

**MAI4CAREU**

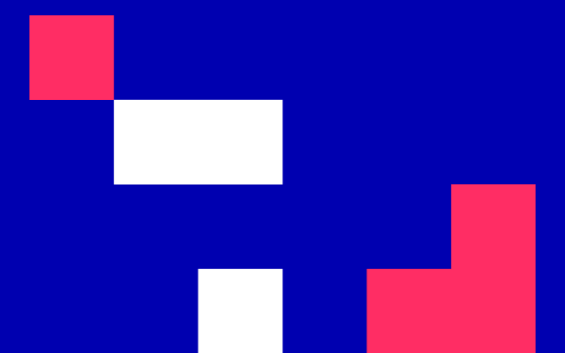
Master programmes in Artificial  
Intelligence 4 Careers in Europe

University of Ruse

# INTELLIGENT COMPUTER SYSTEMS

**Svetlana Stefanova**

September, 2022



**LECTURE 8****REASONING WITH UNCERTAINTY**

1. Introduction
2. Uncertainty of rules
3. Uncertainty of facts
4. Credibility factor
5. Ways to evaluate the certainty of rules

**CONTENT 1**

# Relevance of the knowledge security problem

The problem with reporting and analyzing the security of knowledge is one of the most important in the field of AI.

**Sources of uncertainty:**

- Uncertainty of facts;
- Uncertainty of rules.

**Part of the expertise of experts is to know:**

- when to ignore missing information;
- when to stop to get it.

# Uncertainty of rules

Rules that carry knowledge with uncertain information can be included in the KB.

Example: *IF you have pain in your back, THEN apply warming ointment.*

The pain may be neurological and heating may be contraindicated.

**CONTENT 3**

# Uncertainty of facts

In the real world, despite the uncertainty of the data, we can make completely certain conclusions by adhering to **certain rules**:

- to not apply given rules if the information needed to evaluate their premises is not available.
- the result depends on the type of prerequisites:
  - for **AND** clauses - all clauses must evaluate to true for the rule to apply. If a user answers "unknown" for any part of the prerequisite, the rule fails.
  - for **OR** clauses - unknown information related to one clause of the premise does not make the rule impossible to succeed.

**CONTENT 3**

# Uncertainty of facts

If we use rules in the data structure:

## IF A THEN G

it is important to the inference **G** that the fact **A is not known or not entirely certain**, as it requires the derivation of new propositions from assumed true premises, by repeated application of rules.

**CONTENT 3**

# Uncertainty of facts – where can it come from?

- Ambiguity;
- Incompleteness;
- Unreliability/inaccuracy.

**CONTENT 3**

# Uncertainty of facts - ambiguity

Ambiguity is one of the properties of information in the real world.

**Sources of ambiguity:**

- conflicting knowledge;

Example: *data from a video camera with limited resolution.*

- blurred boundaries of concepts;

Example: *He is an old man.*

- expressions with a multi-valued scale of truth, etc.

Example: *You must be good!*



**CONTENT 3**

# Uncertainty of facts - incompleteness

Incompleteness refers to the content of the information and can be seen as:

- **Absence;**
- **Insufficiency.**

## Sources of incompleteness:

- time - lack of time for a complete analysis;
- financial - need for expensive and risky research.

In reasoning using incomplete data, we get conclusions that are questionable.

Making a decision when facing incomplete data gives reason to thinking.

**CONTENT 3**

# Uncertainty of facts - unreliability/inaccuracy

An inaccuracy is established when the truth of the information is considered.

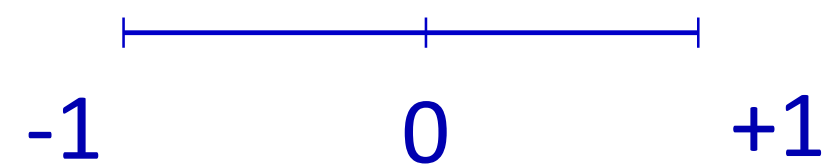
**Sources of inaccuracy:**

- inaccurate data;
- unreliable data.

