let R and D denote a formula in $F(B,V)$ and a set of objects which belong to a decision d. Classification accuracy and coverage(true positive rate) for $R \to d$ is defined as:*

$$\alpha_R(D) = \frac{|R_A \cap D|}{|R_A|} (= P(D|R)), \text{ and}$$

$$\kappa_R(D) = \frac{|R_A \cap D|}{|D|} (= P(R|D)),$$

where $|S|$, $\alpha_R(D)$, $\kappa_R(D)$ and $P(S)$ denote the cardinality of a set S, a classification accuracy of R as to classification of D and coverage (a true positive rate of R to D), and probability of S, respectively.

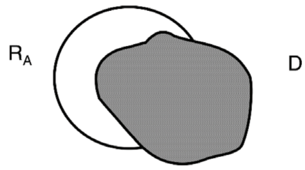


Tsumoto (2006)

Slide 7-36 Probabilistic Rules – modus ponens

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

$$R \overset{\alpha,\kappa}{\to} d \quad s.t. \; R = \wedge_j [a_j = v_k], \alpha_R(D) \geq \delta_\alpha$$
$$\text{and } \kappa_R(D) \geq \delta_\kappa,$$



$$\boxed{R \to D \quad s.t. \quad \alpha_R(D) > \delta_\alpha, \; \kappa_R(D) > \delta_\kappa}$$

Tsumoto (2006)

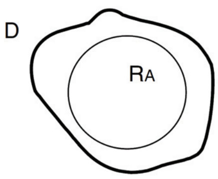
Slide 7-37 Positive Rules

A positive rule is defined as a rule supported by only positive examples, the classification accuracy of which is equal to 1.0. It is notable that the set supporting this rule corresponds to a subset of the lower approximation of a target concept, which is introduced in rough sets [1]. Thus, a positive rule is represented as:

$$R \to d \quad s.t. \qquad R = \wedge_j [a_j = v_k], \quad \alpha_R(D) = 1.0$$

Figure 4 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 exactly equivalent to the classic proposition $R \to d$. In the above example, one positive rule of "m.c.h." (muscle contraction headache) is:

$$[nausea = no] \to m.c.h. \quad \alpha = 3/3 = 1.0.$$



Tsumoto (2006)

Slide 7-38 Exclusive Rules

Before defining a negative rule, let us first introduce an exclusive rule, the contrapositive of a negative rule [7]. 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 represents the necessity condition of a decision. It is notable that the set supporting an exclusive rule corresponds to the upper approximation of a target concept, which is introduced in rough sets [1]. Thus, an exclusive rule is represented as:

$$R \to d \quad s.t. \qquad R = \vee_j [a_j = v_k], \quad \kappa_R(D) = 1.0.$$

Figure 4 shows the Venn diagram of a exclusive rule. As shown in this figure, the meaning of R is a superset of that of D. This diagram is exactly equivalent to the classic proposition $d \to R$. In the above example, the exclusive rule of "m.c.h." is:

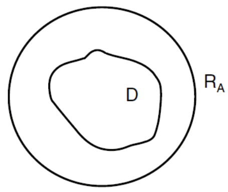
$$[M1 = yes] \vee [nau = no] \to m.c.h$$

From the viewpoint of propositional logic, an exclusi as:

$$d \to \vee_j [a_j = v_k],$$

because the condition of an exclusive rule correspor of conclusion d. Thus, it is easy to see that a ne contrapositive of an exclusive rule:

$$\wedge_j \neg [a_j = v_k] \to \neg d,$$



Tsumoto (2006)

Slide 7-39 Negative Rule

of conclusion d. Thus, it is easy to see that a negative rule is defined as the contrapositive of an exclusive rule:

$$\wedge_j \neg [a_j = v_k] \to \neg d,$$

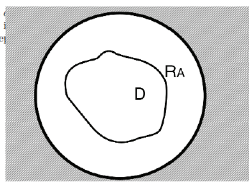
which means that if a case does not satisfy any attribute value pairs in the condition of a negative rules, then we can exclude a decision d from candidates. For example, the negative rule of m.c.h. is:

$$\neg [M1 = yes] \wedge \neg [nausea = no] \to \neg m.c.h.$$

In summary, a negative rule is defined as:

$$\wedge_j \neg [a_j = v_k] \to \neg d \quad s.t. \quad \forall [a_j = v_k] \; \kappa_{[a_j = v_k]}(D) = 1.0,$$

where D denotes a set of samples which belong to a class (Venn diagram of a negative rule. As shown in this figure, i negative region is the "positive region" of "negative concej



Tsumoto (2006)

Slide 7-40 Example: Algorithms for Rule Induction

procedure *Exclusive and Negative Rules*;
  var
    L : *List*;
      /* A list of elementary attribute-value pairs */
  begin
    $L := P_0$;
    /* P_0: A list of elementary attribute-value pairs given in a database */
    while ($L \neq \{\}$) **do**
      begin
        Select one pair $[a_i = v_j]$ from L;
        if ($[a_i = v_j]_A \cap D \neq \phi$) **then do** /* D: positive examples of a target class d */
          begin
          $L_{ir} := L_{ir} + [a_i = v_j]$; /* Candidates for Positive Rules */
          if ($\kappa_{[a_i = v_j]}(D) = 1.0$)
          then $R_{er} := R_{er} \wedge [a_i = v_j]$;
            /* Include $[a_i = v_j]$ into the formula of Exclusive Rule */
          end
      $L := L - [a_i = v_j]$;
      end
    Construct Negative Rules:
        Take the contrapositive of R_{er}.
  end {*Exclusive and Negative Rules*};

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

$$P(\mathbf{d} \mid h) = P(a_1, \ldots, a_T \mid h) = \prod_t P(a_t \mid h)$$

  - 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