185.A83 Machine Learning for Health Informatics
2021S, VU, 2.0 h, 3.0 ECTS
Andreas Holzinger, Rudolf Freund
Marcus Bloice, Florian Endel, Anna Saranti

From data to probabilistic information and knowledge

Contact: andreas.holzinger AT tuwien.ac.at

What will we learn in this lecture?

- 00 Reflection
- 01 **Data** – the underlying physics of data
- 02 Biomedical data sources – taxonomy of data
- 03 Data integration, mapping, fusion
- 04 **Information** - Theory – Entropy
- 05 **Knowledge** Representation – Ontologies – Medical Classifications
00 Reflection
Warm-up Quiz

1. ... 
2. Uncertainty
3. $p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta) \cdot p(\theta)}{p(\mathcal{D})}$
4. ... 
5. ... 
6. Medical Decision Making
7. ... 
8. context
9. ...
Where is the Biologist in this image?

This image is used according UrhG §42 lit. f Abs 1 as “Belegfunktion” for discussion with students

What happens if you feed in data into your ML pipeline?

Image Source: Randall Munroe [https://xkcd.com]

This image is used according UrhG §42 lit. f Abs 1 as “Belegfunktion” for discussion with students.
### How to ensure good data quality assessment?

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessibility</td>
<td>the extent to which data is available, or easily and quickly retrievable</td>
</tr>
<tr>
<td>Appropriate Amount of Data</td>
<td>the extent to which the volume of data is appropriate for the task at hand</td>
</tr>
<tr>
<td>Believability</td>
<td>the extent to which data is regarded as true and credible</td>
</tr>
<tr>
<td>Completeness</td>
<td>the extent to which data is not missing and is of sufficient breadth and depth for the task at hand</td>
</tr>
<tr>
<td>Concise Representation</td>
<td>the extent to which data is compactly represented</td>
</tr>
<tr>
<td>Consistent Representation</td>
<td>the extent to which data is presented in the same format</td>
</tr>
<tr>
<td>Ease of Manipulation</td>
<td>the extent to which data is easy to manipulate and apply to different tasks</td>
</tr>
<tr>
<td>Free-of-Error</td>
<td>the extent to which data is correct and reliable</td>
</tr>
<tr>
<td>Interpretability</td>
<td>the extent to which data is in appropriate languages, symbols, and units, and the definitions are clear</td>
</tr>
<tr>
<td>Objectivity</td>
<td>the extent to which data is unbiased, unprejudiced, and impartial</td>
</tr>
<tr>
<td>Relevancy</td>
<td>the extent to which data is applicable and helpful for the task at hand</td>
</tr>
<tr>
<td>Reputation</td>
<td>the extent to which data is highly regarded in terms of its source or content</td>
</tr>
<tr>
<td>Security</td>
<td>the extent to which access to data is restricted appropriately to maintain its security</td>
</tr>
<tr>
<td>Timeliness</td>
<td>the extent to which the data is sufficiently up-to-date for the task at hand</td>
</tr>
<tr>
<td>Understandability</td>
<td>the extent to which data is easily comprehended</td>
</tr>
<tr>
<td>Value-Added</td>
<td>the extent to which data is beneficial and provides advantages from its use</td>
</tr>
</tbody>
</table>

“The value of data lies in reusability”.

What are the attributes that make data reusable?

**Findable:** metadata -persistent identifier

**Accessible:** retrievable by humans and machines through standards, open and free by default; authentication and authorization where necessary

**Interoperable:** metadata use a ‘formal, accessible, shared, and broadly applicable language for knowledge representation’.

**Reusable:** metadata provide rich and accurate information; clear usage license; detailed provenance.

[https://www.go-fair.org/fair-principles](https://www.go-fair.org/fair-principles)
01 The underlying physics of data
What is this?
What are the key problems in (medical) data science?

- Heterogeneous, distributed, inconsistent data sources (need for data integration & fusion) [1]
- **Complex data** (high-dimensionality – challenge of dimensionality reduction and visualization) [2]
- Noisy, uncertain, missing, dirty, and imprecise, imbalanced data (challenge of pre-processing)
- The discrepancy between data-information-knowledge (various definitions)
- **Big data** sets in high-dimensions (manual handling of the data is often impossible) [3]

Why can data in ML often not be represented by a simple model?

- Data in traditional Statistics
  - Low-dimensional data ($< \mathbb{R}^{100}$)
  - Problem: Much noise in the data
  - Not much structure in the data but it can be represented by a simple model

- Data in Machine Learning
  - High-dimensional data ($\gg \mathbb{R}^{100}$)
  - Problem: not noise, but complexity
  - Much structure, but the structure can **not** be represented by a simple model

Why is the curse of dimensionality for us relevant?

What is the difference between Data – Information – Knowledge?

Knowledge := a set of expectations
Biomedical informatics (BMI) is the interdisciplinary field that studies and pursues the effective use of biomedical data, information, and knowledge for scientific problem solving, and decision making, motivated by efforts to improve human health.

So, where does the data come from?

What types of data do we have in health?

http://www.nytimes.com/2012/05/06/books/review/turings-cathedral-by-george-dyson.html
How can we classify traditional data structures?


Aggregated attribute = a homomorphic map $H$ from a relational system $<A; \approx>$ into a relational system $<B; \rightarrow>$; where $A$ and $B$ are two distinct sets of data elements. This is in contrast with other attributes since the set $B$ is the set of data elements instead of atomic values.
What was the original work on the theory of scales claiming?

