Managing Uncertainty and Vagueness in Description Logics for the Semantic Web

Scuola di Dottorato in Ingegneria dell'Informazione Napoli 2009

Thomas Lukasiewicz

Computing Laboratory, University of Oxford
thomas.lukasiewicz@comlab.ox.ac.uk

Outline

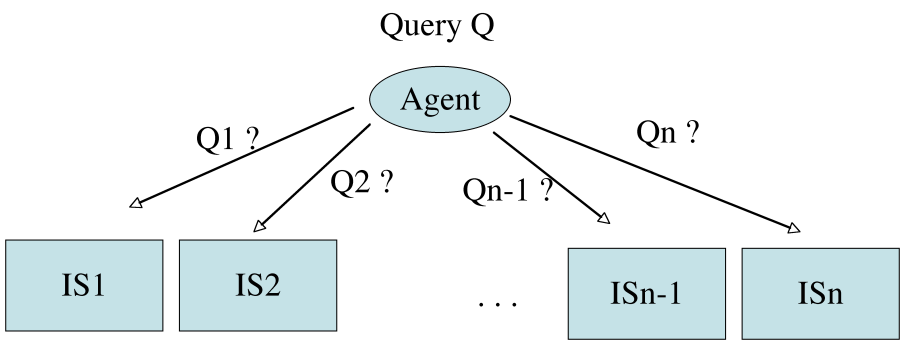
Outline

Sources of Uncertainty and Vagueness on the Web

- (Multimedia) information retrieval:
  - To which degree is a Web site/page, a text passage, an image region, a video segment, ... relevant to my information need?
- Matchmaking
  - To which degree does an object match my requirements?
    - if I'm looking for a car, and my budget is *about* 20.000 €, to which degree does a car's price of 20.500 € match my budget?
- Semantic annotation
  - To which degree does, e.g., an image object represent a dog?

- Information extraction
  - To which degree am I sure that, e.g., SW is an acronym of "Semantic Web"?
- Ontology alignment (schema mapping)
  - To which degree do two concepts of two ontologies represent the same, or are disjoint, or are overlapping?
- Representation of background knowledge
  - To some degree, birds fly.
  - To some degree, Jim is blond and young.

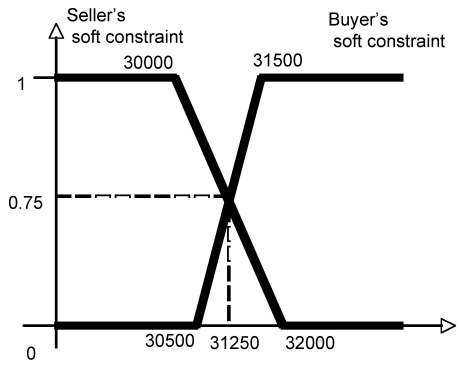
Distributed Information Retrieval



Then, the agent has to perform automatically the following steps:

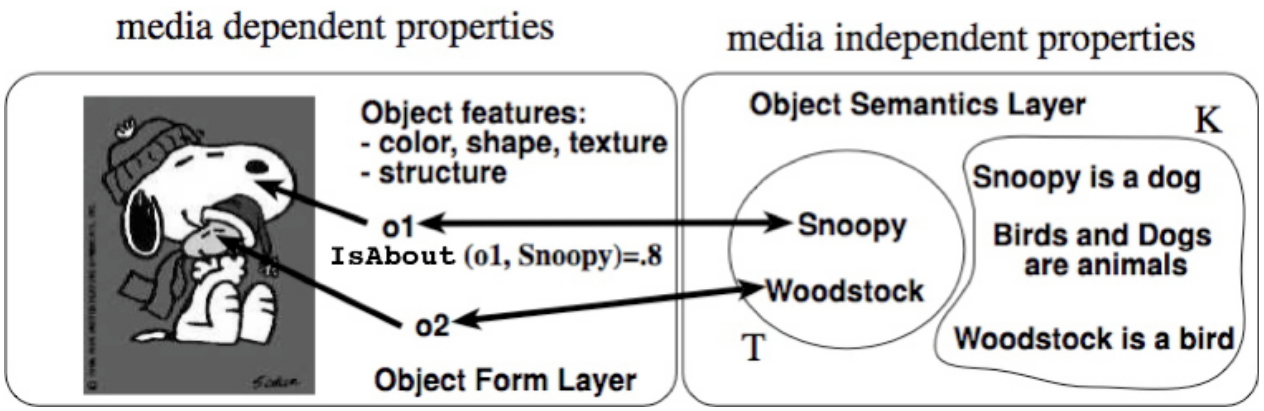
1. The agent has to select a subset of relevant resources $\mathscr{S}' \subseteq \mathscr{S}$, as it is not reasonable to assume to access to and query all resources (resource selection/discovery);
2. For every selected source $\mathcal{S}_i \in \mathscr{S}'$, the agent has to reformulate its information need Q_A into the query language $\mathcal{L}_i$ of the resource (schema mapping/ontology alignment);
3. The results from the selected resources have to be merged (data fusion/rank aggregation).

Negotiation



- A car seller sells an Audi TT for 31500 €, as from the catalog price.
- A buyer is looking for a sports car, but wants to pay not more than around 30000 €
- Classical DLs: the problem relies on the crisp conditions on price.
- More fine-grained approach: to consider prices as vague constraints (fuzzy sets) (as usual in negotiation)
  - Seller would sell above 31500 €, but can go down to 30500 €.
  - The buyer prefers to spend less than 30000 €, but can go up to 32000 €.
  - Highest degree of matching is 0.75. The car may be sold at 31250 €.

Logic-Based Information Retrieval



IsAbout		
ImageRegion	*Object ID*	*degree*
o1	*snoopy*	0.8
o2	*woodstock*	0.7
⋮	⋮	

"Find top-k image regions about animals"

$$Query(x) \leftarrow ImageRegion(x) \wedge isAbout(x, y) \wedge Animal(y)$$

Top-k Database Querying

HotelID	*hasLoc*
$h1$	$hl1$
$h2$	$hl2$
⋮	⋮

ConferenceID	*hasLoc*
$c1$	$cl1$
$c2$	$cl2$
⋮	⋮

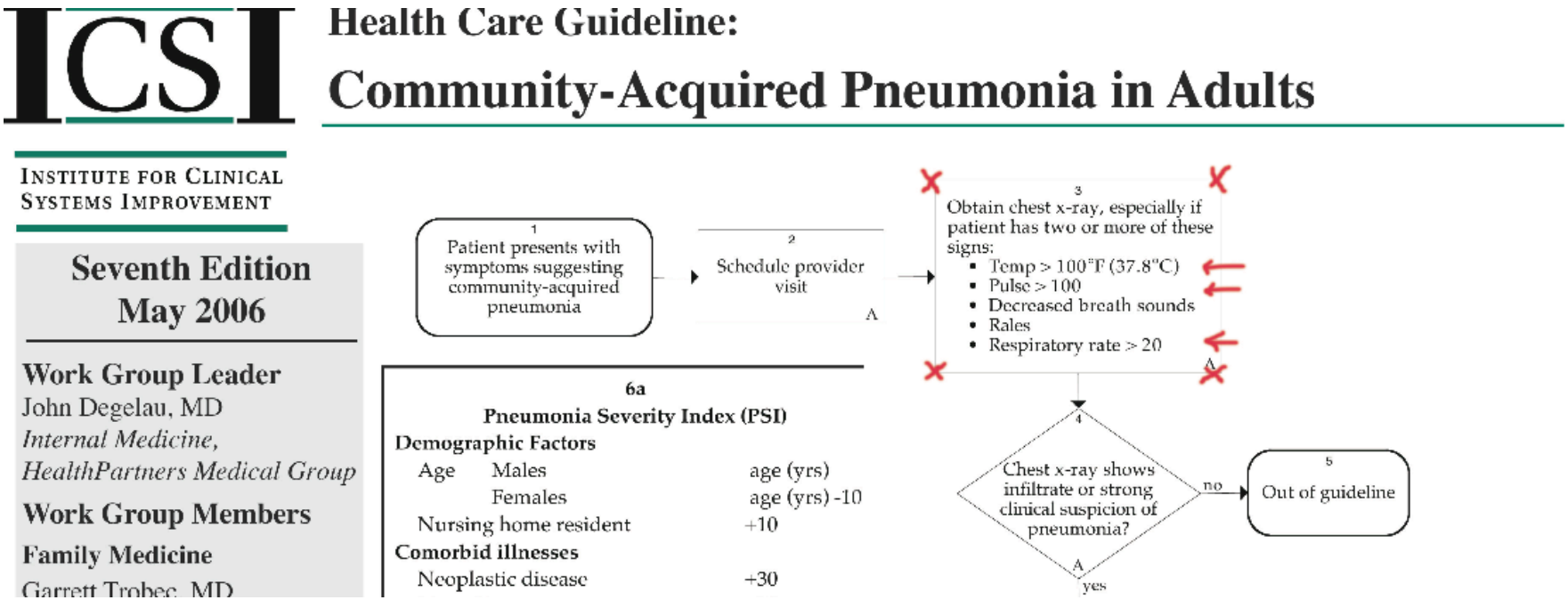
hasLoc	*hasLoc*	*distance*
$hl1$	$cl1$	300
$hl1$	$cl2$	500
$hl2$	$cl1$	750
$hl2$	$cl2$	800
⋮	⋮	

hasLoc	*hasLoc*	*close*	*cheap*
$hl1$	$cl1$	0.7	0.3
$hl1$	$cl2$	0.5	0.5
$hl2$	$cl1$	0.25	0.8
$hl2$	$cl2$	0.2	0.9
⋮	⋮	⋮	

"Find top-k cheapest hotels close to the train station"

$$q(h) \leftarrow hasLocation(h, hl) \wedge hasLocation(\texttt{train}, cl) \wedge close(hl, cl) \wedge cheap(h)$$

Health-Care: Diagnosis of Pneumonia



- Given that $Temp = 37.5$, $Pulse = 98$, $RespiratoryRate = 18$, which is the probability that a patient has pneumonia?
- $Temp = 37.5$, $Pulse = 98$, $RespiratoryRate = 18$ are in the "danger zone".
- These constraints are rather imprecise than crisp.

Uncertainty vs. Vagueness: A Clarification

- What does the degree mean?
- There is often a misunderstanding between interpreting a degree as a measure of uncertainty or as a measure of vagueness.
- The value 0.83 has a different interpretation in "Birds fly to degree 0.83" from that in "Hotel Verdi is close to the train station to degree 0.83".