**CONTENT 4**

# Credibility factor

The degree of certainty can be quantified by a number called the **credibility factor or safety factor**:



**Important:** Credibility factor is not a probability, but it can be compared to a probability without being expected to obey the laws of probability

**CONTENT 4**

# Credibility factor - application

According to the nature of the sources of uncertainty, the credibility factor can be used:

- **In case of uncertainty of the facts** - With the help of composite premises, the uncertain facts can be checked;
- **In case of uncertainty of the rules** - The rules themselves may not be completely defined, and it is also possible that some facts-inferences may be obtained as a result of more than 1 rule. A combining function combines their credibility factors.

**CONTENT 4**

# Credibility factor – of facts

A user sometimes thinks a fact is true, but is not completely sure.

Prerequisites are evaluated differently depending on the number of clauses and logical connections they contain. Rules succeed if the premises have a credibility factor greater than a certain **threshold**.

An uncertain premise always leads to an uncertain conclusion, and it is affected by its uncertainty.

**CONTENT 4**

# Credibility factor – of rules

When the rules are not completely certain they may also have an associated "credibility factor".

**Example:**

IF you have pain in your back, THEN apply warming ointment.

**CONTENT 4**

# Credibility factor – example

*“If Boko is green*

*then he is probably a frog”.*

But he can also be a chameleon?

This type of reasoning can be mimicked using numeric values:

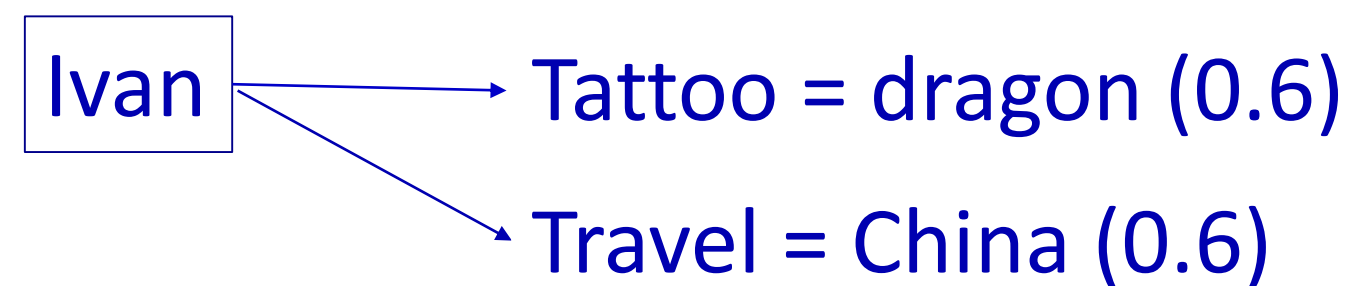
“If it is known that Boko is green, it can be inferred with certainty 0.85 that he is a frog”

“If it is known that Boko is a frog, it can be inferred with certainty 0.95 that he jumps”.



**CONTENT 4**

# Credibility factor – example



*„If Ivan's tattoo is a dragon, then Ivan traveled to China“.*

**Credibility:** We can only be partially (0.6) confident that a given tattoo is of one type.

This credibility factor is propagated by the rule that concludes that Ivan has been in China (0.6).



**CONTENT 4**

# Credibility factor of the inference

- Incompletely certain information will never be combined to form a certain conclusion.
- As more positive information emerges, confidence in the conclusion increases.
- The order in which the information is combined does not matter for credibility.

**CONTENT 4**

# Credibility factor of the inference

- **If the premise is simple** and succeeds with some certainty, then the value is accepted with its given credibility factor;
- **If the premise is complex** and succeeds with some certainty, then the final credibility factor of the conclusion is the product of the credibility factors of the premise.