<table>
<thead>
<tr>
<th>Scale</th>
<th>Empirical Operation</th>
<th>Mathem. Group Structure</th>
<th>Transf. in ( \mathbb{R} )</th>
<th>Basic Statistics</th>
<th>Mathematical Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOMINAL</td>
<td>Determination of equality</td>
<td>Permutation ( x' = f(x) ) ( x \ldots \text{1-to-1} )</td>
<td>( x \mapsto f(x) )</td>
<td>Mode, contingency correlation</td>
<td>=, ≠</td>
</tr>
<tr>
<td>ORDINAL</td>
<td>Determination of more/less</td>
<td>Isotonic ( x' = f(x) ) ( x \ldots \text{monotonic incr.} )</td>
<td>( x \mapsto f(x) )</td>
<td>Median, Percentiles</td>
<td>=, ≠, &gt;, &lt;</td>
</tr>
<tr>
<td>INTERVAL</td>
<td>Determination of equality of intervals or differences</td>
<td>General linear ( x' = ax + b )</td>
<td>( x \mapsto rx+s )</td>
<td>Mean, Std.Dev. Rank-Order Corr., Prod.-Moment Corr.</td>
<td>=, ≠, &gt;, &lt;, -, +</td>
</tr>
<tr>
<td>RATIO</td>
<td>Determination of equality or ratios</td>
<td>Similarity ( x' = ax )</td>
<td>( x \mapsto rx )</td>
<td>Coefficient of variation</td>
<td>=, ≠, &gt;, &lt;, -, +, *, ÷</td>
</tr>
</tbody>
</table>
What levels of data taxonomy can we identify?

- **Physical level** -> bit = binary digit = basic indissoluble unit (= Shannon, Sh), ≠ Bit (!)
in Quantum Systems -> qubit

- **Logical Level** -> integers, booleans, characters, floating-point numbers, alphanumeric strings, ...

- **Conceptual (Abstract) Level** -> data-structures, e.g. lists, arrays, trees, graphs, ...

- **Technical Level** -> Application data, e.g. text, graphics, images, audio, video, multimedia, ...

- **“Hospital Level”** -> Narrative (textual) data, numerical measurements (physiological data, lab results, vital signs, ...), recorded signals (ECG, EEG, ...), Images (x-ray, MR, CT, PET, ...); -omics
Where do data come from at Hospital Level?

- **Clinical workplace data sources**
  - Medical documents: text (non-standardized (“free-text”), semi-structured, standard terminologies (ICD, SNOMED-CT)
  - Measurements: lab, time series, ECG, EEG, EOG, ...
  - Surveys, Clinical study data, trial data

- **Image data sources**
  - Radiology: MRI (256x256, 200 slices, 16 bit per pixel, uncompressed, ~26 MB); CT (512x512, 60 slices, 16 bit per pixel, uncompressed ~32MB; MR, US;
  - Digital Microscopy : WSI (15mm slide, 20x magn., 24 bits per pixel, uncompressed, 2,5 GB, WSI 10 GB; confocal laser scanning, etc.

- **-omics data sources**
  - Sanger sequencing, NGS whole genome sequencing (3 billion reads, read length of 36) ~ 200 GB; NGS exome sequencing (“only” 110,000,000 reads, read length of 75) ~7GB; Microarray, mass-spectrometry, gas chromatography, ...

What are typical examples of imaging data?

Image Source: Laboratory of Neuro Imaging, USC
Why is Digital Pathology interesting?

How is a WSI produced?


(Image Sources: Pathology Graz)
What is the current state of the art in machine learning for pathology?

What about the ground truth?

Why is Neonatal Screening a good example for data generation?

What is an example for the Data Structure “list”?

Why is the data structure graph so versatile?

Evolutionary dynamics act on populations. Neither genes, nor cells, nor individuals evolve; only populations evolve.

\[
W = \begin{bmatrix}
0 & w_{12} & w_{13} & 0 & 0 \\
0 & 0 & w_{23} & w_{24} & 0 \\
w_{31} & 0 & 0 & 0 & w_{35} \\
0 & w_{42} & 0 & 0 & 0 \\
0 & 0 & 0 & w_{54} & 0
\end{bmatrix}
\]

What is a tree?

02 Biomedical data sources: Taxonomy of data
What are origins of health-related data?

Atom
Molecule
Virus
Bacteria
Cell
Tissue
Individual
Collective
Ecosystem

Why is data integration in health an unsolved problem?

Private Health vault data
Electronic health record data
Physiological data
Laboratory results

Exposome
Environmental data
Air pollution
Exposure (toxicants)

Collective data
Social data
Fitness, Wellness data
Ambient Assisted Living data
(Non-medical) personal data

Metabolomics
Chemical processes
Cellular reactions
Enzymatic reactions

Foodomics, Lipidomics
Nutrition data (Nutrigenomics)
Diet data (allergenics)

Microbiomes
Microorganisms processes
Plants, Fungi, ...

Imaging data
X-Ray, ultrasound, MR, CT, PET,
cams, observation (e.g. sleep
laboratory), gait (child walking)

Proteomics
Protein-Protein Interactions

Transcriptomics
RNA, mRNA, rRNA, tRNA

Epigenetics
Epigenetic modifications

Genomics
Why is the human exposome important?

Internal
- metabolism, endogenous hormones, body morphology, physical activity, gut micro flora, inflammation, aging etc.

General external
- social capital, education, financial status, psychological stress, urban-rural environment, climate, etc.

Specific external
- radiation, infectious agents, chemical contaminants and pollutants, diet, lifestyle factors (e.g. tobacco, alcohol), occupation, medical interventions, etc.


https://human-centered.ai/project/eu-project-heap-human-exposome-assessment-platform
What is a good example for the level “cell”? 

What is life according to Erwin Schrödinger?

to reproduce …

to grow …

to evolve …

to self-replicate …

to generate/utilize energy …

to process information …

Where do we get open data sets?