Uncertainty

- Uncertainty: statements are true or false. But, due to lack of knowledge we can only estimate to which probability/possibility/necessity degree they are true or false.
  - For instance, a bird flies or does not fly. The probability/possibility/necessity degree that it flies is 0.83.
- Usually, we have a possible world semantics, with a distribution over possible worlds:

$$W = \{I \text{ classical interpretation}\}, \quad I(\varphi) \in \{0,1\}$$

$$\mu\colon W \to [0,1], \quad \mu(I) \in [0,1]$$

$$Pr(\phi) = \sum_{I \models \phi} \mu(I)$$

$$Poss(\phi) = \sup_{I \models \phi} \mu(I)$$

$$Necc(\phi) = \inf_{I \not\models \phi} \mu(I) = 1 - Poss(\neg\phi)$$

Vagueness

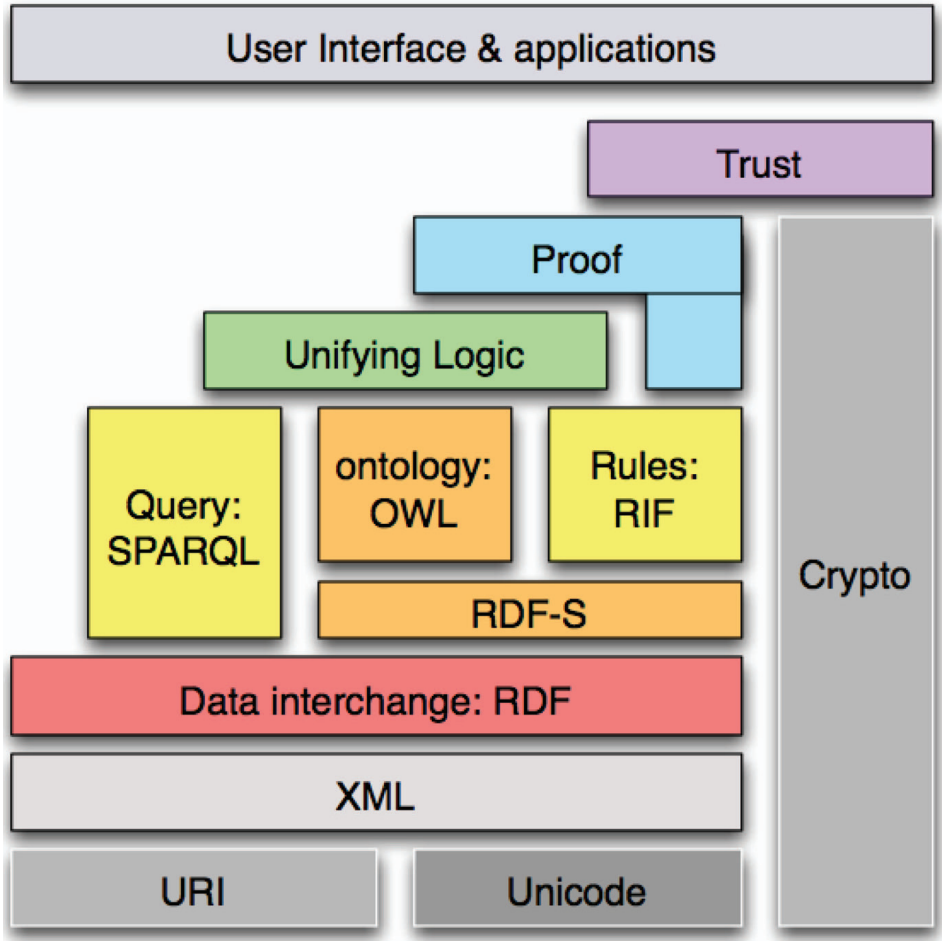
- Vagueness: statements involve concepts for which there is no exact definition, such as tall, small, close, far, cheap, expensive, isAbout, similarTo. Statements are true to some degree, which is taken from a truth space.
  - E.g., "Hotel Verdi is close to the train station to degree 0.83"
- Truth space: set of truth values L and a partial order $\leq$.
- Many-valued Interpretation: a function I mapping formulas into L, i.e., $I(\varphi) \in L$.
- Fuzzy Logic: $L = [0, 1]$.
- Uncertainty and Vagueness: "It is possible/probable to degree 0.83 that it will be hot tomorrow".

Outline

The Semantic Web...

- ...aims at an extension of the current WWW by standards and technologies that help machines to understand the information on the Web to support richer discovery, data integration, navigation, and automation of tasks.
- ...consists of several *hierarchical layers*, including
  - the Ontology layer: *OWL Web Ontology Language*: *OWL Lite* $\approx \mathcal{SHIF}(\mathbf{D})$, *OWL DL* $\approx \mathcal{SHOIN}(\mathbf{D})$, *OWL Full*;
  - the Rules, Logic, and Proof layers, which should offer sophisticated representation and reasoning capabilities.

Layers of the Semantic Web



OWL

- Three sublanguages of OWL:
  - OWL Full is union of OWL syntax and RDF (Undecidable)
  - OWL DL restricted to FOL fragment (decidable in NEXPTIME)
  - OWL Lite is "easier to implement" subset of OWL DL (decidable in EXPTIME)
- Semantic layering:
  - OWL DL within Description Logic (DL) fragment
- OWL DL based on $\mathcal{SHOIN}(D_n)$ DL
- OWL Lite based on $\mathcal{SHIF}(D_n)$ DL

Description Logics (DLs)

- The logics behind OWL DL and OWL Lite, `http://dl.kr.org/`.
- Concept/Class: names are equivalent to unary predicates
  - In general, concepts equiv. to formulas with one free variable
- Role or attribute: names are equivalent to binary predicates
  - In general, roles equiv. to formulas with two free variables
- Taxonomy: concept and role hierarchies can be expressed
- Individual: names are equivalent to constants
- Operators: restricted so that:
  - Language is decidable and, if possible, of low complexity
  - No need for explicit use of variables
    - Restricted form of $\exists$ and $\forall$
  - Features such as counting can be succinctly expressed

The DL Family

- A given DL is defined by set of concept and role forming operators
- Basic language: $\mathcal{ALC}$ ($\mathcal{A}$ttributive $\mathcal{L}$anguage with $\mathcal{C}$omplement)

Syntax				*Semantics*	*Example*
C, D	$\rightarrow$	$\top$	$\mid$	$\top(x)$	
		$\bot$	$\mid$	$\bot(x)$	
		A	$\mid$	$A(x)$	$Human$
		$C \sqcap D$	$\mid$	$C(x) \wedge D(x)$	$Human \sqcap Male$
		$C \sqcup D$	$\mid$	$C(x) \vee D(x)$	$Nice \sqcup Rich$
		$\neg C$	$\mid$	$\neg C(x)$	$\neg Meat$
		$\exists R.C$	$\mid$	$\exists y.R(x,y) \wedge C(y)$	$\exists has\_child.Blond$
		$\forall R.C$		$\forall y.R(x,y) \Rightarrow C(y)$	$\forall has\_child.Human$
$C \sqsubseteq D$				$\forall x.C(x) \Rightarrow D(x)$	$Happy\_Father \sqsubseteq Man \sqcap \exists has\_child.Female$
$a{:}C$				$C(a)$	$John{:}Happy\_Father$

Example

$$
\begin{aligned}
Sex &= Male \sqcup Female \\
Male \sqcap Female &\sqsubseteq \bot \\
Person &\sqsubseteq Human \sqcap \exists hasSex.Sex \\
MalePerson &\sqsubseteq Person \sqcap \exists hasSex.Male
\end{aligned}
$$

$$john{:}Person \sqcap \exists hasSex.\neg Female$$

$$KB \models john{:}MalePerson$$

DL Naming

$\mathcal{AL}$: $C, D \longrightarrow \top \mid \bot \mid A \mid C \sqcap D \mid \neg A \mid \exists R.\top \mid \forall R.C$

$\mathcal{C}$: Concept negation: $\neg C$. Thus, $\mathcal{ALC} = \mathcal{AL} + \mathcal{C}$

$\mathcal{S}$: Used for $\mathcal{ALC}$ with transitive roles $\mathcal{R}_+$

$\mathcal{U}$: Concept disjunction: $C_1 \sqcup C_2$

$\mathcal{E}$: Existential quantification: $\exists R.C$

$\mathcal{H}$: Role inclusion axioms: $R_1 \sqsubseteq R_2$,
e.g., $is\_component\_of \sqsubseteq is\_part\_of$

$\mathcal{N}$: Number restrictions: $(\geq n\ R)$ and $(\leq n\ R)$,
e.g., $(\geq 3\ has\_Child)$ (has at least 3 children)

$\mathcal{Q}$: Qualified number restrictions: $(\geq n\ R.C)$ and $(\leq n\ R.C)$,
e.g., $(\leq 2\ has\_Child.Adult)$ (has at most 2 adult children)

$\mathcal{O}$: Nominals (singleton class): $\{a\}$, e.g., $\exists has\_child.\{mary\}$.
Note: $a{:}C$ equiv. to $\{a\} \sqsubseteq C$, and $(a, b){:}R$ equiv. to $\{a\} \sqsubseteq \exists R.\{b\}$

$\mathcal{I}$: Inverse role: R^-, e.g., $isPartOf = hasPart^-$

$\mathcal{F}$: Functional role: f, e.g., $functional(hasAge)$

$\mathcal{R}_+$: transitive role: e.g., $transitive(isPartOf)$

For instance,

$$\mathcal{SHIF} = \mathcal{S} + \mathcal{H} + \mathcal{I} + \mathcal{F} = \mathcal{ALCR}_+\mathcal{HIF}$$ OWL-Lite (EXPTIME)

$$\mathcal{SHOIN} = \mathcal{S} + \mathcal{H} + \mathcal{O} + \mathcal{I} + \mathcal{N} = \mathcal{ALCR}_+\mathcal{HOIN}$$ OWL-DL (NEXPTIME)

Semantics of Additional Constructs

$\mathcal{H}$: Role inclusion axioms: $\mathcal{I} \models R_1 \sqsubseteq R_2$ iff ${R_1}^{\mathcal{I}} \subseteq {R_1}^{\mathcal{I}}$

$\mathcal{N}$: Number restrictions:
$(\geq n\ R)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} : |\{y \mid \langle x, y\rangle \in R^{\mathcal{I}}\}| \geq n\}$,
$(\leq n\ R)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} : |\{y \mid \langle x, y\rangle \in R^{\mathcal{I}}\}| \leq n\}$

$\mathcal{Q}$: Qualified number restrictions:
$(\geq n\ R.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} | \{y \mid \langle x, y\rangle \in R^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}| \geq n\}$,
$(\leq n\ R.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} : |\{y \mid \langle x, y\rangle \in R^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}| \leq n\}$

$\mathcal{O}$: Nominals (singleton class): $\{a\}^{\mathcal{I}} = \{a^{\mathcal{I}}\}$

$\mathcal{I}$: Inverse role: $(R^{-})^{\mathcal{I}} = \{\langle x, y\rangle \mid \langle y, x\rangle \in R^{\mathcal{I}}\}$

$\mathcal{F}$: Functional role: $I \models \mathit{functional}(f)$ iff
$\forall z \forall y \forall z$ if $\langle x, y\rangle \in f^{\mathcal{I}}$ and $\langle x, z\rangle \in f^{\mathcal{I}}$ then $y = z$

$\mathcal{R}_{+}$: Transitive role: $I \models \mathit{transitive}(r)$ iff $r^{\mathcal{I}}$ is transitive

Concrete Domains

- Concrete domains: reals, integers, strings, ...