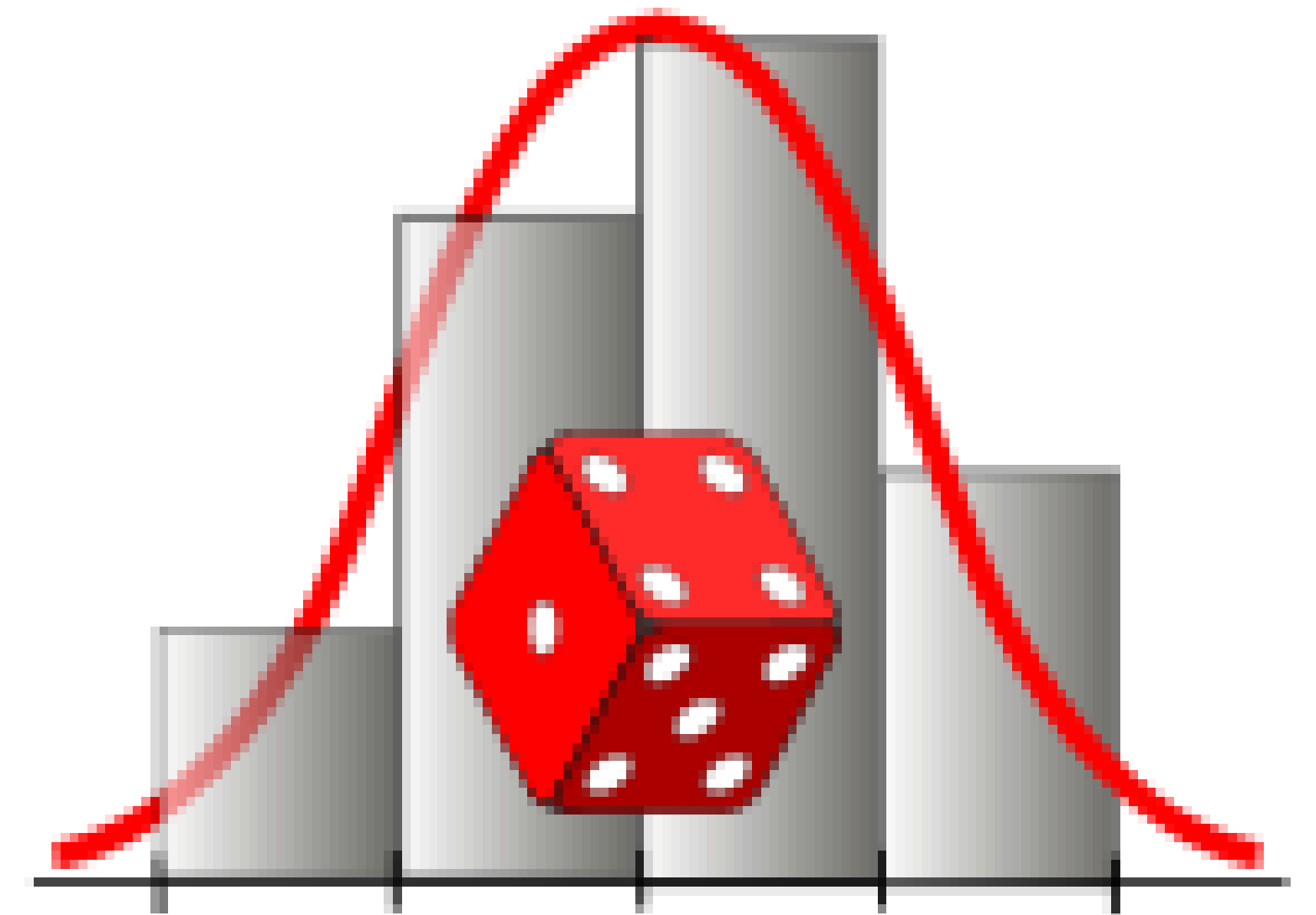
**CONTENT 5**

# Ways to evaluate the certainty of rules

- probability theory (Bayes's theory);
- evidence theory (Dempster-Shafer theory);
- possibility theory (Zadeh's theory);

## CONTENT 5

# Probability theory



**CONTENT 5**

# Probability theory

- **Probability** – it is established experimentally or statistically as a way to determine the subjective confidence in a given event.
- **Unconditional probability** - the confidence that is assigned to a statement A without requiring the fulfillment of any condition.

**P(A)**, where P(A) is a non-negative number

- **Conditional probability** – it is used when, in order to establish the probability of a certain statement A, the realization of a certain event B is taken into account. It is used only when we know everything about the event B.

**P(A|B)** (i.e. IF B THEN A)

When event C becomes known (additional information), the conditional probability is calculated: **P (A | B∧C)**.