- Billions of biological data sets are openly available, here only some examples:
  - General Repositories:
    - GenBank, EMBL, HMCA, ...
  - Specialized by data types:
    - UniProt/SwissProt, MMMP, KEGG, PDB, ...
  - Specialized by organism:
    - WormBase, FlyBase, NeuroMorpho, ...
  - [https://human-centered.ai/open-data-sets](https://human-centered.ai/open-data-sets)
What are the most important *omics data?

- **Genomics** (sequence annotation)
- **Transcriptomics** (microarray)
- **Proteomics** (Proteome Databases)
- **Metabolomics** (enzyme annotation)
- **Fluxomics** (isotopic tracing, metabolic pathways)
- **Phenomics** (biomarkers)
- **Epigenomics** (epigenetic modifications)
- **Microbiomics** (microorganisms)
- **Lipidomics** (pathways of cellular lipids)
What is *omics data integration?

<table>
<thead>
<tr>
<th>omics data integration</th>
<th>Genomics (sequence annotation)</th>
<th>Transcriptomics (microarray, SAGE)</th>
<th>Proteomics (abundance, post-translational modification)</th>
<th>Metabolomics (metabolite abundance)</th>
<th>Protein-DNA interactions (ChIP-chip)</th>
<th>Protein-protein interactions (yeast 2H, coAF-MS)</th>
<th>Fluxomics (isotopic tracing)</th>
<th>Phenomics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomics</td>
<td>ORF validation</td>
<td>SNP effect on protein activity or abundance</td>
<td>Enzyme annotation</td>
<td>Metabolic-transcriptional response</td>
<td>Signaling cascades</td>
<td>Protein-protein interactions</td>
<td>Dynamic network responses</td>
<td>Functional annotation</td>
</tr>
<tr>
<td></td>
<td><em>Regulatory element identify</em></td>
<td><em>Protein transcript correlation</em></td>
<td><em>Enzyme annotation</em></td>
<td><em>Metabolic pathway bottlenecks</em></td>
<td><em>Dynamic network responses</em></td>
<td><em>Pathway identification activity</em></td>
<td><em>Functional annotation</em></td>
<td><em>Biomarkers</em></td>
</tr>
<tr>
<td></td>
<td><em>identification</em>^4^</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>^7,11^,13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>^9,10^2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>^9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>^9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>^9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>^9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>^9</td>
</tr>
</tbody>
</table>

Examples for lower dimensional data?

- **0-D data** = a **data point** existing isolated from other data, e.g. integers, letters, Booleans, etc.
- **1-D data** = consist of a **string** of 0-D data, e.g. Sequences representing nucleotide bases and amino acids, SMILES etc.
- **2-D data** = having **spatial component**, such as images, NMR-spectra etc.
- **2.5-D data** = can be stored as a 2-D matrix, but can represent biological entities in three or more dimensions, e.g. **PDB records**
- **3-D data** = having **3-D spatial component**, e.g. image voxels, e-density maps, etc.
- **H-D Data** = data having arbitrarily **high dimensions**
SMILES (Simplified Molecular Input Line Entry Specification)

... is a compact machine and human-readable chemical nomenclature:

e.g. Viagra:
CCc1nn(C)c2c(=O)[nH]c(nc12)c3cc(ccc3OCC)S(=O)(=O)N4CCN(C)CC4

...is Canonicalizable
...is Comprehensive
...is Well Documented

What is a typical example for 2-D data (bivariate data)?

Example: 2.5-D data (structural information & metadata)?
What are 3-D Voxel data (volumetric picture elements) ?

03 Data Integration, mapping, fusion
What do we mean with data integration – information fusion?

What is the goal of data integration?

**Goal:**
Unified View for decision support (“what is relevant?”)

What is information fusion?

Constructing a multi-modal interaction & correspondence graph (ICG)

Interesting signals from each modality (time-based, image, structured & unstructured) are connected according to pre-defined rules. Each modality’s features lie in their own, un-aligned concept spaces.

https://doi.org/10.1016/j.inffus.2021.01.008
Why may information bridge the gap between both worlds?

Our central hypothesis:
Information may bridge this gap

Why is imaging data along not enough?

What is translational health?

Translational Medicine Continuum

Bench ← T1 → Bedside ← T2 → Community ← T3 → Policy

Translational Bioinformatics

Clinical Research Informatics

Bio-informatics

Imaging Informatics

Clinical Informatics

Public Health Informatics

Molecules & Cells

Tissues & Organs

Individuals

Populations

Biomedical Informatics Continuum

What are the key problems in medical data management?
What is the combining link?

Biomedical R&D data (e.g. clinical trial data)

Clinical patient data (e.g. EPR, lab, reports etc.)

The combining link is text

Health business data (e.g. costs, utilization, etc.)

Private patient data (e.g. AAL, monitoring, etc.)

Why is medical text important?

Digression: Medical Communication
Why is medical work relying on team communication?

- ... and requires a lot of information exchange..

What are problems of the medical report?

Radiologischer Befund

Kurzamnese: St.p. SHT
Fragestellung: -
Untersuchung: Thorax eine Ebene liegend

SB
Bewegungsartefakte. Zustand nach Schädelhirmtrauma.

Das Cor in der Größennorm, keine akuten Stauungszeichen.
Fragliches Infiltrat parahilär li. im UF, RW-Erguss li.

Zustand nach Anlage eines ET, die Spitze ca. 5cm cranial der Bifurkation, lieg. MS, orthotop
positioniert. ZVK über re., die Spitze in Proj. auf die VCS. Kein Hinweis auf Pneumothorax.
Der re. Rezessus frei.

Mit kollegialen Grüßen

***Elektronische Freigabe durch am 09.05.2006 ***

„die Antrumschleimhaut ist durch Lymphozyten infiltriert“
„lymphozytäre Infiltration der Antrummukosa“
„Lymphozyteninfiltration der Magenschleimhaut im Antrumbereich“
Why does Language Understanding require knowledge?