  $(tim, 14)$:*hasAge*
  $(sf, \text{“}SoftComputing\text{”})$:*hasAcronym*
  $(source1, \text{“}ComputerScience\text{”})$:*isAbout*
  $(service2, \text{“}InformationRetrievalTool\text{”})$:*Matches*
  $Minor = Person \sqcap \exists hasAge. \leq_{18}$

- Semantics: a clean separation between “object” classes and concrete domains
  - $D = \langle \Delta_D, \Phi_D \rangle$
  - Δ_D is an interpretation domain
  - Φ_D is the set of concrete domain predicates d with a predefined arity n and fixed interpretation $d^D \subseteq \Delta_D^n$
  - Concrete properties: $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta_D$
- Notation: (D). E.g., $\mathcal{ALC}(D)$ is $\mathcal{ALC}$ + concrete domains

OWL DL as Description Logic

Concept/Class constructors:

Abstract Syntax	DL Syntax	Example
Descriptions (C)		
A (URI reference)	A	Conference
owl:Thing	$\top$	
owl:Nothing	$\bot$	
intersectionOf(C_1 C_2 ...)	$C_1 \sqcap C_2$	Reference $\sqcap$ Journal
unionOf(C_1 C_2 ...)	$C_1 \sqcup C_2$	Organization $\sqcup$ Institution
complementOf(C)	$\neg C$	$\neg$ MasterThesis
oneOf(o_1 ...)	$\{o_1, \ldots\}$	{"WISE","ISWC",...}
restriction(R someValuesFrom(C))	$\exists R.C$	$\exists$parts.InCollection
restriction(R allValuesFrom(C))	$\forall R.C$	$\forall$date.Date
restriction(R hasValue(o))	$\exists R.\{o\}$	$\exists$date.{2005}
restriction(R minCardinality(n))	$(\geq n\ R)$	($\geqslant$ 1 location)
restriction(R maxCardinality(n))	$(\leq n\ R)$	($\leqslant$ 1 publisher)
restriction(U someValuesFrom(D))	$\exists U.D$	$\exists$issue.integer
restriction(U allValuesFrom(D))	$\forall U.D$	$\forall$name.string
restriction(U hasValue(v))	$\exists U. =_v\}$	$\exists$series.="LNCS"
restriction(U minCardinality(n))	$(\geq n\ U)$	($\geqslant$ 1 title)
restriction(U maxCardinality(n))	$(\leq n\ U)$	($\leqslant$ 1 author)

R is an abstract role, while U is a concrete property of arity 2.

Axioms:

Abstract Syntax	DL Syntax	Example
Axioms		
Class(A partial $C_1 \ldots C_n$)	$A \sqsubseteq C_1 \sqcap \ldots \sqcap C_n$	$Human \sqsubseteq Animal \sqcap Biped$
Class(A complete $C_1 \ldots C_n$)	$A = C_1 \sqcap \ldots \sqcap C_n$	$Man = Human \sqcap Male$
EnumeratedClass(A $o_1 \ldots o_n$)	$A = \{o_1\} \sqcup \ldots \sqcup \{o_n\}$	$RGB = \{r\} \sqcup \{g\} \sqcup \{b\}$
SubClassOf(C_1 C_2)	$C_1 \sqsubseteq C_2$	
EquivalentClasses($C_1 \ldots C_n$)	$C_1 = \ldots = C_n$	
DisjointClasses($C_1 \ldots C_n$)	$C_i \sqcap C_j = \bot, i \neq j$	$Male \sqcap Female \sqsubseteq \bot$
ObjectProperty(R super (R_1) ... super (R_n)	$R \sqsubseteq R_i$	$HasDaughter \sqsubseteq hasChild$
domain(C_1) ... domain(C_n)	$(\geq 1\ R) \sqsubseteq C_i$	$(\geq 1\ hasChild) \sqsubseteq Human$
range(C_1) ... range(C_n)	$\top \sqsubseteq \forall R.C_i$	$\top \sqsubseteq \forall hasChild.Human$
[inverseof(P)]	$R = P^-$	$hasChild = hasParent^-$
[symmetric]	$R = R^-$	$similar = similar^-$
[functional]	$\top \sqsubseteq (\leq 1\ R)$	$\top \sqsubseteq (\leq 1\ hasMother)$
[Inversefunctional]	$\top \sqsubseteq (\leq 1\ R^-)$	
[Transitive])	$Tr(R)$	$Tr(ancestor)$
SubPropertyOf(R_1 R_2)	$R_1 \sqsubseteq R_2$	
EquivalentProperties($R_1 \ldots R_n$)	$R_1 = \ldots = R_n$	$cost = price$
AnnotationProperty(S)		

Abstract Syntax	DL Syntax	Example
DatatypeProperty(U super (U_1) ... super (U_n)	$U \sqsubseteq U_i$	
domain(C_1) ... domain(C_n)	$(\geq 1\ U) \sqsubseteq C_i$	$(\geq 1\ hasAge) \sqsubseteq Human$
range(D_1) ... range(D_n)	$\top \sqsubseteq \forall U.D_i$	$\top \sqsubseteq \forall hasAge.posInteger$
[functional])	$\top \sqsubseteq (\leq 1\ U)$	$\top \sqsubseteq (\leq 1\ hasAge)$
SubPropertyOf($U_1\ U_2$)	$U_1 \sqsubseteq U_2$	$hasName \sqsubseteq hasFirstName$
EquivalentProperties($U_1 \ldots U_n$)	$U_1 = \ldots = U_n$	
Individuals		
Individual(o type (C_1) ... type (C_n))	$o{:}C_i$	$tim{:}Human$
value($R_1\ o_1$) ... value($R_n\ o_n$)	$(o, o_i){:}R_i$	$(tim, mary){:}hasChild$
value($U_1\ v_1$) ... value($U_n\ v_n$)	$(o, v_1){:}U_i$	$(tim, 14){:}hasAge$
SameIndividual($o_1 \ldots o_n$)	$o_1 = \ldots = o_n$	$president\_Bush = G.W.Bush$
DifferentIndividuals($o_1 \ldots o_n$)	$o_i \neq o_j, i \neq j$	$john \neq peter$
Symbols		
Object Property R (URI reference)	R	hasChild
Datatype Property U (URI reference)	U	hasAge
Individual o (URI reference)	U	tim
Data Value v (RDF literal)	U	"ESWC07"

Outline

Uncertainty and Logic

- Any statement is either true or false
- Due to lack of knowledge, we can only estimate to which probability/possibility/necessity degree they are true or false
- Ususally, we have a possible world semantics with a distribution over possible worlds
- Possible world: any classical interpretation I, mapping any statement ϕ into $\{0, 1\}$:

$$W = \{\text{I classical interpretation}\},\ I(\phi) \in \{0, 1\}$$

- Distribution: a mapping

$$\mu\colon W \to [0, 1],\ \mu(I) \in [0, 1]$$

  satisfying some additional conditions (probability/possibility)
- $\mu(I)$ is the probability/possibility that the world I is indeed the actual one

- The probability of a statement ϕ is determined as

$$Pr(\phi) \;=\; \sum_{I \models \phi} \mu(I)$$

- The possibility of a statement ϕ is determined as

$$Poss(\phi) \;=\; \sup_{I \models \phi} \mu(I)$$

- The necessity of a statement ϕ is determined as

$$Nec(\phi) \;=\; 1 - Poss(\neg\phi) \;=\; \inf_{I \not\models \phi} 1 - \mu(I)$$

Probabilistic setting:

$$\phi = sprinklerOn \vee wet$$

W	sprinklerOn	wet	μ
I_1	0	0	0.1
I_2	0	1	0.2
I_3	1	0	0.4
I_4	1	1	0.3

$$1 = \sum_{I \in W} \mu(I)$$

$$Pr(\phi) = 0.2 + 0.4 + 0.3 = 0.9$$

Possibilistic setting:

$$\phi = sprinklerOn \vee wet$$

W	sprinklerOn	wet	μ
I_1	0	0	0.3
I_2	0	1	1.0
I_3	1	0	0.8
I_4	1	1	1.0

$$1 = \sup_{I \in W} \mu(I)$$

$$Poss(\phi) = \sup(1, 0.8, 1) = 1$$

$$Nec(\phi) = 1 - Poss(\neg\phi) = 1 - 0.3 = 0.7$$

Properties of Probabilities

$$
\begin{array}{lcl}
Pr(\phi \wedge \psi) & = & Pr(\phi) + Pr(\psi) - Pr(\phi \vee \psi) \\
Pr(\phi \wedge \psi) & \leq & \min(Pr(\phi), Pr(\psi)) \\
Pr(\phi \wedge \psi) & \geq & \max(0, Pr(\phi) + Pr(\psi) - 1) \\
Pr(\phi \vee \psi) & = & Pr(\phi) + Pr(\psi) - Pr(\phi \wedge \psi) \\
Pr(\phi \vee \psi) & \leq & \min(1, Pr(\phi) + Pr(\psi)) \\
Pr(\phi \vee \psi) & \geq & \max(Pr(\phi), Pr(\psi)) \\
Pr(\neg\phi) & = & 1 - Pr(\phi) \\
Pr(\bot) & = & 0 \\
Pr(\top) & = & 1
\end{array}
$$

Properties of Possibilities/Necessities

$$
\begin{array}{lcl}
Poss(\phi \wedge \psi) & \leq & \min(Poss(\phi), Poss(\psi)) \\
Poss(\phi \vee \psi) & = & \max(Poss(\phi), Poss(\psi)) \\
Poss(\neg\phi) & = & 1 - Nec(\phi) \\
Poss(\bot) & = & 0 \\
Poss(\top) & = & 1 \quad \text{(in the normalized case)} \\
\\
Nec(\phi \wedge \psi) & = & \min(Nec(\phi), Nec(\psi)) \\
Nec(\phi \vee \psi) & \geq & \max(Nec(\phi), Nec(\psi)) \\
Nec(\neg\phi) & = & 1 - Poss(\phi) \\
Nec(\bot) & = & 0 \quad \text{(in the normalized case)} \\
Nec(\top) & = & 1
\end{array}
$$

Probabilistic Logic

- Integration of (propositional) logic- and probability-based representation and reasoning formalisms.
- Reasoning from logical constraints and interval restrictions for conditional probabilities (also called *conditional constraints*).
- Reasoning from convex sets of probability distributions.
- Model-theoretic notion of logical entailment.

Syntax of Probabilistic Knowledge Bases

- Finite nonempty set of basic events $\Phi = \{p_1, \ldots, p_n\}$.
- Event ϕ: Boolean combination of basic events
- Logical constraint $\psi \Leftarrow \phi$: events ψ and ϕ: "ϕ implies ψ".
- Conditional constraint $(\psi|\phi)[l, u]$: events ψ and ϕ, and $l, u \in [0, 1]$: "conditional probability of ψ given ϕ is in $[l, u]$".
- Probabilistic knowledge base $KB = (L, P)$:
  - finite set of logical constraints L,
  - finite set of conditional constraints P.