**CONTENT 5**

# Probability theory – conditional and unconditional probability

Unconditional probability can be seen as a special case of conditional probability, where the absence of facts about the event **B** is indicated by its absence.

$$P(A)$$

Conditional probabilities can be determined from unconditional ones by the probability product theorem

$$P(A|B) = P(A \cap B) / P(B), \text{ where } P(B) > 0$$

i.e. written back

$$P(A \cap B) = P(A|B) * P(B)$$

**CONTENT 5**

# Probability theory - axioms

- all probabilities are with a value **between 0 and 1**;
- true statements have a probability of **1** and false statements have a probability of **0**;
- the disjunction probability is:

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

**CONTENT 5**

# Probability theory - features

The following property can be deduced from the axioms:

$$P(A|B) * P(B) = P(B|A) * P(A)$$

If we divide by  $P(A)$  we get **Bayes's rule**:

$$P(B|A) = P(A|B) * P(B) / P(A),$$

which allows for an unknown probability to be calculated with given known others.



**CONTENT 5**

# Probability theory - example

**Task:** making a medical diagnosis (where there are many conditional probabilities). What is the probability that a patient has meningitis if he has a stiff neck?

A is the statement that the patient has a stiff neck and B is the probability that he has meningitis.  $P(B|A)$  means that the patient with a stiff neck may have meningitis.

$$P(B|A) = (1/2 * 1/50000) / 1/20 = 1/5000.$$

The doctor knows that meningitis is a diagnosis in 50% of cases when the complaint is a stiff neck (the conditional probability). According to the statistics (unconditionally):

- the probability of someone suffering from meningitis is 1/50 000;
- the probability of a person having a stiff neck is 1/20.

**CONTENT 5**

# Probability theory – credibility measurement

In Bayes's theory, probability is often used to measure the credibility of a statement.

Laplace (1812r.) reaches the same conclusion and used it to solve problems in celestial mechanics, medicine, and court practice.

**Example:** Laplace calculated the mass of Saturn based on existing astronomical observations of its orbit. He announces the results, along with their uncertainty: "I bet 11'000:1, that the error in this result is not more than 1/100 of its value."

Laplace would win the bet, because 150 years later, based on new data, his result has been corrected by only 0,63%.

**CONTENT 5**

# Probability theory – evaluating hypotheses

Bayes's rule is also used for hypothesis evaluations.

When new information comes in that imposes changes on the probabilities, the changes are propagated to the nodes and the new situation is established, i.e. it reflects the fact that statements influence each other.

**Bayesian nets** - they represent dependencies between individual quantities and specify the probability distribution. A confidence network is a graph including:

- **quantities**, forming the nodes in the network;
- **direct connections**, connecting pairs of nodes, where the intuitive meaning of arrows is to indicate the direct influence of one quantity on another.
- each node has a **table of probabilities** quantifying the effect on it that the so-called causal quantities have (those from which the arrows start).

**CONTENT 5**

# Probability theory – uncertain inference

There are two reasons for the uncertainty of statement **B**:

- the user has said, that **B** is uncertain;
- the program derives **B** by a plausible rule, i.e. with probability.

If we want to use **B** in a rule for  $P(A|B)$  the question arises: how do we reduce the certainty of the conclusion due to the uncertainty associated with **B**?

**CONTENT 5**

# Probability theory – uncertain inference

Let **E** denote the evidence used to establish **B**, and  $P(A|E)$  be the current probability of **A** with given **E**.

Under some assumptions:

$$P(A|E) = P(A|B) \cdot P(B|E) + P(A|\sim B) \cdot [1 - P(B|E)]$$

This formula is valid in the extreme cases of complete certainty, i.e.

if we know that **B** is true ( $P(B|E)=1$ ), we get  $P(A|B)$ , and

if we know that **B** is false ( $P(B|E)=0$ ), we get  $P(A|\sim B)$ .

**CONTENT 5**

# Probability theory – uncertain inference

Unfortunately, a problem arises in the intermediate cases.

In particular, let's suppose that E does not provide information about B, i.e.  
 **$P(B|E) = P(B)$** .

