Introdution to medical informatics. Data and information.

O.O.Bogomolets National Medical University

February 18, 2013

Outline

- Course overview
- Medical informatics. Why? What for?
- Fundamental role of informatics in modern science.
- Information. Data. Knowledge.

Course overview

- Information. Data. Knowledge
- Biostatistics
- Formal Logic. Expert and decision support systems in medicine
- Evidence-based medicine

Medical informatics – the application of informatics to the field of biology and medicine/healthcare.

Informatics, Computers science

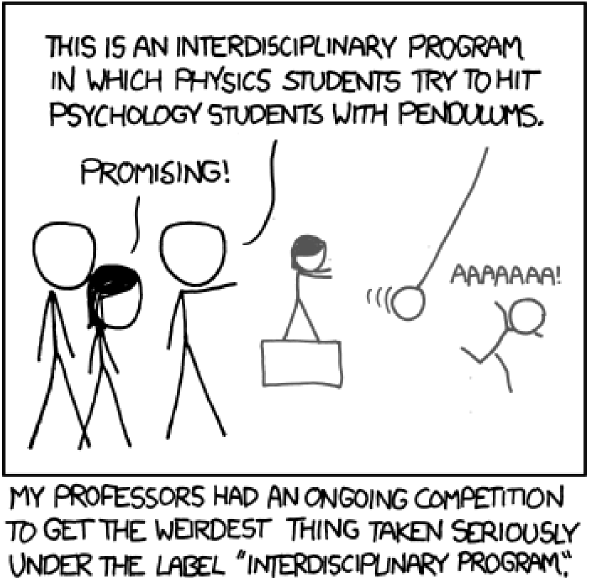
- **Informatics** – the study of the processing, management, and retrieval of information
- **Computers science** – is the study of the theoretical foundations of information and computation and of practical techniques for their implementation and application in computer systems
- **Cybernetics** is the interdisciplinary study of the structure of regulatory systems.

Medical informatics. Why? What for?

Medical informatics borrows concepts/ideas from:

- Medicine
- Informatics
- Logic
- Statistics

Medical informatics – interdisciplinary study



Fundamental role of informatics in modern science

- Any object in physical reality can be described with some information

Fundamental role of informatics in modern science

- Any object in physical reality can be described with some information
- Any physical process can be described as informational process
- Physical laws $\sim$ program/algorithm

Fundamental role of informatics in modern science

- Any object in physical reality can be described with some information
- Any physical process can be described as informational process
- Physical laws $\sim$ program/algorithm
- Universe is a gigantic computer!
  - Seth Lloyd, Programming the Universe
  - Roger Penrose, The Emperor's New Mind

Fundamental role of informatics in modern society

That's why informatics is an essential part of science disciplines in any University.
You will need basic informatics skills in modern society:

- We live in so called information society. (or at least tend to develop it)
- You may need to analyze systems/products
- You may need apply informatics skills in your research projects
- You may need to implement prototypes
- You may need to oversee programmers in your future job

Fundamental concepts of informatics

- Information
- Computation
- Data
- Knowledge

Information – primitive notion[1] of modern science.

So it is not defined formally in terms of previously defined concepts, but is only motivated informally, usually by an appeal to intuition and everyday experience

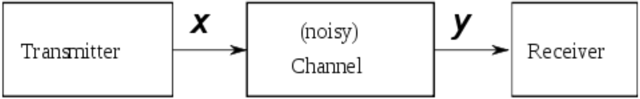
[1] fundamental undefined concept; primitive, undefined, term

Properties

Properties of information:

- **Objectivity** – information doesn't depend upon storage method or your opinion
- **Authenticity** – how trustworsy it is
- **Accuracy**
- **Actual**

Transmission of Information



All communicative situations include the presence of the sender information, its recipient and channel. Between any sender and a receiver there is always a transmission channel through which the message must pass.

Measure

Function $\mu : \Omega \to \mathbb{R}$ is called **measure**, if:

- $\forall A \in \Omega : \mu(A) \geq 0$
- $\forall A, B \in \Omega, A \cap B = \varnothing : \mu(A \cup B) = \mu(A) + \mu(B)$
- $A_n \searrow \varnothing : \mu(A_n) \longrightarrow 0$

Measure is mathematics's formalization/generalisation of such trivial common notion as **volume**.

Measures of Information. Shannon's measure

Well, we don't *exactly* know what the information is. But for our practical goals we can define *measures of information*, depending on our goals.

Shannon's measure[2]: $I(x) = \log(\frac{1}{P(x)}) = -\log(P(x))$ Here $x \in X$ – some message from the source X, $P(x)$ – probability of receiving message x.

Example:
On tossing a coin, the chance of 'tail' is 0.5. When it is proclaimed that indeed 'tail' occurred, this amounts to $I('tail') = \log_2(\frac{1}{0.5}) = \log_2 2 = 1$ bits of information.

[2] also called self-information

Shannon's measure. WTF?