Example

Probabilistic knowledge base $KB = (L, P)$:

- $L = \{bird \Leftarrow eagle\}$:

  "All eagles are birds".

- $P = \{(have\_legs \mid bird)[1, 1], (fly \mid bird)[0.95, 1]\}$:

  "All birds have legs".
  "Birds fly with a probability of at least 0.95".

Semantics of Probabilistic Knowledge Bases

- World I: truth assignment to all basic events in Φ.
- $\mathcal{I}_\Phi$: all worlds for Φ.
- Probabilistic interpretation Pr: probability function on $\mathcal{I}_\Phi$.
- $Pr(\phi)$: sum of all $Pr(I)$ such that $I \in \mathcal{I}_\Phi$ and $I \models \phi$.
- $Pr(\psi|\phi)$: if $Pr(\phi) > 0$, then $Pr(\psi|\phi) = Pr(\psi \wedge \phi) \,/\, Pr(\phi)$.
- Truth under Pr:
  - $Pr \models \psi \Leftarrow \phi$ iff $Pr(\psi \wedge \phi) = Pr(\phi)$
    (iff $Pr(\psi \Leftarrow \phi) = 1$).
  - $Pr \models (\psi|\phi)[l, u]$ iff $Pr(\psi \wedge \phi) \in [l, u] \cdot Pr(\phi)$
    (iff either $Pr(\phi) = 0$ or $Pr(\psi|\phi) \in [l, u]$).

Example

- Set of basic propositions $\Phi = \{bird, fly\}$.
- $\mathcal{I}_\Phi$ contains exactly the worlds I_1, I_2, I_3, and I_4 over Φ:

	fly	$\neg fly$
$bird$	I_1	I_2
$\neg bird$	I_3	I_4

- Some probabilistic interpretations:

Pr_1	fly	$\neg fly$
$bird$	19/40	1/40
$\neg bird$	10/40	10/40

Pr_2	fly	$\neg fly$
$bird$	0	1/3
$\neg bird$	1/3	1/3

- $Pr_1(fly \wedge bird) = 19/40$ and $Pr_1(bird) = 20/40$.
- $Pr_2(fly \wedge bird) = 0$ and $Pr_2(bird) = 1/3$.
- $\neg fly \Leftarrow bird$ is false in Pr_1, but true in Pr_2.
- $(fly \mid bird)[.95, 1]$ is true in Pr_1, but false in Pr_2.

Satisfiability and Logical Entailment

- Pr is a model of $KB = (L, P)$ iff $Pr \models F$ for all $F \in L \cup P$.
- KB is satisfiable iff a model of KB exists.
- $KB \Vdash (\psi|\phi)[l, u]$: $(\psi|\phi)[l, u]$ is a logical consequence of KB iff every model of KB is also a model of $(\psi|\phi)[l, u]$.
- $KB \Vdash_{tight} (\psi|\phi)[l, u]$: $(\psi|\phi)[l, u]$ is a tight logical consequence of KB iff l (resp., u) is the infimum (resp., supremum) of $Pr(\psi|\phi)$ subject to all models Pr of KB with $Pr(\phi) > 0$.

Hence, inference tasks in probabilistic logic: deciding satisfiability, deciding logical consequence, computing tight logically entailed intervals for conditional events.

Example

- Probabilistic knowledge base:

$$KB = (\{bird \Leftarrow eagle\}, \{(have\_legs \mid bird)[1,1], (fly \mid bird)[0.95,1]\}).$$

- KB is satisfiable, since

  Pr with $Pr(bird \wedge eagle \wedge have\_legs \wedge fly) = 1$ is a model.

- Some conclusions under logical entailment:

$$KB \models (have\_legs \mid bird)[0.3,1], \quad KB \models (fly \mid bird)[0.6,1].$$

- Tight conclusions under logical entailment:

$$KB \models_{tight} (have\_legs \mid bird)[1,1], \quad KB \models_{tight} (fly \mid bird)[0.95,1],$$
$$KB \models_{tight} (have\_legs \mid eagle)[1,1], \quad KB \models_{tight} (fly \mid eagle)[0,1].$$

Deciding Model Existence / Satisfiability

Theorem: The probabilistic knowledge base $KB = (L, P)$ has a model Pr with $Pr(\alpha) > 0$ iff the following system of linear constraints over the variables y_r $(r \in R)$, where $R = \{I \in \mathcal{I}_\Phi \mid I \models L\}$, is solvable:

$$\begin{aligned}
\sum_{r \in R,\, r \models \neg\psi \wedge \phi} -l\, y_r + \sum_{r \in R,\, r \models \psi \wedge \phi} (1-l)\, y_r &\geq 0 \quad (\forall (\psi|\phi)[l,u] \in P) \\
\sum_{r \in R,\, r \models \neg\psi \wedge \phi} u\, y_r + \sum_{r \in R,\, r \models \psi \wedge \phi} (u-1)\, y_r &\geq 0 \quad (\forall (\psi|\phi)[l,u] \in P) \\
\sum_{r \in R,\, r \models \alpha} y_r &= 1 \\
y_r &\geq 0 \quad (\text{for all } r \in R)
\end{aligned}$$

Computing Tight Logical Consequences

Theorem: Suppose $KB = (L, P)$ has a model Pr such that $Pr(\alpha) > 0$. Then, l (resp., u) such that $KB \Vdash_{tight} (\beta|\alpha)[l, u]$ is given by the optimal value of the following linear program over the variables y_r ($r \in R$), where $R = \{I \in \mathcal{I}_\Phi \mid I \models L\}$:

$$\text{minimize (resp., maximize)} \sum_{r \in R,\, r \models \beta \wedge \alpha} y_r \quad \text{subject to}$$

$$\begin{aligned}
\sum_{r \in R,\, r \models \neg\psi \wedge \phi} -l\, y_r + \sum_{r \in R,\, r \models \psi \wedge \phi} (1 - l)\, y_r &\geq 0 \quad (\forall (\psi|\phi)[l, u] \in P) \\
\sum_{r \in R,\, r \models \neg\psi \wedge \phi} u\, y_r + \sum_{r \in R,\, r \models \psi \wedge \phi} (u - 1)\, y_r &\geq 0 \quad (\forall (\psi|\phi)[l, u] \in P) \\
\sum_{r \in R,\, r \models \alpha} y_r &= 1 \\
y_r &\geq 0 \quad (\text{for all } r \in R)
\end{aligned}$$

Towards Stronger Notions of Entailment

Problem: Inferential weakness of logical entailment.

Solutions:

- Probability selection techniques: Perform inference from a representative distribution of the encoded convex set of distributions rather than the whole set, e.g.,
  - distribution of maximum entropy,
  - distribution in the center of mass.
- Probabilistic default reasoning: Perform constraining rather than conditioning and apply techniques from default reasoning to resolve local inconsistencies.
- Probabilistic independencies: Further constrain the convex set of distributions by probabilistic independencies.
  ($\Rightarrow$ adds nonlinear equations to linear constraints)

Entailment under Maximum Entropy

- Entropy of a probabilistic interpretation Pr, denoted $H(Pr)$:

$$H(Pr) \quad = \quad -\sum_{I \in \mathcal{I}_\Phi} Pr(I) \cdot \log Pr(I)\,.$$

- The ME model of a satisfiable probabilistic knowledge base KB is the unique probabilistic interpretation Pr that is a model of KB and that has the greatest entropy among all the models of KB.
- $KB \models^{me} (\psi|\phi)[l, u]$: $(\psi|\phi)[l, u]$ is a ME consequence of KB iff the ME model of KB is also a model of $(\psi|\phi)[l, u]$.
- $KB \models^{me}_{tight} (\psi|\phi)[l, u]$: $(\psi|\phi)[l, u]$ is a tight ME consequence of KB iff for the ME model Pr of KB, it holds either (a) $Pr(\phi) = 0$, $l = 1$, and $u = 0$, or (b) $Pr(\phi) > 0$ and $Pr(\psi|\phi) = l = u$.

Logical vs. Maximum Entropy Entailment

Probabilistic knowledge base:

$$KB = (\{bird \Leftarrow eagle\}, \{(have\_legs \mid bird)[1,1], (fly \mid bird)[0.95,1]\}).$$

Tight conclusions under logical entailment:

$$KB \mid\!\models_{tight} (have\_legs \mid bird)[1,1], \quad KB \mid\!\models_{tight} (fly \mid bird)[0.95,1],$$
$$KB \mid\!\models_{tight} (have\_legs \mid eagle)[1,1], \quad KB \mid\!\models_{tight} (fly \mid eagle)[0,1].$$

Tight conclusions under maximum entropy entailment:

$$KB \mid\!\sim^{me}_{tight} (have\_legs \mid bird)[1,1], \quad KB \mid\!\sim^{me}_{tight} (fly \mid bird)[0.95,0.95],$$
$$KB \mid\!\sim^{me}_{tight} (have\_legs \mid eagle)[1,1], \quad KB \mid\!\sim^{me}_{tight} (fly \mid eagle)[0.95,0.95].$$

Lexicographic Entailment

- Pr verifies $(\psi|\phi)[l,u]$ iff $Pr(\phi)=1$ and $Pr \models (\psi|\phi)[l,u]$.
- P tolerates $(\psi|\phi)[l,u]$ under L iff $L \cup P$ has a model that verifies $(\psi|\phi)[l,u]$.
- $KB=(L,P)$ is consistent iff there exists an ordered partition $(P_0,\ldots,P_k)$ of P such that each P_i is the set of all $C \in P \setminus \bigcup_{j=0}^{i-1} P_j$ tolerated under L by $P \setminus \bigcup_{j=0}^{i-1} P_j$.
- This (unique) partition is called the z-partition of KB.

Let $KB = (L, P)$ be consistent, and $(P_0, \ldots, P_k)$ be its z-partition.

- Pr is *lex*-preferable to Pr' iff some $i \in \{0, \ldots, k\}$ exists such that
  - $|\{C \in P_i \mid Pr \models C\}| > |\{C \in P_i \mid Pr' \models C\}|$ and
  - $|\{C \in P_j \mid Pr \models C\}| = |\{C \in P_j \mid Pr' \models C\}|$ for all $i < j \leq k$.
- A model Pr of $\mathcal{F}$ is a *lex*-minimal model of $\mathcal{F}$ iff no model of $\mathcal{F}$ is *lex*-preferable to Pr.
- $KB \mid\!\sim^{lex} (\psi|\phi)[l, u]$: $(\psi|\phi)[l, u]$ is a *lex*-consequence of KB iff every *lex*-minimal model Pr of L with $Pr(\phi)=1$ satisfies $(\psi|\phi)[l, u]$.
- $KB \mid\!\sim^{lex}_{tight} (\psi|\phi)[l, u]$: $(\psi|\phi)[l, u]$ is a tight *lex*-consequence of KB iff l (resp., u) is the infimum (resp., supremum) of $Pr(\psi)$ subject to all *lex*-minimal models Pr of L with $Pr(\phi) = 1$.