The previous formula promises to add  **$P(A)$** , when the calculation is based on expert-derived values, but the resulting value for  **$P(A|E)$**  usually does not agree with the expert estimate of  **$P(A)$** . This means that  **$P(A)$** ,  **$P(B)$** ,  **$P(A|B)$**  and  **$P(A|\sim B)$**  are not independent, and the expert's subjective estimates of them are almost certainly numerically untenable.

**CONTENT 5**

# Probability theory – uncertain inference

In this particular case, the problem can be solved by, for example, not asking the expert about  $P(A)$ , but calculating it

$$P(A) = P(A|B)*P(B) + P(A|\sim B)*P(\sim B).$$

However, this makes the parameters of only one rule compatible, and the solution is not obvious when given a network of rules with conflicting values of the probability parameters.

**CONTENT 5**

# Probability theory - assessing the uncertainty of the rules

- **the reliability coefficient of a conjunction** is the minimum of the coefficients of the elements included in the conjunction  
e.g.  $(P1(\text{with } K1) \wedge P2(\text{with } K2)) \text{ with } K3 \rightarrow K3 = \min(K1, K2)$
- **the reliability coefficient of a disjunction** is the maximum of the coefficients of the elements included in the disjunction;
- **the confidence coefficient of the negation** is the difference between a one and the confidence coefficient for the statement;
- **the certainty of an entire rule** is considered as the product of the coefficients of the premises with the certainty of the consequence, i.e. (If A with K1, then C with K2) with K3  $\rightarrow K3 = K1 * K2$



**CONTENT 5**

# Probability theory - conclusions

- Using these rules, it is possible to combine partial information based on mathematical approaches.
- Typically, a rule has a security threshold value related to the security of the premise. Below this threshold value, the rule is not activated.

**Bayes' theory does not make it possible to distinguish uncertainty from ignorance.**

**CONTENT 5**

# Probability theory - problems in quantifying uncertainty

- **the need to give numerical values for a priori probabilities** - The expert can say how certain he feels about conclusion A when condition B is present (If B Then A), but may find it difficult to assign the probability of A in the absence of any given condition.
- **the expert cannot distinguish probability from utility or importance** and expresses some undetermined measure of importance.

## CONTENT 5

# Evidence theory

### I. TOOLS

#### 1. NATURAL DEDUCTION

Proofs for Conditionals  
Normal Proofs  
Strong Normalisation & Terms

#### 2. SEQUENT CALCULUS

Derivations for  $\wedge/\vee$  ( $A \rightarrow B$ )  
Eliminating Id & Cut  
 $X \rightarrow A, X \rightarrow Y$  Sequents for  $\wedge, \vee, @, @, \rightarrow, \neg, \text{f}, \text{T}, \perp$   
Consequences of Cut Elimination

#### 3. FROM PROOFS TO MODELS

Positions & Valuations  
Soundness & Completeness  
Cut Admissibility  
The Significance of Valuations

### II. THE CORE ARGUMENT

#### 4. TONK

Prior's Challenge  
What Could Count as a Response?  
Answering with Model Theory  
Conservative Extension  
Uniqueness  
Harmony

#### 5. POSITIONS

Language  
Assertion & its Norms  
Assertion, Denial & Other Speech Acts  
Positions & Structural Rules  
Bounds, Cut & Inference  
Challenges

#### 6. DEFINING RULES

Defining a Biconditional  
Defining Rules Defined  
Defining Rules & QR Rules  
Eliminating Cut  
Answering Prior's Question

### III. INSIGHTS

#### 7. MEANING & PROOF

Connectives  
Necessity  
Proof & Meaning  
Warrant

#### 8. QUANTIFIERS & OBJECTS

Generality  
Identity  
Defining Rules for Quantifiers  
Positions & Models  
Arithmetic, Realism & Anti-Realism

#### 9. MODALITY & WORLDS

Hypersequents  
Solving Prior's other Problem  
Positions & Worlds  
Quantifiers & Identity  
Two Dimensions



**CONTENT 5**

# Comparison between the two theories

**Both:**

- account for degrees in measuring uncertainty using probabilities.
- set a confidence function on a set of hypotheses;
- set a mechanism for updating the current set of probabilities when new information arrives.