- Syntax
- Semantics
- Pragmatics
- Context
- (Emotion)


Image Source: https://cs.stanford.edu/people/karpathy/deepimagesent/
Why is text a good example for Non-Standardized Data

What are the typical challenges in data-driven medicine?

- Increasingly large data sets due to **data-driven medicine** [1]
- Increasing amounts of **non-standardized** data and **un-structured information** (e.g. “free text”)
- **Data quality**, data **integration**, universal **access**
- **Privacy**, security, safety, data protection, data ownership, fair use of data [2]
- **Time** aspects in databases [3]

---

Digression:
Data Augmentation
What is data augmentation?

- Generation of artificial data via expansion of your dataset
- Why?
  - Neural networks require “big data” so augmentation is now basically part of most all deep learning projects
  - It is also used to address issues with class imbalance
  - It is a cheap and relatively easy way to get more data, which will almost certainly improve the accuracy of a trained model
  - It improves model generalisation, model accuracy, and can control overfitting
- Image augmentation is most common, because text augmentation is much harder, and DL is applied to images
- done by making label-preserving transformations to the original images (e.g. rotation, zooming, cropping, ...)

How does image augmentation work?

04 Information Theory & Entropy
Why are probabilistic models so important for medicine?

- **Boolean models**
- **Algebraic models**
- **Probabilistic models (*)**

*) Our probabilistic models describes data which we can observe from our environment – and if we use the mathematics of probability theory, in order to express the uncertainties around our model then the inverse probability allows us to infer unknown unknowns ... learning from data and making predictions – the core essence of machine learning and of vital importance for health informatics.

Why is life complex information?


What is information?
Why is the work of Bayes, Price, Laplace so important for us?

Bayes’ Rule in words

d ... data; h ... hypothesis

$H = \{H_1, H_2, \ldots, H_n\}$ ... Hypothesis space

$$p(h|d) = \frac{p(d|h)p(h)}{\sum_{h' \in H} p(d|h')p(h')}$$

- **Posterior Probability**
- **Likelihood**
- **Prior Probability**
- **Evidence = marginal likelihood**
- **Sum over space of alternative hypotheses**
- Information is the reduction of uncertainty
- If something is 100 % certain its uncertainty = 0
- Uncertainty is max. if all choices are equally probable (I.I.D)
- Uncertainty (as information) sums up for independent sources
Entropy as measure for disorder

http://www.scottaaronson.com
What are the origins of Entropy?

Bernoulli (1713)  
Principle of Insufficient Reason

Maxwell (1859), Boltzmann (1871), Gibbs (1902)  
Statistical Modeling of problems in physics

Bayes (1763), Laplace (1770)  
How to calculate the state of a system with a limited number of expectation values

Jeffreys, Cox (1939-1948)  
Statistical Inference

Pearson (1900)  
Goodness of Fit measure

Fisher (1922)  
Maximum Likelihood

Shannon (1948)  
Information Theory

Bayesian Statistics  
Entropy Methods  
Generalized Entropy

See next slide

What current Entropy methods can we use?

Entropic Methods

- Jaynes (1957)
  - Maximum Entropy (MaxEn)

- Adler et al. (1965)
  - Topology Entropy (TopEn)

- Pincus (1991)
  - Approximate Entropy (ApEn)

- Richman (2000)
  - Sample Entropy (SampEn)

Generalized Entropy

- Mowshowitz (1968)
  - Graph Entropy (MinEn)

- Posner (1975)
  - Minimum Entropy (MinEn)

- Rubinstein (1997)
  - Cross Entropy (CE)

How does Approximate Entropy work?

Let: \( \langle x_n \rangle = \{x_1, x_2, \ldots, x_N \} \)

\[
\tilde{X}_i = (x_i, x_{(i+1)}, \ldots, x_{(i+m-1)})
\]

\[
\|\tilde{X}_i, \tilde{X}_j\| = \max_{k=1,2,\ldots,m} (|x_{(i+k-1)} - x_{(j+k-1)}|)
\]

\[
\tilde{H}(m, r) = \lim_{N \to \infty} [\phi^m(r) - \phi^{m+1}(r)]
\]

\[
C_r^m(i) = \frac{N^m(i)}{N - m + 1} \quad \phi^m(r) = \frac{1}{N - m + 1} \sum_{t=1}^{N-m+1} \ln C_r^m(i)
\]

What do we have to consider when measuring entropy?
What is the main advantage of entropy measures?

Cross-Entropy
Kullback-Leibler
Divergence
What is the difference between Entropy – KL-divergence and MI?

- **Entropy:**
  - Measure for the **uncertainty** of random variables
- **Kullback-Leibler divergence:**
  - comparing two distributions
- **Mutual Information:**
  - measuring the **correlation** of two random variables
ON INFORMATION AND SUFFICIENCY

BY S. KULLBACK AND R. A. LEIBLER

The George Washington University and Washington, D. C.

1. Introduction. This note generalizes to the abstract case Shannon’s definition of information [15], [16]. Wiener’s information (p. 75 of [18]) is essentially the same as Shannon’s although their motivation was different (cf. footnote 1, p. 95 of [16]) and Shannon apparently has investigated the concept more completely. R. A. Fisher’s definition of information (intrinsic accuracy) is well known (p. 709 of [6]). However, his concept is quite different from that of Shannon and Wiener, and hence ours, although the two are not unrelated as is shown in paragraph 2.