Meaning behind the stages of all this math. Why Shannon has designed this measure to look so? His motives were simple:

1. $P(x) \searrow \Rightarrow I(x) \nearrow$ – more unexpected message contains more information.
2. For independent messages we can add their informational measures: $\forall x, y \in X; x \cap y = \varnothing : I(x \cup y) = I(x) + I(y)$
3. $I(x) = 0 \Rightarrow P(x) = 1$ – If we already know smt, then we don't receive any information at all.

Unfortunately Shannon's measure is not suitable for complex[3] messages.

[3] In practise "complex" means more then 1–2 symbols or smt like so

Entropy.

Entropy – amount of disorder in a system
In *information theory*, **entropy** is a measure of the uncertainty associated with a random variable(in our case received message); Entropy is an average amount of information in one message.

$$H(X) = \mathrm{M}(I(x)) = \sum_{x \in X} p(x) \log p(x)$$

Entropy. Example.

Message	Mark	Number of students
x_1	5	125
x_2	4	250
x_3	3	500
x_4	2	125

x_i	$P(x_i)$	$I(x_i)$
x_1	$\frac{1}{8}$	3
x_2	$\frac{1}{4}$	2
x_3	$\frac{1}{2}$	1
x_4	$\frac{1}{8}$	3

$$H(X) = \frac{1}{8} \cdot 3 + \frac{1}{4} \cdot 2 + \frac{1}{2} \cdot 1 + \frac{1}{8} \cdot 3$$

Break

Let's have a **break**!

Data

Data refers to qualitative or quantitative attributes of a variable or set of variables.

- For example – results of measurements: $(35.7, 36.7, 42, 38, 39.4, 37.5, 40.1)$
- Raw data is the lowest level of abstraction. Raw data on its own carries no meaning.

Data

Data refers to qualitative or quantitative attributes of a variable or set of variables.

- For example – results of measurements: $(35.7, 36.7, 42, 38, 39.4, 37.5, 40.1)$
- Raw data is the lowest level of abstraction. Raw data on its own carries no meaning.
- Raw data needs some way of interpretation.

Data

Data refers to qualitative or quantitative attributes of a variable or set of variables, stored in some way.

- For example – results of measurements: $(35.7, 36.7, 42, 38, 39.4, 37.5, 40.1)$
- Raw data is the lowest level of abstraction. Raw data on its own carries no meaning.
- Raw data must be interpreted and take on a meaning
- Example: $(35.7, 36.7, 42, 38, 39.4, 37.5, 40.1)$ – results of patients thermometrics

Knowledge

Knowledge is a familiarity with someone or something, that can include facts/data, descriptions, information, and/or skills acquired through experience or education.

Main difference between knowledge and data: knowledge has dynamic nature by itself, it can change via learning. This the highest level of information abstraction.

Computation

- **Computation** – any type of information processing.

Computation

- **Computation** – any type of information processing.
- A computation can be seen as a purely physical phenomenon occurring inside a closed physical system called a computer.

- **Computation** – any type of information processing.
- A computation can be seen as a purely physical phenomenon occurring inside a closed physical system called a computer.
- To describe a computation we need some formal way to state rules for information processing - **algorithm**.

Algorithm

Algorithm is an effective method expressed as a finite list of *well-defined*[4] instructions for computation.

Starting from an initial state and initial input (perhaps empty), the instructions describe a computation that, when executed, will proceed through a finite number of well-defined successive states,using finite amount of resources, eventually producing "output"

[4] Because "well-defined" is not strict, formal term – the whole definition, presented here, is **informal**. Though it is sufficient for basic intuitive understanding of the topic

Properties of Algorithms

- **Finiteness** – An algorithm must always terminate after a finite number of steps
- **Definiteness** – Each step of an algorithm must be precisely defined; the actions to be carried out must be rigorously and unambiguously specified for each case
- **Discrete** – moving in separate(discrete) steps: 1, 2, 3, ...
- **Input**
- **Output**

Computability and computational complexity

Some task is called **computable** if there exists an algorithm for it.[5] Not all tasks are computable. **Computational complexity** – informally,

amount of resources we'll need to complete the task using selected algorithm. This characteristics helps us do determinate the most optimal algorithm to solve our problems.

[5] Literally: it can be solved in general case.

Example – herbal drug foalfoot

1. 10 g of herbal drug (2 spoons) place to enamel ware,
2. Add 200 ml (1 cup) of boiled room-temperature water
3. close cover and infuse on the water bath for 15 min
4. Cool down for 45 min at the room temperature
5. strain
6. press out the rest to the strained infusion
7. The infusion adds up with boiled water to 200 ml
8. Take 0, 5 cup of warm infusion two times a time 15 min before meal

Flowcharts

Name	Drawing
Start and end symbols	
Input/Output	
Generic processing steps	
Conditional or decision	

Types of Algorithms and Their Flowcharts

1. Linear algorithm
2. Branch algorithm
3. Cycle algorithm

Dixi