**Evidence theory can distinguish ignorance from uncertainty.**

**CONTENT 5**

# Evidence theory - advantage

With Bayes, the degree of confidence in hypothesis H can be defined by the probability of its negation:

$$P(H) = 1 - P(\sim H)$$

Here, the degree of confidence in the hypothesis H together with the degree of confidence in the complement does not always give full certainty of 1:

$$P(H) + P(\sim H) \leq 1.$$

The difference to 1 defines the **degree of ignorance**.

Dempster and Shaffer insist on a fundamental distinction between uncertainty and ignorance.

In probability theory, one expresses the degree of one's knowledge or belief in the statement A with a number P(A).

Dempster and Shaffer believe that the classic Bayesian failure of a prior probabilities is often due to the fact that one does not know the values of the prior probabilities, and this makes every choice arbitrary and unjustified.

**CONTENT 5**

# Evidence theory – confidence functions

Evidence theory addresses the distinction between uncertainty and ignorance by introducing **confidence functions**.

They satisfy axioms that are weaker than those for probability functions. Thus, probability functions become a subclass of confidence functions, and proof theory reduces to probability theory when the probability values are known.

**The Dempster-Shafer theory does not make it possible to distinguish uncertainty from lack of specificity.**

**CONTENT 5**

# Possibility theory



**CONTENT 5**

# Possibility theory

In 1973 professor Zadeh created a theory of possibilities/fuzzy logic, which is **not actually fuzzy**, but to a large extent precise.

**Applications:**

- creation of algorithms for recognition of images, pictures and sounds;
- signal processing;
- quantitative analysis in economics - study of financial operations, etc.;
- decision-making systems - expert systems for diagnostics, planning, forecasting, etc.;
- information processing/databases.