Logical vs. Lexicographic Entailment

Probabilistic knowledge base:

$$KB = (\{bird \Leftarrow eagle\}, \{(have\_legs \mid bird)[1,1], (fly \mid bird)[0.95,1]\}).$$

Tight conclusions under logical entailment:

$$KB \Vdash_{tight} (have\_legs \mid bird)[1,1], \quad KB \Vdash_{tight} (fly \mid bird)[0.95,1],$$
$$KB \Vdash_{tight} (have\_legs \mid eagle)[1,1], \quad KB \Vdash_{tight} (fly \mid eagle)[0,1].$$

Tight conclusions under probabilistic lexicographic entailment:

$$KB \mid\!\sim^{lex}_{tight} (have\_legs \mid bird)[1,1], \quad KB \mid\!\sim^{lex}_{tight} (fly \mid bird)[0.95,1],$$
$$KB \mid\!\sim^{lex}_{tight} (have\_legs \mid eagle)[1,1], \quad KB \mid\!\sim^{lex}_{tight} (fly \mid eagle)[0.95,1].$$

Probabilistic knowledge base:

$$KB = (\{bird \Leftarrow penguin\}, \{(have\_legs \mid bird)[1,1], (fly \mid bird)[1,1], (fly \mid penguin)[0,0.05]\}).$$

Tight conclusions under logical entailment:

$$KB \mid\!\models_{tight} (have\_legs \mid bird)[1,1], \; KB \mid\!\models_{tight} (fly \mid bird)[1,1],$$
$$KB \mid\!\models_{tight} (have\_legs \mid penguin)[1,0], \; KB \mid\!\models_{tight} (fly \mid penguin)[1,0].$$

Tight conclusions under probabilistic lexicographic entailment:

$$KB \mid\!\sim^{lex}_{tight} (have\_legs \mid bird)[1,1], \; KB \mid\!\sim^{lex}_{tight} (fly \mid bird)[1,1],$$
$$KB \mid\!\sim^{lex}_{tight} (have\_legs \mid penguin)[1,1], \; KB \mid\!\sim^{lex}_{tight} (fly \mid penguin)[0,0.05].$$

Probabilistic knowledge base:

$$KB = (\{bird \Leftarrow penguin\}, \{(have\_legs \mid bird)[0.99, 1], (fly \mid bird)[0.95, 1], (fly \mid penguin)[0, 0.05]\}).$$

Tight conclusions under logical entailment:

$$KB \mid\!\models_{tight} (have\_legs \mid bird)[0.99, 1], \; KB \mid\!\models_{tight} (fly \mid bird)[0.95, 1],$$
$$KB \mid\!\models_{tight} (have\_legs \mid penguin)[0, 1], \; KB \mid\!\models_{tight} (fly \mid penguin)[0, 0.05].$$

Tight conclusions under probabilistic lexicographic entailment:

$$KB \mid\!\sim^{lex}_{tight} (have\_legs \mid bird)[0.99, 1], \; KB \mid\!\sim^{lex}_{tight} (fly \mid bird)[0.95, 1],$$
$$KB \mid\!\sim^{lex}_{tight} (have\_legs \mid penguin)[0.99, 1], \; KB \mid\!\sim^{lex}_{tight} (fly \mid penguin)[0, 0.05].$$

Bayesian Networks

Well-structured, exact conditional constraints plus conditional independencies specify exactly one joint probability distribution.

Joint probability distributions can answer any queries, but can be very large and are often hard to specify.

Bayesian network (BN): compact specification of a joint distribution, based on a graphical notation for conditional independencies:

- a set of nodes; each node represents a random variable
- a directed, acyclic graph (link $\approx$ "directly influences")
- a conditional distribution for each node given its parents: $\mathbf{P}(X_i|Parents(X_i))$

Any joint distribution can be represented as a BN.

Example

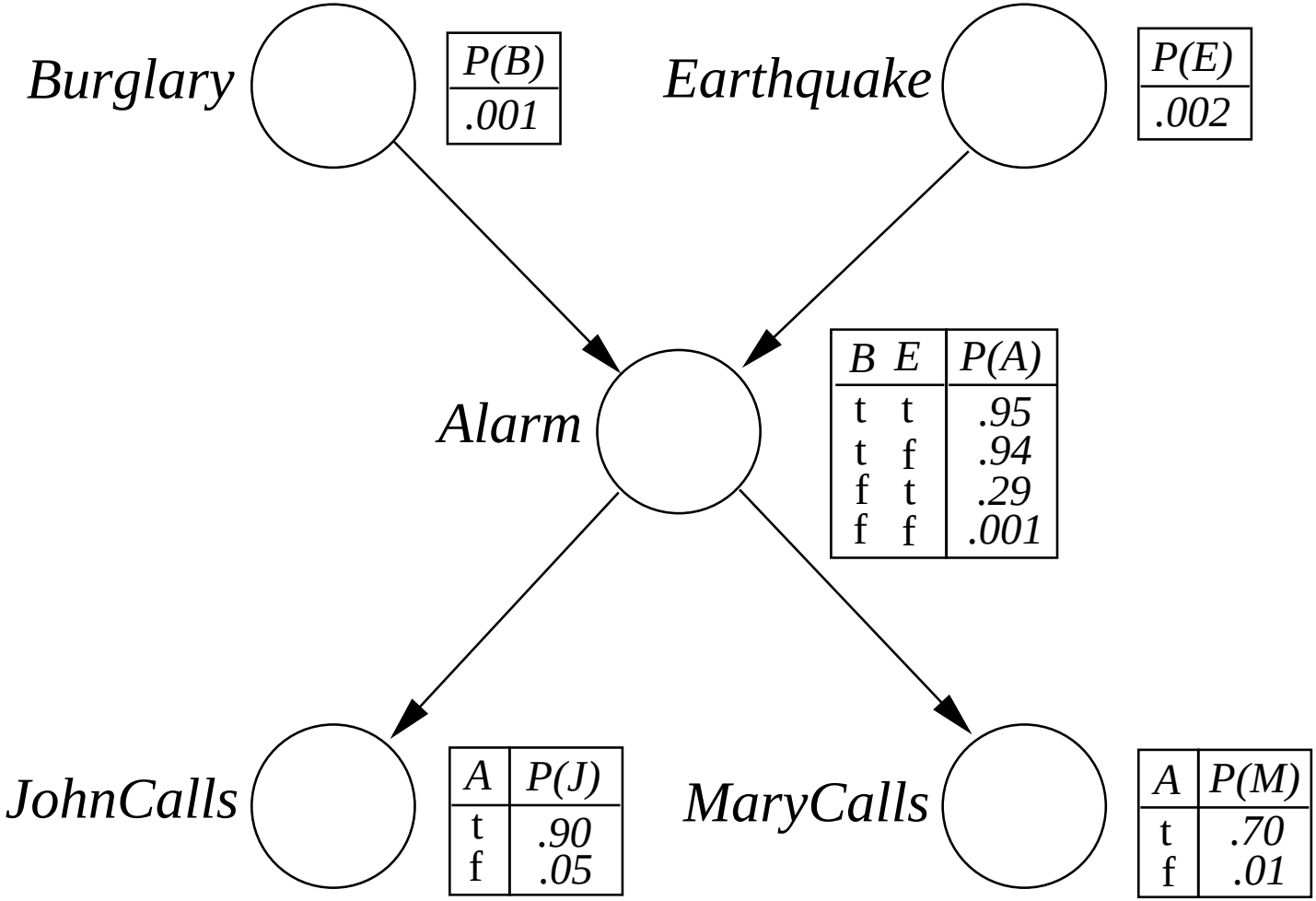
I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?

Variables: *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*

Network topology reflects "causal" knowledge:

- a burglar can set the alarm off
- an earthquake can set the alarm off
- the alarm can cause Mary to call
- the alarm can cause John to call

John sometimes confuses the telephone ringing with the alarm.
Mary likes rather loud music and sometimes misses the alarm.

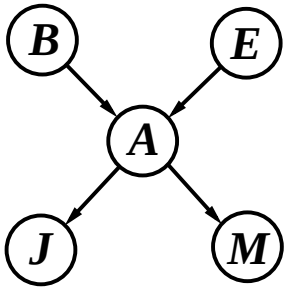

Burglary
P(B)
.001
Earthquake
P(E)
.002
Alarm
B E P(A)
t t .95
t f .94
f t .29
f f .001
JohnCalls
A P(J)
t .90
f .05
MaryCalls
A P(M)
t .70
f .01

Global Semantics

"Global" semantics defines the full joint distribution as the product of the local conditional distributions:

$$\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i))$$

e.g.,



$$\begin{aligned} P(j \wedge m \wedge a \wedge \neg b \wedge \neg e) &= P(j|a)P(m|a)P(a|\neg b, \neg e)P(\neg b)P(\neg e) \\ &= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 \\ &= 0.00062 \end{aligned}$$

Inference Tasks

- Computing unconditional and conditional probabilities $Pr(\phi)$ and $Pr(\phi|\psi)$ from a Bayesian network:
  - Simple queries: compute posterior marginal $\mathbf{P}(X_i|\mathbf{E}=\mathbf{e})$, e.g., $P(Burglary|Alarm=true,\ John=true,\ Mary=false)$.
  - Conjunctive queries: $\mathbf{P}(X_i, X_j|\mathbf{E}=\mathbf{e}) = \mathbf{P}(X_i|\mathbf{E}=\mathbf{e})\mathbf{P}(X_j|X_i, \mathbf{E}=\mathbf{e})$.
- Many algorithms: probability propagation (runs in linear time for polytrees), clustering, variable elimination, sampling, ...

Possibilistic Knowledge Bases

- Possibilistic formulas have the form $\mathrm{P}\ \phi \geq l$ or $\mathrm{N}\ \phi \geq l$ with $l \in [0, 1]$
- Encode to what extent ϕ is possibly resp. necessarily true
- Possibilistic interpretation: $\pi : \mathcal{I}_\Phi \rightarrow [0, 1]$
- $\pi(I)$ is the degree to which the world I is possible
- π is normalized: $\pi(I) = 1$ for some $I \in \mathcal{I}_\Phi$
- Possibility/Necessity of an event ϕ under π:

$$\begin{aligned} Poss(\phi) &= \sup\{\pi(I) | I \in \mathcal{I}_\Phi, I \models \phi\} \text{ (where } \max \emptyset = 0) \\ Necc(\phi) &= 1 - Poss(\neg\phi) \end{aligned}$$

- Truth under π:

$$\begin{aligned} \pi \models \mathrm{P}\ \phi \geq l \quad &\text{iff} \quad Poss(\phi) \geq l \\ \pi \models \mathrm{N}\ \phi \geq l \quad &\text{iff} \quad Nec(\phi) \geq l \end{aligned}$$

Deciding Entailment

- Reduction to classical entailment
- Let

$$\begin{aligned} KB_l &= \{\phi \mid \mathrm{N}\,\phi \geq l' \in KB,\ l' \geq l\} \\ KB^l &= \{\phi \mid \mathrm{N}\,\phi \geq l' \in KB,\ l' > l\} \end{aligned}$$

- Then:

$$\begin{aligned} KB \models \mathrm{N}\,\phi \geq l \quad &\text{iff} \quad KB_l \models \phi \\ KB \models \mathrm{P}\,\phi \geq l \quad &\text{iff} \quad KB^0 \models \phi \text{ or} \\ &\qquad \text{there is } \mathrm{P}\,\psi \geq l' \in KB \\ &\qquad \text{with } l' \geq l \text{ and } KB^{1-l'} \cup \{\psi\} \models \phi \end{aligned}$$

First-Order Logics of Probability

Rough classification of first-order logics of probability:

- types of probability structures
- underlying probabilistic formalism

Types of Probability Structures

- Probabilities on a set of possible worlds:
  a probability structure is a tuple (D, S, π, μ), where D is a domain, S is a set of possible worlds, π associates with every $s \in S$ an interpretation of the function and predicate symbols over D, and μ is a discrete probability function on S.