R. A. Fisher, in his original introduction of the criterion of sufficiency, required “that the statistic chosen should summarize the whole of the relevant information supplied by the sample,” (p. 316 of [8]). Halmos and Savage in a recent paper, one of the main results of which is a generalization of the well known Fisher-Neyman theorem on sufficient statistics to the abstract case, conclude, “We think that confusion has from time to time been thrown on the subject by . . . , and (c) the assumption that a sufficient statistic contains all the information in only the technical sense of ‘information’ as measured by variance,” (p. 241 of [8]). It is shown in this note that the information in a sample as defined herein, that is, in the Shannon-Wiener sense cannot be increased by any statistical operations and is invariant (not decreased) if and only if sufficient statistics are employed. For a similar property of Fisher’s information see p. 717 of [6], Doob [19].

We are also concerned with the statistical problem of discrimination (3), (17)), by considering a measure of the “distance” or “divergence” between statistical populations [11], [2], [15]) in terms of our measure of information. For the statistician two populations differ more or less according as to how difficult it is to discriminate between them with the best test [14]. The particular measure of divergence we use has been considered by Jeffreys ([10], [11]) in another connection. He is primarily concerned with its use in providing an invariant density of a priori probability. A special case of this divergence is Mahalanobis’ generalized distance [13].
Why should we remember Shannon Entropy?

\[ H[x] = - \sum_x p(x) \log_2 p(x) \]


Important quantity in
• coding theory
• statistical physics
• machine learning
What is Conditional Entropy?

\[
H[y|x] = - \int \int p(y, x) \ln p(y|x) \, dy \, dx
\]

\[
H[x, y] = H[y|x] + H[x]
\]
When do we need the Kullback-Leibler Divergence?

KL\((p \| q)\) = \(- \int p(x) \ln q(x) \, dx - \left( - \int p(x) \ln p(x) \, dx \right)\)

= \(- \int p(x) \ln \left( \frac{q(x)}{p(x)} \right) \, dx\)

KL\((p \| q) \approx \frac{1}{N} \sum_{n=1}^{N} \left\{ - \ln q(x_n | \theta) + \ln p(x_n) \right\}\)

KL\((p \| q) \geq 0\)

KL-divergence is often used to measure the distance between two distributions.
What is important to note when using KL divergence?

$q^* = \arg\min_q D_{KL}(p||q)$

$q^* = \arg\min_q D_{KL}(q||p)$

$KL(p||q) \neq KL(q||p)$

In summary: Why do we use Entropy measures generally?

- ... are robust against noise;
- ... can be applied to complex time series with good replication;
- ... is finite for stochastic, noisy, composite processes;
- ... the values correspond directly to irregularities – good for detecting anomalies
05 Knowledge Representation
What is medical knowledge? Where does the ground truth come from?

Why is logic insufficient for solving complex real-world problems?

- Logical representations are based on
  - Facts about the world (true or false)
  - These facts can be combined with logical operators
  - Logical inference is based on certainty

What are examples for famous knowledge representations?

<table>
<thead>
<tr>
<th>Mathematical Logic</th>
<th>Psychology</th>
<th>Biology</th>
<th>Statistics</th>
<th>Economics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aristotle</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Descartes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boole</td>
<td>James</td>
<td></td>
<td>Laplace</td>
<td>Bentham Pareto</td>
</tr>
<tr>
<td>Hebb</td>
<td>Lashley</td>
<td></td>
<td>Bernoulli</td>
<td>Friedman</td>
</tr>
<tr>
<td>Bruner</td>
<td>Rosenblatt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miller</td>
<td>Ashby</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Newell, Simon</td>
<td>Lettvin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McCulloch, Pitts</td>
<td>Heubel, Weisel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayes</td>
<td>Tversky, Kahneman</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frege, Peano</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goedel, Post, Church</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davis, Putnam, Robinson</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Logic

<table>
<thead>
<tr>
<th>Logic</th>
<th>SOAR</th>
<th>Connectionism</th>
<th>Causal</th>
<th>Rational</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROLOG</td>
<td>KBS, Frames</td>
<td></td>
<td>Networks</td>
<td>Agents</td>
</tr>
</tbody>
</table>

What does Formalization versus Expressivity mean?

What do you need for developing clinical decision support systems?

Why is the history of “Deep Learning” interesting for us?

Image source: Andrew Beam, Department of Biomedical Informatics, Harvard Medical School
https://slides.com/beamandrew/deep-learning-101/#/12
This image is used according UrhG §42 lit. f Abs 1 as “Belegfunktion” for discussion with students
Why were early Decision Support Systems no success?

What is the connection of early expert systems with current xAI?

How did an expert system work?

Shortliffe & Buchanan (1984)
What was the certainty factor in the MYCIN System?

- 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 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.
  - \( CF[h] = +1 \rightarrow h \) is 100% true
  - \( CF[h] = -1 \rightarrow h \) is 100% false
How does an example of the Certainty Factor look like?

\( h_1 = \text{The identity of ORGANISM-1 is streptococcus} \)

\( h_2 = \text{PATIENT-1 is febrile} \)

\( h_3 = \text{The name of PATIENT-1 is John Jones} \)

\[ \begin{align*}
\text{CF}[h_1,E] &= .8 \quad \text{: There is strongly suggestive evidence (.8) that the identity of ORGANISM-1 is streptococcus} \\
\text{CF}[h_2,E] &= -.3 \quad \text{: There is weakly suggestive evidence (.3) that PATIENT-1 is not febrile} \\
\text{CF}[h_3,E] &= +1 \quad \text{: It is definite (1) that the name of PATIENT-1 is John Jones} 
\end{align*} \]

Ontologies
What happens if you put the word “Jaguar” into a search engine?

Image Sources: The images are in the public domain and are used according UrhG §42 lit. f Abs 1 as “Belegfunktion” for discussion with students.
Who created the first ontology?