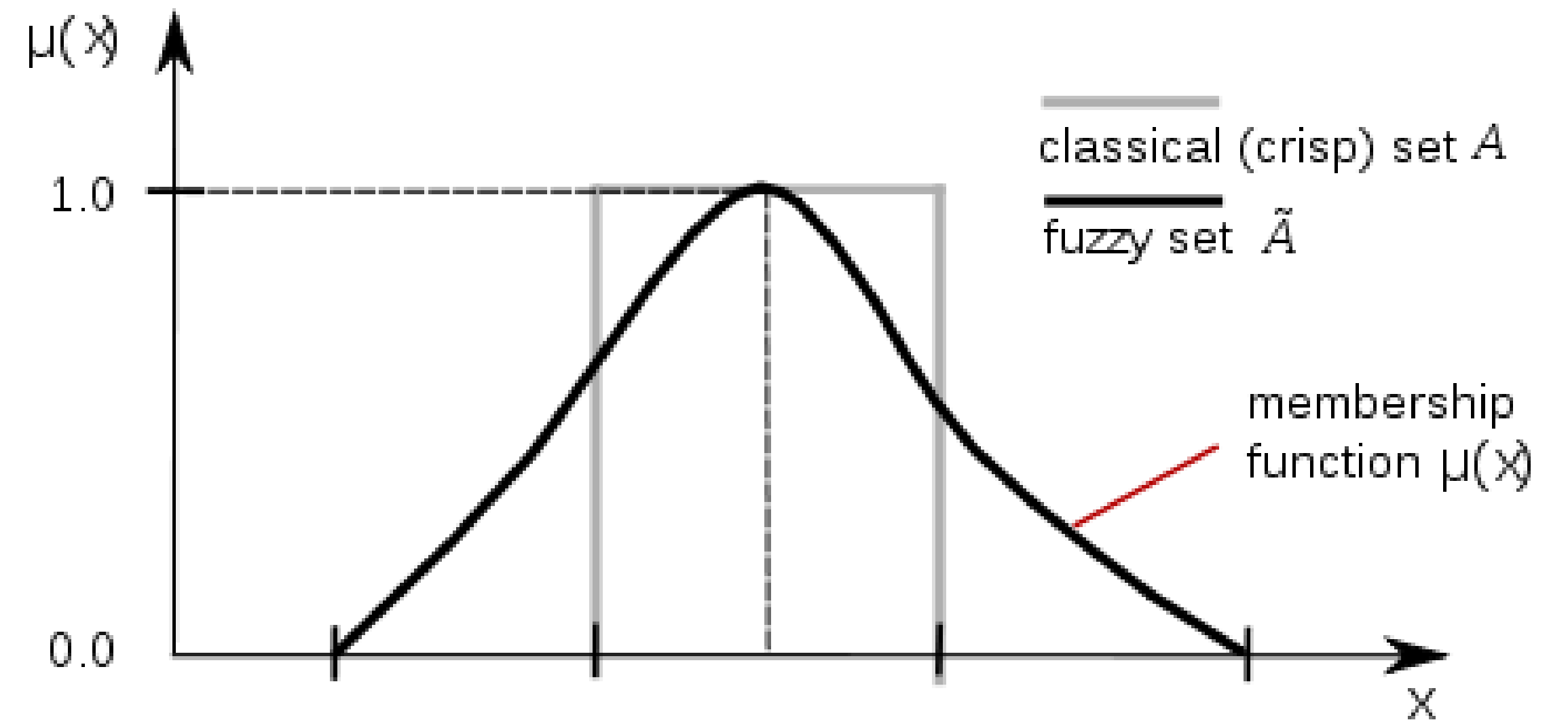


**CONTENT 5**

# Possibility theory – fuzzy sets

Fuzzy sets are classes with imprecise boundaries.

**Example:** a class of beautiful women, a class of honest men, and a class of tall mountains.



**CONTENT 5**

# Possibility theory – fuzzy sets

According to Zadeh, probability theory is suitable for tasks related to measuring information, but it is inappropriate for **tasks related to the meaning of the information**.

Much of the uncertainty surrounding the use of English terms and expressions is due to vagueness rather than coincidence.

Possibility theory quantifies this kind of uncertainty by introducing **membership functions** with values between 0 and 1.

**CONTENT 5**

# Possibility theory – fuzzy sets and membership degree

- **In classical set theory** - a given element either belongs or does not belong to a given set, i.e. its membership is evaluated as 1 or 0 and there is no third possibility (Law of the Excluded Middle).
- **In fuzzy set theory**, an element belongs to a given set with a degree calculated according to a specified membership function as a number in the interval  $[0, 1]$ . The membership function is an extension of Cantor's characteristic function and expresses the degree of truth of a statement. In classical logic, there are only two degrees of truth: "false" and "true";

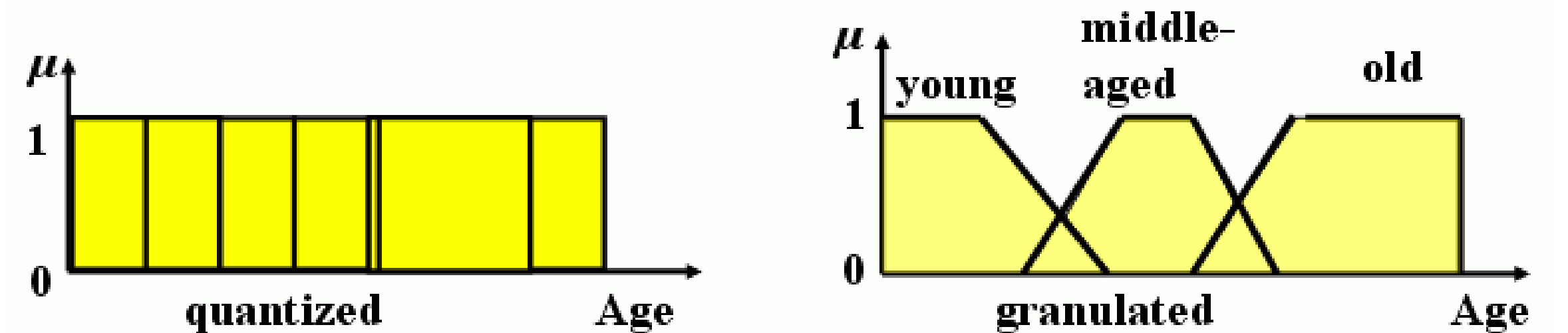
## CONTENT 5

# Possibility theory - overlapping

In fuzzy logic, everything is or is allowed to be overlapping.