  "The probability that Tweety (a particular bird) flies is ≥ 0.9"

- Probabilities on the domain:
  a probability structure is a tuple (D, π, μ), where D is a domain, π is an interpretation of the function and predicate symbols over D, and μ is a discrete probability function on D, which is extended to D^n by $\mu(d_1, \ldots, d_n) = \Pi_{i=1}^{n} \mu(d_i)$.
  "The probability that a randomly chosen bird flies is at ≥ 0.9"
- Probabilities on both a set of possible worlds and the domain:
  a probability structure is a tuple $(D, S, \pi, \mu_D, \mu_S)$, where D is a domain, S is a set of possible worlds, π maps every $s \in S$ to an interpretation of the function and predicate symbols over D, and μ_D (resp., μ_S) is a discrete prob. function on S (resp., D).

Underlying Probabilistic Formalism

By far the most first-order probabilistic formalisms have a semantics in probabilities on a set of possible worlds.

Such a semantics can be easily based on one of the many existing semantics of propositional probabilistic formalisms:

- model-theoretic probabilistic logic
- probabilistic logic under maximum entropy
- strong nonmonotonic probabilistic logics
- Bayesian networks

A query on such a first-order probabilistic knowledge base KB is then evaluated by (i) constructing a relevant propositional prob. knowledge base KB', and (ii) evaluating the query on KB.

Very few first-order formalisms with probabilities on the domain and with probabilities on both a set of possible worlds and the domain.

Probabilistic Ontologies

Main types of encoded probabilistic knowledge:

- Terminological probabilistic knowledge about concepts and roles: "Birds fly with a probability of at least 0.95".
- Assertional probabilistic knowledge about instances of concepts and roles: "Tweety is a bird with a probability of at least 0.9".

Main types of reasoning problems:

- Satisfiability of the terminological probabilistic knowledge.
- Tight conclusions about generic objects (from the terminological probabilistic knowledge).
- Satisfiability of the assertional probabilistic knowledge.
- Tight conclusions about concrete objects (from both the terminological and the assertional probabilistic knowledge).

Probabilities about Generic and Concrete Objects

Combining generic and concrete probability distributions:

- Conditioning: Generic distributions are conditioned on the (classical) information about concrete distributions.
- Probabilistic default reasoning: Generic distributions are constrained by the (not necessarily classical) information about the concrete distributions, and techniques from default reasoning resolve local inconsistencies.
- Minimum cross entropy: Generic and concrete distributions are combined via cross entropy minimization.

Probabilistic DLs

R. Giugno, T. Lukasiewicz. P-$\mathcal{SHOQ}(\mathbf{D})$: A probabilistic extension of $\mathcal{SHOQ}(\mathbf{D})$ for probabilistic ontologies in the SW. In *Proc. JELIA-2002*.

- probabilistic generalization of the description logic $\mathcal{SHOQ}(\mathbf{D})$ (recently also extended to $\mathcal{SHIF}(\mathbf{D})$ and $\mathcal{SHOIN}(\mathbf{D})$)
- terminological probabilistic knowledge about concepts and roles
- assertional probabilistic knowledge about instances of concepts and roles
- terminological probabilistic inference based on lexicographic entailment in probabilistic logic (stronger than logical entailment)
- assertional probabilistic inference based on lexicographic entailment in probabilistic logic (for combining assertional and terminological probabilistic knowledge)
- terminological and assertional probabilistic inference problems reduced to sequences of linear optimization problems

M. Jaeger. Probabilistic reasoning in terminological logics. In *Proceedings KR-1994*.

- probabilistic generalization of the description logic $\mathcal{ALC}$
- terminological probabilistic knowledge about concepts and roles
- assertional probabilistic knowledge about concept instances, but no assertional probabilistic knowledge about role instances
- terminological probabilistic inference based on logical entailment in probabilistic logic (by solving linear optimization problems)
- assertional probabilistic inference based on cross entropy minimization relative to terminological probabilistic knowledge (by an approximation algorithm; no exact algorithm known so far)

D. Koller, A. Levy, and A. Pfeffer. P-Classic: A tractable probabilistic description logic. In *Proceedings AAAI-1997*.

- probabilistic generalization (of a variant) of the description logic Classic
- so-called *p-classes* express terminological probabilistic knowledge about concepts, roles, and attributes
- but assertional classical and probabilistic knowledge about instances of concepts and roles is not supported
- probabilistic semantics based on Bayesian networks
- determines exact probabilities for conditionals between concept expressions in canonical form
- probabilistic inference can be done in polynomial time, when the underlying Bayesian network is a polytree

Probabilistic OWL

P. C. G. da Costa. *Bayesian Semantics for the Semantic Web*. PhD thesis, George Mason University, Fairfax, VA, USA, 2005.

P. C. G. da Costa and K. B. Laskey. PR-OWL: A framework for probabilistic ontologies. In *Proceedings FOIS-2006*.

- probabilistic extension of OWL
- probabilistic semantics based on multi-entity Bayesian networks (MEBNs), which are a Bayesian logic that combines first-order logic with Bayesian probabilities:
  - represents knowledge as parameterized fragments of Bayesian networks
  - expresses repeated structure
  - represents probability distribution on interpretations of associated first-order theory

Use of Probabilistic Ontologies

- Representation of terminological and assertional probabilistic knowledge (e.g., in the medical domain or at the stock exchange market).
- Information retrieval, for an increased recall (e.g., Udrea et al.: Probabilistic ontologies and relational databases. In *Proc. CoopIS/DOA/ODBASE-2005*).
- Ontology matching (e.g., Mitra et al.: OMEN: A probabilistic ontology mapping tool. In *Proc. ISWC-2005*).
- Probabilistic data integration, especially for handling ambiguous and controversial pieces of information.

Possibilistic DLs

Generalization of DLs by possibilistic uncertainty, which is based on possibilistic interpretations rather than probabilistic interpretations.

Directly extends propositional possibilistic logic

Expressions: $\mathsf{P}\,\alpha \geq l$ or $\mathsf{N}\,\alpha \geq l$, where α is a classical description logic axiom and $l \in [0,1]$

Example:

$$\begin{aligned}
&\mathsf{N}\,(\exists \mathit{owns}.\mathit{Porsche} \sqsubseteq \mathit{richPerson} \sqcup \mathit{carFanatic}) \geq 0.8\\
&\mathsf{P}\,(\mathit{richPerson} \sqsubseteq \mathit{golfer}) \geq 0.7\\
&\mathsf{N}\,((\mathit{Tom}, 911) \colon \mathit{owns}) \geq 1\\
&\mathsf{N}\,(911 \colon \mathit{Porsche}) \geq 1,\\
&\mathsf{N}\,(\mathit{Tom} \colon \neg \mathit{carFanatic}) \geq 0.7
\end{aligned}$$

logically implies $\mathsf{P}\,(\mathit{Tom} \colon \mathit{golfer}) \geq 0.7$.

Outline

Vagueness and Logic

- Statements involve concepts with no exact definition, e.g.,
  - tall, small, close, far, cheap, expensive, is about, similar to, ...
- A statements is true to some degree, which is taken from a truth space, e.g.,
  - "Hotel Verdi is close to the train station to degree 0.83"
  - "The image is about a sun set to degree 0.75"
- Truth space: set of truth values L and a partial order $\leq$; usually: $L = [0, 1]$, but also $\{\frac{0}{n}, \frac{1}{n}, \ldots, \frac{n}{n}\}$ for an integer $n \geq 1$.
- Interpretation: a function I mapping atoms A into L, i.e., $I(A) \in L$.

- Problem: what is the interpretation of, e.g., $\phi \wedge \psi$?
  - E.g., given $I(\phi) = 0.83$ and $I(\psi) = 0.2$, what is the result of $I(\phi \wedge \psi) = 0.83 \wedge 0.2$?
  - E.g., in multimedia retrieval: if an image region is white to degree 0.8 and the object is about a dog to degree 0.4, to which degree is the image about a "white dog"? That is, what is $0.8 \wedge 0.4$?
- More generally, what is the result of $n \wedge m$, for $n, m \in [0, 1]$?
- The choice cannot be any arbitrary computable function, but has to reflect some basic properties that one expects to hold for a "conjunction".
- Norms: functions that are used to interpret connectives such as $\wedge$, $\vee$, $\neg$, and $\rightarrow$
  - t-norm: interprets conjunction
  - s-norm: interprets disjunction
- Norms are compatible with classical two-valued logic.

Axioms for Norms

Axioms for t-norms and s-norms:

Axiom Name	Conjunction Strategy	Disjunction Strategy
Tautology / Contradiction	$a \wedge 0 = 0$	$a \vee 1 = 1$
Identity	$a \wedge 1 = a$	$a \vee 0 = a$
Commutativity	$a \wedge b = b \wedge a$	$a \vee b = b \vee a$
Associativity	$(a \wedge b) \wedge c = a \wedge (b \wedge c)$	$(a \vee b) \vee c = a \vee (b \vee c)$
Monotonicity	if $b \leqslant c$, then $a \wedge b \leqslant a \wedge c$	if $b \leqslant c$, then $a \vee b \leqslant a \vee c$

Axioms for implication and negation functions:

Axiom Name	Implication Strategy	Negation Strategy
Tautology / Contradiction	$0 \rightarrow b = 1,\ \ a \rightarrow 1 = 1,\ \ 1 \rightarrow 0 = 0$	$\neg 0 = 1,\ \ \neg 1 = 0$
Antitonicity	if $a \leqslant b$, then $a \rightarrow c \geqslant b \rightarrow c$	if $a \leqslant b$, then $\neg a \geqslant \neg b$
Monotonicity	if $b \leqslant c$, then $a \rightarrow b \leqslant a \rightarrow c$	

Usually, $a \rightarrow b = \sup\{c \colon a \wedge c \leq b\}$ is used. It is called r-implication and depends on the t-norm only.