* 384 BC † 322 BC


Later: Porphyry (≈ 234-305) tree
Aristotle attempted to **classify the things in the world** - where it is employed to describe the existence of beings in the world;

Artificial Intelligence and Knowledge Engineering deals also with **reasoning about models of the world**.

Therefore, AI researchers adopted the term 'ontology' to describe **what can be computationally represented** of the world within a program.

"An ontology is a formal, explicit specification of a shared conceptualization".

- A 'conceptualization' refers to an **abstract model** of some phenomenon in the world by having identified the relevant concepts of that phenomenon.
- 'Explicit' means that the type of concepts used, and the constraints on their use are **explicitly defined**.
What are the principles and methods of Knowledge Engineering?

Where are ontologies used today?

What is a Gene Ontology?


http://geneontology.org/
What is the difference between an ontology and a terminology?

- Ontology = a structured description of a domain in form of concepts ↔ relations;
- The IS-A relation provides a taxonomic skeleton;
- Other relations reflect the domain semantics;
- Formalizes the terminology in the domain;
- Terminology = terms definition and usage in the specific context;
- Knowledge base = instance classification and concept classification;
- Classification provides the domain terminology ...
What are the conditions an ontology may satisfy?

1. In addition to the IS-A relationship, partitive (meronomic) relationships may hold between concepts, denoted by PART-OF. Every PART-OF relationship is irreflexive, asymmetric and transitive. IS-A and PART-OF are also called hierarchical relationships.

2. In addition to hierarchical relationships, associative relationships may hold between concepts. Some associative relationships are domain-specific (e.g., the branching relationship between arteries in anatomy and rivers in geography).

3. Relationships $r$ and $r'$ are inverses if, for every pair of concepts $x$ and $y$, the relations $(x, r, y)$ and $(y, r', x)$ hold simultaneously. A symmetric relationship is its own inverse. Inverses of hierarchical relationships are called INVERSE-IS-A and HAS-PART, respectively.

4. Every non-taxonomic relation of $x$ to $z$, $(x, r, z)$, is either inherited $(y, r, z)$ or refined $(y, r, z')$ where $z'$ is more specific than $z$) by every child $y$ of $x$. In other words, every child $y$ of $x$ has the same properties ($z$) as it parent or more specific properties ($z'$).

What is a semantic relationship?

What are typical medical ontologies?

<table>
<thead>
<tr>
<th>Name</th>
<th>Ref.</th>
<th>Scope</th>
<th># concepts</th>
<th># concept names</th>
<th>Subs. Hier.</th>
<th>Version / Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNOMED CT</td>
<td>[21]</td>
<td>Clinical medicine (patient records)</td>
<td>310,314</td>
<td>1 37 2</td>
<td>yes</td>
<td>July 31, 2007</td>
</tr>
<tr>
<td>LOINC</td>
<td>[24]</td>
<td>Clinical observations and laboratory tests</td>
<td>46,406</td>
<td>1 3 3</td>
<td>no</td>
<td>Version 2.21 (no “natural language” names)</td>
</tr>
<tr>
<td>FMA</td>
<td>[25]</td>
<td>Human anatomical structures</td>
<td>~72,000</td>
<td>1 2 1</td>
<td>yes</td>
<td>(not yet in the UMLS)</td>
</tr>
<tr>
<td>RxNorm</td>
<td>[31]</td>
<td>Standard names for prescription drugs</td>
<td>93,426</td>
<td>1 2 1</td>
<td>no</td>
<td>Aug. 31, 2007</td>
</tr>
<tr>
<td>NCI Thesaurus</td>
<td>[34]</td>
<td>Cancer research, clinical care, public information</td>
<td>58,868</td>
<td>1 100 2</td>
<td>yes</td>
<td>2007.05E</td>
</tr>
<tr>
<td>ICD-10</td>
<td>[36]</td>
<td>Diseases and conditions (health statistics)</td>
<td>12,318</td>
<td>1 1 1</td>
<td>no</td>
<td>1998 (tabular)</td>
</tr>
<tr>
<td>MeSH</td>
<td>[38]</td>
<td>Biomedicine (descriptors for indexing the literature)</td>
<td>24,767</td>
<td>1 208 5</td>
<td>no</td>
<td>Aug. 27, 2007</td>
</tr>
<tr>
<td>UMLS Meta</td>
<td>[41]</td>
<td>Terminology integration in the life sciences</td>
<td>1,4 M</td>
<td>1 339 2</td>
<td>n/a</td>
<td>2007AC (English only)</td>
</tr>
</tbody>
</table>

What notations of ontologies do we use?

1) Graph notations
   - Semantic networks
   - Topic Maps (ISO/IEC 13250)
   - Unified Modeling Language (UML)
   - Resource Description Framework (RDF)

2) Logic based
   - Description Logics (e.g., OIL, DAML+OIL, OWL)
   - Rules (e.g. RuleML, LP/Prolog)
   - First Order Logic (KIF – Knowledge Interchange Format)
   - Conceptual graphs
   - (Syntactically) higher order logics (e.g. LBase)
   - Non-classical logics (e.g. Flogic, Non-Mon, modalities)

3) Probabilistic/fuzzy
How does a graphical notation look like?