**Example:** the term age is an overlapped term when its values are described as young, middle-aged, and old.

• continuous  $\longrightarrow$  quantized  $\longrightarrow$  granulated



Young, middle-aged and old are fuzzy sets

**CONTENT 5**

# Possibility theory – fuzzy sets

If  $U$  is a set and  $u$  is an element of  $U$ , a fuzzy subset  $A$  of  $U$  is defined by a membership function  $m_A(u)$ , which measures the degree to which  $A$  belongs to  $U$ .

Mathematically,  $A$  is defined as the set of ordered pairs:

$$A = \{u, m_A(u)\}$$

$m_A(u)$  maps to each element of  $U$  a number from the interval  $[0, 1]$ .

**Example:**

if  $S$  is the set of positive integers,

$F$  is the fuzzy subset of small integers,

we can have  $m_F(1) = 1$ ,  $m_F(2) = 1$ ,  $m_F(3) = 0,8$ ... $m_F(20) = 0,01$  etc.

**CONTENT 5**

# Possibility theory – representation of fuzzy sets

- **by enumeration** - if  $U$  consists of a finite number of elements, then  $A$  is presented:

$$A(u_j) = m_1/u_1 + \dots + m_n/u_n = \sum m_j/u_j, \text{ където } j = 1 \text{ до } n.$$

- **graphically;**
- **analytically.**

Composite membership functions are obtained by operations on simple membership functions, for example addition or subtraction.

**CONTENT 5**

# Possibility theory and fuzzy sets

**Probability theory** is concerned not so much with how the numerical values of probabilities are determined, but with the rules for calculating the probability of expressions containing random variables.

Similarly, **possibility theory** is not concerned with obtaining numerical values for probability distributions, but with rules for computing the probability of expressions containing fuzzy variables.

**Conclusion:** Most of the concepts in probability theory have counterparts in possibility theory. Therefore, possibility theory can also be used to quantify the uncertainty coming from the ambiguity of input information — no matter whether the input is data or rules.

**CONTENT 5**

# Possibility theory – linguistic approach

To account for the factors affecting a given problem, which have qualitative characteristics and are expressed by insufficiently defined and ambiguous terms from natural language and professional terminology in a given subject area, special formalization methods have to be applied.

**Linguistic approach** uses the theory of fuzzy sets and allows to present subjective linguistic evaluations of specialists and thus account for the inaccuracy of information contained in natural language.



**CONTENT 5**

# Possibility theory – linguistic variable

**Determined by:**

- name;
- multitude of meanings;
- field of meanings;
- a syntactic procedure for forming new meanings of a linguistic variable;
- a semantic procedure for rewriting a formed new meaning to some semantics by forming a corresponding fuzzy set.

**Example:** linguistic variable HEIGHT, formalizing the concept of *human height* has meanings SHORT, MEDIUM, TALL and a range of meanings from 1m to 2 meters. Formally, this can be written:  $\langle \text{height}\{\text{short, medium, tall}\}, [1, 2], \text{Proc1}, \text{Proc2} \rangle$ .

**CONTENT 5**

# Possibility theory – linguistic variable

The concept of **degree of affiliation** serves to denote the degree of correspondence of a given value of a linguistic variable to the subjective meaning attached to it.

**Example:**

set **A** corresponds to the fuzzy concept "*small stock in the warehouse*",

exponent of **A** is a finite set of values  $S\{10,11,12,\dots,40\}$ , where the elements from 10 to 40 are separate quantities of materials.

If we ask a specialist to express with a number how true the statement "a small stock of m-l in storage" is for each of the elements of **S**, then the set can be represented:

$$A=\{0,05/10;0.1/11;0.2/12;\dots;0.1/40\}.$$

According to the formalized concepts of a specialist, some of the following meanings correspond to the most complete degree of "small stock in the warehouse":

max 20-23; 14-19; 30-40 min.

Each of the meanings of linguistic variables can be graphically expressed by means of a membership function.

# MAI4CAREU

Master programmes in Artificial  
Intelligence 4 Careers in Europe



Co-financed by the European Union  
Connecting Europe Facility

This Master is run under the context of Action  
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom  
under GA nr. INEA/CEF/ICT/A2020/2267423