Typical Norms

	Łukasiewicz Logic	Gödel Logic	Product Logic	Zadeh Logic
$a \wedge b$	$\max(a+b-1, 0)$	$\min(a, b)$	$a \cdot b$	$\min(a, b)$
$a \vee b$	$\min(a+b, 1)$	$\max(a, b)$	$a+b-a \cdot b$	$\max(a, b)$
$a \rightarrow b$	$\min(1-a+b, 1)$	$\begin{cases} 1 & \text{if } a \leqslant b \\ b & \text{otherwise} \end{cases}$	$\min(1, b/a)$	$\max(1-a, b)$
$\neg a$	$1-a$	$\begin{cases} 1 & \text{if } a = 0 \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} 1 & \text{if } a = 0 \\ 0 & \text{otherwise} \end{cases}$	$1-a$

- Any other t-norm can be obtained as a combination of Lukasiewicz, Gödel, and Product t-norm.
- Zadeh Logic: not interesting for fuzzy logicians: it is a sublogic of Łukasiewicz:

$$\begin{aligned} \neg_Z \phi &= \neg_Ł \phi \\ \phi \wedge_Z \psi &= \phi \wedge_Ł (\phi \rightarrow_Ł \psi) \\ \phi \rightarrow_Z \psi &= \neg_Ł \phi \vee_Ł \psi \end{aligned}$$

Some additional properties of t-norms, s-norms, implication functions, and negation functions of various fuzzy logics.

Property	Łukasiewicz Logic	Gödel Logic	Product Logic	Zadeh Logic
$a \wedge \neg a = 0$	•	•	•	
$a \vee \neg a = 1$	•			
$a \wedge a = a$		•		•
$a \vee a = a$		•		•
$\neg\neg a = a$	•			•
$a \rightarrow b = \neg a \vee b$	•			•
$\neg(a \rightarrow b) = a \wedge \neg b$	•			•
$\neg(a \wedge b) = \neg a \vee \neg b$	•	•	•	•
$\neg(a \vee b) = \neg a \wedge \neg b$	•	•	•	•
$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$		•		•
$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$		•		•

If all conditions in the upper part of a column are satisfied, then we collapse to classical two-valued logic, i.e., $L = \{0, 1\}$.

Propositional Fuzzy Logics

- Formulas: propositional formulas
- Truth space is $[0, 1]$
- Formulas have a degree of truth in $[0, 1]$
- Interpretation: is a mapping $\mathcal{I} : \mathit{Atoms} \rightarrow [0, 1]$
- Interpretations are extended to formulas using norms to interpret connectives $\wedge, \vee, \neg, \rightarrow$

$$\begin{array}{lcl} I(\phi \wedge \psi) & = & I(\phi) \wedge I(\psi)\,; \\ I(\phi \vee \psi) & = & I(\phi) \vee I(\psi)\,; \\ I(\phi \rightarrow \psi) & = & I(\phi) \rightarrow I(\psi)\,; \\ I(\neg\phi) & = & \neg\, I(\phi)\,, \end{array}$$

- $r \in [0, 1]$ may appear as atom in formula, where $I(r) = r$.

Example

In Lukasiewicz Logic:

$$\phi = Cold \wedge Cloudy$$

I	$Cold$	$Cloudy$	$I(\phi)$
I_1	0	0.1	$\max(0, 0 + 0.1 - 1) = 0$
I_2	0.3	0.4	$\max(0, 0.3 + 0.4 - 1) = 0$
I_3	0.7	0.9	$\max(0, 0.7 + 0.9 - 1) = 0.6$
I_4	1	1	$\max(0, 1 + 1 - 1) = 1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

- Note:

$$I(r \rightarrow \phi) = 1 \quad \text{iff} \quad I(\phi) \geq r$$
$$I(\phi \rightarrow r) = 1 \quad \text{iff} \quad I(\phi) \leq r$$

- Semantics:

$$\begin{aligned} I \models \phi \quad &\text{iff} \quad I(\phi) = 1 \\ I \models KB \quad &\text{iff} \quad I \models \phi \text{ for all } \phi \in KB \\ KB \models \phi \quad &\text{iff} \quad \text{for all } I: \text{ if } I \models KB \text{ then } I \models \phi \end{aligned}$$

- Let:

$$\|\phi\|_{KB} = \inf\{I(\phi) \mid I \models KB\} \text{ (truth degree)}$$

$$|\phi|_{KB} = \sup\{r \mid KB \models r \rightarrow \phi\} \text{ (provability degree)}$$

then $\|\phi\|_{KB} = |\phi|_{KB}$.

- Also, in Lukasiewicz Logic:

$$|\neg\phi|_{KB} = 1 - |\phi|_{KB}$$

Proposition

$|\phi|_{KB} = \min x$ *such that* $KB \cup \{\phi \rightarrow x\}$ *satisfiable.*

Decision Algorithm (for Lukasiewicz Logic)

- We use MILP (Mixed Integer Linear Programming) to compute $|\phi|_{KB}$
- Let $r \in [0,1]$, variable or expresson $1 - r'$ (r' variable), admitting solution in $[0,1]$, $\neg r = 1 - r$, $\neg\neg r = r$
- For each propositional letter p, let x_p be a variable denoting the degree of truth of p
- Apply inference rules:

$r \rightarrow p$	$\mapsto$	$x_p \geq r, x_p \in [0,1]$
$p \rightarrow r$	$\mapsto$	$x_p \leq r, x_p \in [0,1]$
$r \rightarrow \neg\phi$	$\mapsto$	$\phi \rightarrow \neg r$
$\neg\phi \rightarrow r$	$\mapsto$	$\neg r \rightarrow \phi$
$r \rightarrow (\phi \wedge \psi)$	$\mapsto$	$x_1 \rightarrow \phi, x_2 \rightarrow \psi, y \leq 1 - r, x_i \leq 1 - y, x_1 + x_2 = r + 1 - y,$ $x_i \in [0,1], y \in \{0,1\}$
$(\phi \wedge \psi) \rightarrow r$	$\mapsto$	$x_1 \rightarrow \neg\phi, x_2 \rightarrow \neg\psi, x_1 + x_2 = 1 - r, x_i \in [0,1]$
$r \rightarrow (\phi \rightarrow \psi)$	$\mapsto$	$\phi \rightarrow x_1, x_2 \rightarrow \psi, r + x_1 - x_2 = 1, x_i \in [0,1]$
$(\phi \rightarrow \psi) \rightarrow r$	$\mapsto$	$x_1 \rightarrow \phi, \psi \rightarrow x_2, y - r \leq 0, y + x_1 \leq 1, y \leq x_2, y + r + x_1 - x_2 = 1,$ $x_i \in [0,1], y \in \{0,1\}$

- Now we have to solve a MILP problem of the form

$$\min \mathbf{c} \cdot \mathbf{x} \text{ s.t. } A\mathbf{x} + B\mathbf{y} \geq \mathbf{h}$$

where $a_{ij}, b_{ij}, c_l, h_k \in [0,1]$, x_i admits solutions in $[0,1]$, while y_j admits solutions in $\{0,1\}$

Example

- Consider $KB = \{0.6 \rightarrow p, 0.7 \rightarrow (p \rightarrow q)\}$
- Let us show that $|q|_{KB} = 0.6 \wedge 0.7 = \max(1, 0.6 + 0.7 - 1) = 0.3$
- Recall that $|q|_{KB} = \min x.$ such that $KB \cup \{q \rightarrow x\}$ satisfiable:

$$
\begin{aligned}
& KB \cup \{q \rightarrow x\} = \{0.6 \rightarrow p, 0.7 \rightarrow (p \rightarrow q), q \rightarrow x, x \in [0,1]\} \\
\mapsto \quad & \{x_p \geq 0.6, x_q \leq x, 0.7 \rightarrow (p \rightarrow q), \{x, x_p\} \subseteq [0,1]\} \\
\mapsto \quad & \{x_p \geq 0.6, x_q \leq x, p \rightarrow x_1, x_2 \rightarrow q, 0.7 + x_1 - x_2 = 1, \{x, x_p, x_i\} \subseteq [0,1]\} \\
\mapsto \quad & \{x_p \geq 0.6, x_q \leq x, x_p \leq x_1, x_q \geq x_2, 0.7 + x_1 - x_2 = 1, \{x, x_p, x_i\} \subseteq [0,1]\} = S
\end{aligned}
$$

It follows that $0.3 = \min x$ such that $Sat(S)$ satisfiable.

Predicate Fuzzy Logics

- Formulae: First-Order Logic formulae, *terms* are either variables or constants
  - we may introduce functions symbols as well, with crisp semantics (but uninteresting), or we need to discuss also fuzzy equality (which we leave out here)
- Truth space is $[0, 1]$
- Formulae have a a degree of truth in $[0, 1]$
- Interpretation: is a mapping $\mathcal{I} : \mathit{Atoms} \rightarrow [0, 1]$
- Interpretations are extended to formulae as follows:

$$
\begin{array}{rcl}
\mathcal{I}(\neg\phi) & = & \mathcal{I}(\phi) \rightarrow 0 \\
\mathcal{I}(\phi \wedge \psi) & = & \mathcal{I}(\phi) \wedge \mathcal{I}(\psi) \\
\mathcal{I}(\phi \rightarrow \psi) & = & \mathcal{I}(\phi) \rightarrow \mathcal{I}(\psi) \\
\mathcal{I}(\exists x \phi) & = & \sup\limits_{c \in \Delta^{\mathcal{I}}} \mathcal{I}_x^c(\phi) \\
\mathcal{I}(\forall x \phi) & = & \inf\limits_{c \in \Delta^{\mathcal{I}}} \mathcal{I}_x^c(\phi)
\end{array}
$$

  where $\mathcal{I}_x^c$ is as $\mathcal{I}$, except that variable x is mapped into individual c
- Definitions of $\mathcal{I} \models \phi$, $\mathcal{I} \models \mathcal{T}$, $\models \phi$, $\mathcal{T} \models \phi$, $||\phi||_{\mathcal{T}}$ and $|\phi|_{\mathcal{T}}$ are as for the propositional case

Fuzzy DLs

- In classical DLs, a concept C is interpreted by an interpretation $\mathcal{I}$ as a set of individuals
- In fuzzy DLs, a concept C is interpreted by $\mathcal{I}$ as a fuzzy set of individuals
- Each individual is instance of a concept to a degree in $[0, 1]$
- Each pair of individuals is instance of a role to a degree in $[0, 1]$

Fuzzy $\mathcal{ALC}$

The semantics is an immediate consequence of the First-Order-Logic translation of DLs expressions

Interpretation:

$$\begin{array}{lcl} \mathcal{I} & = & \Delta^{\mathcal{I}} \\ C^{\mathcal{I}} & : & \Delta^{\mathcal{I}} \to [0,1] \\ R^{\mathcal{I}} & : & \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \to [0,1] \end{array} \qquad \begin{array}{lcl} \wedge & = & \text{t-norm} \\ \vee & = & \text{s-norm} \\ \neg & = & \text{negation} \\ \to & = & \text{implication} \end{array}$$

Concepts:

Syntax			*Semantics*		
C, D	$\longrightarrow$	$\top$	$\top^{\mathcal{I}}(x)$	$=$	1
		$\bot$	$\bot^{\mathcal{I}}(x)$	$=$	0
		A	$A^{\mathcal{I}}(x)$	$\in$	$[0,1]$
		$C \sqcap D$	$(C_1 \sqcap C_2)^{\mathcal{I}}(x)$	$=$	${C_1}^{\mathcal{I}}(x) \wedge {C_2}^{\mathcal{I}}(x)$
		$C \sqcup D$	$(C_1 \sqcup C_2)^{\mathcal{I}}(x)$	$=$	${C_1}^{\mathcal{I}}(x) \vee {C_2}^{\mathcal{I}}(x)$
		$\neg C$	$(\neg C)^{\mathcal{I}}(x)$	$=$	$\neg C^{\mathcal{I}}(x)$
		$\exists R.C$	$(\exists R.C)^{\mathcal{I}}(x)$	$=$	$\sup_{y \in \Delta^{\mathcal{I}}} R^{\mathcal{I}}(x,y) \wedge C^{\mathcal{I}}(y)$
		$\forall R.C$	$(\forall R.C)^{\mathcal{I}}(u)$	$=$	$\inf_{y \in \Delta^{\mathcal{I}}} R^{\mathcal{I}}(x,y) \to C^{\mathcal{I}}(y)\}$

Assertions: $\langle a{:}C, r\rangle$, $\mathcal{I} \models \langle a{:}C, r\rangle$ iff $C^{\mathcal{I}}(a^{\mathcal{I}}) \geq r$ (similarly for roles)

- individual a is instance of concept C at least to degree r, $r \in [0,1] \cap \mathbb{Q}$

Inclusion axioms: $\langle C \sqsubseteq D, r\rangle$,

- $\mathcal{I} \models \langle C \sqsubseteq D, r\rangle$ iff $\inf_{x \in \Delta^{\mathcal{I}}}(C^{\mathcal{I}}(x) \to D^{\mathcal{I}}(x)) \geq r$.

Basic Inference Problems

Consistency: Check if knowledge is meaningful

- Is KB consistent, i.e., satisfiable?

Graded instantiation: Check if individual a instance of class C to degree at least r

- $KB \models \langle a{:}C, r\rangle$?

BTVB: Best Truth Value Bound problem

- $|a{:}C|_{KB} = \sup\{r \mid KB \models \langle a{:}C, r\rangle\}$?

Top-k retrieval: Retrieve the top-k individuals that instantiate C w.r.t. best truth value bound

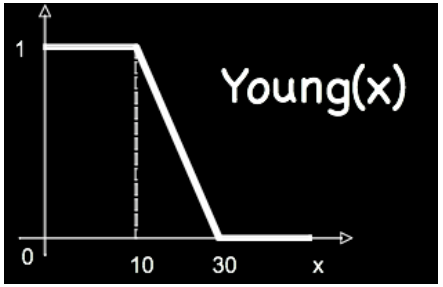
- $ans_{top-k}(KB, C) = Top_k\{\langle a, v\rangle \mid v = |a{:}C|_{KB}\}$

Towards Fuzzy OWL Lite and OWL DL

- Recall that OWL Lite and OWL DL relate to $\mathcal{SHIF}(D)$ and $\mathcal{SHOIN}(D)$, respectively
- We need to extend the semantics of fuzzy $\mathcal{ALC}$ to fuzzy $\mathcal{SHOIN}(D) = \mathcal{ALCHOINR}_+(D)$
- Additionally, we add
  - modifiers (e.g., *very*)
  - concrete fuzzy concepts (e.g., *Young*)
  - both additions have explicit membership functions

Concrete Fuzzy Concepts

- E.g., *Small*, *Young*, *High*, *etc*. with explicit membership function
- Use the idea of concrete domains:
  - $D = \langle \Delta_D, \Phi_D \rangle$
  - Δ_D is an interpretation domain
  - Φ_D is the set of concrete fuzzy domain predicates d with a predefined arity $n = 1, 2$ and fixed interpretation $d^D : \Delta_D^n \rightarrow [0, 1]$
  - For instance,



$$\begin{array}{rcl} Minor & = & Person \sqcap \exists hasAge.\ \leq_{18} \\ YoungPerson & = & Person \sqcap \exists hasAge.Young \\ & & functional(hasAge) \end{array}$$

Modifiers

- *Very*, *moreOrLess*, *slightly*, etc.
- Apply to fuzzy sets to change their membership function
  - $very(x) = x^2$
  - $slightly(x) = \sqrt{x}$
- For instance,

$$SportsCar \quad = \quad Car \sqcap \exists speed.very(High)$$

Fuzzy $\mathcal{SHOIN}(D)$

Concepts:

Syntax			Semantics
C, D	$\longrightarrow$	$\top$	$\top(x)$
		$\bot$	$\bot(x)$
		A	$A(x)$
		$(C \sqcap D)$	$C_1(x) \wedge C_2(x)$
		$(C \sqcup D)$	$C_1(x) \vee C_2(x)$
		$(\neg C)$	$\neg C(x)$
		$(\exists R.C)$	$\exists x\ R(x,y) \wedge C(y)$
		$(\forall R.C)$	$\forall x\ R(x,y) \rightarrow C(y)$
		$\{a\}$	$x = a$
		$(\geq n\ R)$	$\exists y_1, \ldots, y_n.\ \bigwedge_{i=1}^{n} R(x,y_i) \wedge \bigwedge_{1 \leq i < j \leq n} y_i \neq y_j$
		$(\leq n\ R)$	$\forall y_1, \ldots, y_{n+1}.\ \bigwedge_{i=1}^{n+1} R(x,y_i) \rightarrow \bigvee_{1 \leq i < j \leq n+1} y_i = y_j$
		FCC	$\mu_{FCC}(x)$
		$M(C)$	$\mu_M(C(x))$
R	$\longrightarrow$	P	$P(x,y)$
		P^-	$P(y,x)$

Assertions:

Syntax			Semantics
α	$\longrightarrow$	$\langle a{:}C, r\rangle$	$r \rightarrow C(a)$
		$\langle (a,b){:}R, r\rangle$	$r \rightarrow R(a,b)$

Axioms:

Syntax			Semantics
τ	$\longrightarrow$	$\langle C \sqsubseteq D, r\rangle$	$\forall x\ r \rightarrow (C(x) \rightarrow D(x))$, where $\rightarrow$ is r-implication
		$fun(R)$	$\forall x \forall y \forall z\ R(x,y) \wedge R(x,z) \rightarrow y = z$
		$trans(R)$	$(\exists z\ R(x,z) \wedge R(z,y)) \rightarrow R(x,y)$

Example (Graded Entailment)

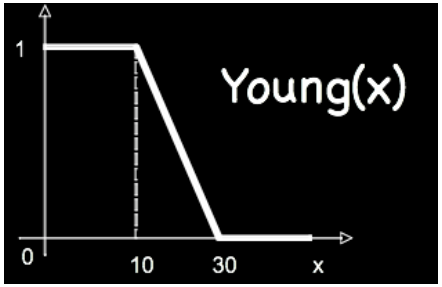
		
audi_tt	*mg*	*ferrari_enzo*

Car	*speed*
audi_tt	243
mg	≤ 170
ferrari_enzo	≥ 350

$$SportsCar \quad = \quad Car \sqcap \exists hasSpeed.very(High)$$

$$\begin{array}{lll} KB & \models & \langle ferrari\_enzo{:}SportsCar, 1\rangle \\ KB & \models & \langle audi\_tt{:}SportsCar, 0.92\rangle \\ KB & \models & \langle mg{:}\neg SportsCar, 0.72\rangle \end{array}$$

Example (Graded Subsumption)

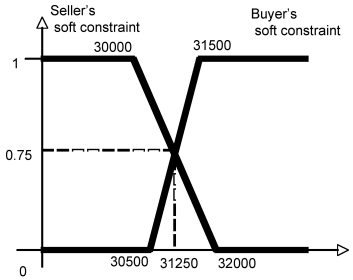


$$\begin{aligned} Minor &= Person \sqcap \exists hasAge. \leq_{18} \\ YoungPerson &= Person \sqcap \exists hasAge. Young \end{aligned}$$

$$KB \models \langle Minor \sqsubseteq YoungPerson, 0.2 \rangle$$

Note: without an explicit membership function of *Young*, this inference cannot be drawn

Example (Simplified Negotiation)



- a car seller sells an Audi TT for 31500 €, as from the catalog price.
- a buyer is looking for a sports-car, but wants to to pay not more than around 30000 €
- classical DLs: the problem relies on the crisp conditions on price

- more fine grained approach: to consider prices as fuzzy sets (as usual in negotiation)
  - seller may consider optimal to sell above 31500 €, but can go down to 30500 €
  - the buyer prefers to spend less than 30000 €, but can go up to 32000 €
    $AudiTT = SportsCar \sqcap \exists hasPrice.R(x; 30500, 31500)$
    $Query = SportsCar \sqcap \exists hasPrice.L(x; 30000, 32000)$
  - highest degree to which the concept
    $C = AudiTT \sqcap Query$
    is satisfiable is 0.75 (the possibility that the Audi TT and the query matches is 0.75)
  - the car may be sold at 31250 €

Reasoning

Depends on the semantics and reasoning method (tableau-based or MILP-based)

Tableaux method: under Zadeh semantics

- a tableau exists for fuzzy $\mathcal{SHIN}$, solving the satisfiability problem
- classical blocking methods apply similarly in the fuzzy variant
- the management of General concept inclusions (GCI's) is more complicated compared to the crisp case
- a translation of fuzzy $\mathcal{SHOIN}$ to crisp $\mathcal{SHOIN}$ also exists (not addressed here)
- the tableaux method is not suitable to deal with fuzzy concrete concepts and modifiers
- the BTVB can be solved, but not efficiently

MILP based method: under Zadeh semantics, Łukasiewicz semantics, and classical semantics

- exists for fuzzy $\mathcal{ALC}$ + linear modifiers + fuzzy concrete concepts
- exists for fuzzy $\mathcal{SHIF}$ + linear modifiers + fuzzy concrete concepts
- solves the BTVB as primary problem

MIQP based method: using Mixed Integer Quadratically Constrained Programming optimization problem (MICQP) for product T-norm

- exists for fuzzy $\mathcal{SHIF}$ + linear modifiers + fuzzy concrete concepts. Important as it simulates probabilistic reasoning under independent event assumption.
- solves the BTVB as primary problem
- the fuzzyDL solver also allows to mix all three semantics

Summary and Outlook

Summary: probabilistic, possibilistic, and fuzzy description logics for the Semantic Web: syntax, semantics, computation problems, solution techniques, and applications in the (Semantic) Web.

Current/Future Research:

- Implementation of systems for probabilistic, possibilistic, and fuzzy description logics. (Recent systems for all of them).
- Development of highly scalable probabilistic, possibilistic, and fuzzy description logics.
- Use of probabilistic, possibilistic, and fuzzy description logics in Web and Semantic Web applications.
- Learning of probabilistic, possibilistic, and fuzzy description logics from data.

Literature

Thomas Lukasiewicz and Umberto Straccia. Managing Uncertainty and Vagueness in Description Logics for the Semantic Web. Journal of Web Semantics, 6(4), 291-308, November 2008.