**What is the purpose of the Web Ontology Language OWL?**

<table>
<thead>
<tr>
<th>Axiom</th>
<th>DL syntax</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub class</td>
<td>$C_1 \sqsubseteq C_2$</td>
<td>Alga $\sqsubseteq$ Plant $\sqsubseteq$ Organism</td>
</tr>
<tr>
<td>Equivalent class</td>
<td>$C_1 \equiv C_2$</td>
<td>Cancer $\equiv$ Neoplastic Process</td>
</tr>
<tr>
<td>Disjoint with</td>
<td>$C_1 \sqsubseteq \neg C_2$</td>
<td>Vertebrate $\sqsubseteq \neg$ Invertebrate</td>
</tr>
<tr>
<td>Same individual</td>
<td>$x_1 \equiv x_2$</td>
<td>Blue_Shark $\equiv$ Prionace_Glaucicola</td>
</tr>
<tr>
<td>Different from</td>
<td>$x_1 \sqsubseteq \neg x_2$</td>
<td>Sea_Horse $\sqsubseteq \neg$ Horse</td>
</tr>
<tr>
<td>Sub property</td>
<td>$P_1 \sqsubseteq P_2$</td>
<td>has_mother $\sqsubseteq$ has_parent</td>
</tr>
<tr>
<td>Equivalent property</td>
<td>$P_1 \equiv P_2$</td>
<td>treated_by $\equiv$ cured_by</td>
</tr>
<tr>
<td>Inverse</td>
<td>$P_1 \equiv P_2^{-}$</td>
<td>location_of $\equiv$ has_location^-</td>
</tr>
<tr>
<td>Transitive property</td>
<td>$P^+ \sqsubseteq P$</td>
<td>part_of$^+$ $\sqsubseteq$ part_of</td>
</tr>
<tr>
<td>Functional property</td>
<td>$T \sqsubseteq 1P$</td>
<td>$T \sqsubseteq 1has_tributary$</td>
</tr>
<tr>
<td>Inverse functional property</td>
<td>$T \sqsubseteq 1P^{-}$</td>
<td>$T \sqsubseteq 1has_scientific_name^-$</td>
</tr>
</tbody>
</table>

DL = Description Logic

Concept equivalence
Speak: C1 is equivalent to C2

Concept inclusion,
Speak: All C1 are C2

---

How do you pronounce all these math expressions?

web.efzg.hr/dok/MAT/vkojic/Larrys_speakeasy.pdf


What are ontological class constructors?

<table>
<thead>
<tr>
<th>Constructor</th>
<th>DL syntax</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intersection</td>
<td>$C_1 \cap \ldots \cap C_n$</td>
<td>Anatomical_Abnormality $\cap$ Pathological_Function</td>
</tr>
<tr>
<td>Union</td>
<td>$C_1 \cup \ldots \cup C_n$</td>
<td>Body_Substance $\cup$ Organic_Chemical</td>
</tr>
<tr>
<td>Complement</td>
<td>$\neg C$</td>
<td>$\neg$Invertebrate</td>
</tr>
<tr>
<td>One of</td>
<td>$x_1 \cup \ldots \cup x_n$</td>
<td>Oestrogen $\cup$ Progesterone</td>
</tr>
<tr>
<td>All values from</td>
<td>$\forall P.C$</td>
<td>$\forall$co_occurs_with.Plant</td>
</tr>
<tr>
<td>Some values</td>
<td>$\exists P.C$</td>
<td>$\exists$co_occurs_with.Animal</td>
</tr>
<tr>
<td>Max cardinality</td>
<td>$\leq nP$</td>
<td>1 has_ingredient</td>
</tr>
<tr>
<td>Min cardinality</td>
<td>$\geq nP$</td>
<td>$\geq 2$ has_ingredient</td>
</tr>
</tbody>
</table>

Bhatt et al. (2009)

Intersection/conjunction of concepts, 
Speak: $C_1$ and ... $C_n$

Universal Restriction
Speak: All $P$-successors are in $C$

Existential Restriction
Speak: An $P$-successor exists in $C$
How can we combine statistical with logic-symbolic approaches?

What is a knowledge graph?

Huge amounts of open, unstructured data on the Web (and structured data in the “Deep Web”)

Heterogeneous data – Real-world entities associated with a wide variety of information
- You – as an individual
- A disease and all of its relationships
- A geographic location, e.g., your home
- Ethical information for a region
- ...

The need to search all of this data and “integrate” data (e.g., Google/Bing search)

Knowledge Representation and Querying Systems
- OWL, SPARQL

Computational power
- Big data – BigQuery, BigTable

Graph databases

https://www.youtube.com/watch?v=fyAHrwHjSck

https://web.stanford.edu/class/cs520
Albert Einstein was a German-born theoretical physicist who developed the theory of relativity.

https://web.stanford.edu/class/cs520
What is a typical example from computer vision?
Medical Classifications
What is classification generally?
Since the classification by Carl von Linne (1735) approx. 100+ various classifications in use:

- **International Classification of Diseases** (ICD)
- **Systematized Nomenclature of Medicine** (SNOMED)
- **Medical Subject Headings** (MeSH)
- **Foundational Model of Anatomy** (FMA)
- **Gene Ontology** (GO)
- **Unified Medical Language System** (UMLS)
- **Logical Observation Identifiers Names & Codes** (LOINC)
- **National Cancer Institute Thesaurus** (NCI Thesaurus)
What is the International Classification of Diseases (ICD)?

ICD-10 was endorsed by the Forty-third World Health Assembly in May 1990 and came into use in WHO Member States as from 1994. The classification is the latest in a series which has its origins in the 1850s. The first edition, known as the International List of Causes of Death, was adopted by the International Statistical Institute in 1893. WHO took over the responsibility for the ICD at its creation in 1948 when the Sixth Revision, which included causes of morbidity for the first time, was published. The World Health Assembly adopted in 1967 the WHO Nomenclature Regulations that stipulate use of ICD in its most current revision for mortality and morbidity statistics by all Member States.

http://www.who.int/classifications/icd/en
How did the International Classification of Diseases evolve?

- 1629 London Bills of Mortality
- 1855 William Farr (London, one founder of medical statistics): List of causes of death, list of diseases
- 1893 von Jacques Bertillot: List of causes of death
- 1900 International Statistical Institute (ISI) accepts Bertillot’s list
- 1938 5th Edition
- 1948 WHO
- 1965 ICD-8
- 1989 ICD-10
- 2015 ICD-11 due
- 2018 ICD-11 adopted
What is SNOMED?

- 1965 SNOP, 1974 SNOMED, 1979 SNOMED II
- 1997 (Logical Observation Identifiers Names and Codes (LOINC) integrated into SNOMED
- 2000 SNOMED RT, 2002 SNOMED CT

INTERNATIONAL HEALTH TERMINOLOGY
STANDARDS DEVELOPMENT ORGANISATION

239 pages

SNOMED CT® Technical Reference Guide
January 2011 International Release
(US English)

How does Hypertension look in SNOMED?

A

24184005|Finding of increased blood pressure (finding) →
38936003|Abnormal blood pressure (finding) AND
roleGroup SOME
  (363714003|Interprets (attribute) SOME
  75367002|Blood pressure (observable entity))

B

12763006|Finding of decreased blood pressure (finding) →
392570002|Blood pressure finding (finding) AND
roleGroup SOME
  (363714003|Interprets (attribute) SOME
  75367002|Blood pressure (observable entity))

What is MeSH?

- MeSH thesaurus is produced by the National Library of Medicine (NLM) since 1960.
- Used for cataloging documents and related media and as an index to search these documents in a database and is part of the metathesaurus of the Unified Medical Language System (UMLS).
- This thesaurus originates from keyword lists of the Index Medicus (today Medline);
- MeSH thesaurus is polyhierarchic, i.e. every concept can occur multiple times. It consists of the three parts:
  - 1. MeSH Tree Structures,
  - 2. MeSH Annotated Alphabetic List and
  - 3. Permutated MeSH.
What are the 16 trees in MeSH?

1. Anatomy [A]
2. Organisms [B]
3. Diseases [C]
4. Chemicals and Drugs [D]
5. Analytical, Diagnostic and Therapeutic Techniques and Equipment [E]
6. Psychiatry and Psychology [F]
7. Biological Sciences [G]
8. Natural Sciences [H]
10. Technology, Industry, Agriculture [J]
11. Humanities [K]
12. Information Science [L]
13. Named Groups [M]
14. Health Care [N]
15. Publication Characteristics [V]
16. Geographicals [Z]
How does the MeSH hierarchy look?

How does the MeSH Example Hypertension look?

National Library of Medicine - Medical Subject Headings

2011 MeSH

MeSH Descriptor Data

Return to Entry Page

Standard View. Go to Concept View; Go to Expanded Concept View

<table>
<thead>
<tr>
<th>MeSH Heading</th>
<th>Hypertension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree Number</td>
<td>C14.907.489</td>
</tr>
<tr>
<td>Annotation</td>
<td>not for intracranial or intraocular pressure; relation to BLOOD PRESSURE; Manual 23.27; Goldblatt kidney is HYPERTENSION, GOLDBLATT see HYPERTENSION, RENOVASCULAR; hypertension with kidney disease is probably HYPERTENSION, RENAL, not HYPERTENSION; venous hypertension: index under VENOUS PRESSURE (IM) &amp; do not coordinate with HYPERTENSION; PREHYPERTENSION is also available</td>
</tr>
<tr>
<td>Scope Note</td>
<td>Persistently high systemic arterial BLOOD PRESSURE. Based on multiple readings (BLOOD PRESSURE DETERMINATION), hypertension is currently defined as when SYSTOLIC PRESSURE is consistently greater than 140 mm Hg or when DIASTOLIC PRESSURE is consistently 90 mm Hg or more.</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Blood Pressure, High</td>
</tr>
<tr>
<td>See Also</td>
<td>Antihypertensive Agents</td>
</tr>
<tr>
<td>See Also</td>
<td>Vascular Resistance</td>
</tr>
<tr>
<td>Allowable Qualifiers</td>
<td>BL, CF, CI, CL, CN, CO, DH, DI, DT, EC, EH, EM, EN, EP, ET, GE, HI, IM, ME, MI, MO, NU, PA, PC, PP, PS, PX, RA, RH, RI, RT, SU, TH, UR, US, VE, VI</td>
</tr>
<tr>
<td>Date of Entry</td>
<td>19990101</td>
</tr>
<tr>
<td>Unique ID</td>
<td>D006973</td>
</tr>
</tbody>
</table>

http://www.nlm.nih.gov/mesh/
How does Mesh look in an interactive tree visualization?

What is UMLS – Unified Medical Language System?
The UMLS integrates and distributes key terminology, classification and coding standards, and associated resources to promote creation of more effective and electronic health records. More information...

New Users
- UMLS Quick Start Guide
- Licensing Information
- Basics Tutorial
- More...

User Education
- Webcasts
- Quick Tours
- Presentations
- More...

UMLS Knowledge Sources
Documentation for:
- Metathesaurus
- Semantic Network
- SPECIALIST Lexicon and Lexical Tools
- More...

Implementation Resources
For advanced users:
- MetamorphoSys
- Database Query Diagrams
- Load Scripts
- More...

UMLS News and Announcements
SNOMED CT RDA Subset available for download...
Subscribe to the UMLS News RSS Feed.

Related Resources
- MetaMap®
- RxNorm
- SNOMED CT®
- SNOMED CT CORE Subset
What subdomains does UMLS Metathesaurus integrate?

How does an Example of proteins and diseases in the UMLS look?

Conclusion
Concluding remark

- Progress in machine learning is driven by the explosion in the availability of **big data** and **low-cost computation** ...
- We need top-quality data and/or robust models to deal with the non-iid character of real-world data
ULTRA-MODERN MEDICINE: EXAMPLES OF MACHINE LEARNING IN HEALTHCARE

July 4, 2019 • Updated: March 25, 2020

Written by Mike Thomas
Thank you!