# Structural subsumption algorithm for Fuzzy Description Logics

Marco Cerami

Palacký University in Olomouc
Department of Computer Science
Olomouc, Czech Republic

Olomouc, 11.12.2014

european social fund in the czech republic
EUROPEAN UNION
MINISTRY OF EDUCATION, YOUTH AND SPORTS
OP Education for Competitiveness
INVESTMENTS IN EDUCATION DEVELOPMENT

# Introduction

# Introduction

- We address today the possibility of **generalizing the structural subsumption algorithm** for the classical description language $\mathcal{FL}^-$ to the finite-valued case.

- This language is interesting for us because, as it has been proved in [Brachman and Levesque, 1984], it has a **polynomial time** subsumption problem.

- As we will see, the classical algorithm is not complete for some FDLs, due to the **lack of idempotence**. Nevertheless, the subsumption problem is **still polynomial**.

- This presentation is based on part of the paper **Complexity Sources in Fuzzy Description Logic**, published in the proceedings of the International Workshop DL 2014. The paper is a joint work with U. Straccia.

# Preliminaries

# The language $\mathcal{FL}^-$

- The name $\mathcal{FL}$ stands for **frame language** because it has more or less the same expressive power of frame-based systems.
- Frame languages were studied in the 80's.
- Below we define the language $\mathcal{FL}^-$:

$$
\begin{array}{lcll}
C, D & \longrightarrow & A & \text{atomic concept} \\
 & & C \sqcap D & \text{conjunction} \\
 & & \forall R.C & \text{value restriction} \\
 & & \exists R.\top & \text{restricted existential quantif.}
\end{array}
$$

# Examples

Some examples of $\mathcal{FL}^-$ concepts:

$$\texttt{Person} \sqcap \forall \texttt{hasChild}.\texttt{Male}$$

"person who has only sons (if (s)he has children)"

$$\texttt{Person} \sqcap \exists \texttt{hasChild}.\top$$

"person who has a child"

$$\texttt{Person} \sqcap \forall \texttt{hasChild}.(\texttt{Male} \sqcap \exists \texttt{hasChild}.\top)$$

"person who has only sons of have a child"

# Classical interpretations

An **interpretation** is a pair $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ where:

- $\Delta^\mathcal{I}$ is a nonempty set, called **domain**;
- $\cdot^\mathcal{I}$ is an **interpretation function** that assigns:
  - to each **individual name** $a$ an element $a^\mathcal{I} \in \Delta^\mathcal{I}$,
  - to each **atomic concept** $A$ a subset of the domain set $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$,
  - to each **role name** $R$ a binary relation on the domain set $R^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$.

The interpretation function can be **inductively extended** to complex concepts in the following way:

$$(C \sqcap D)^\mathcal{I} = C^\mathcal{I} \cap D^\mathcal{I}$$
$$(\forall R.C)^\mathcal{I} = \{v \in \Delta^\mathcal{I} : \text{for every } w \in \Delta^\mathcal{I}, \text{ if } R^\mathcal{I}(v,w) \text{ then } C^\mathcal{I}(w)\}$$
$$(\exists R.\top)^\mathcal{I} = \{v \in \Delta^\mathcal{I} : \text{exists } w \in \Delta^\mathcal{I} \text{ such that } R^\mathcal{I}(v,w)\}$$

# Reasoning in $\mathcal{FL}^-$

- In $\mathcal{FL}^-$ concepts and axioms are **trivially satisfiable**.
- The reason for this is that in $\mathcal{FL}^-$ there is **no negation**.
- A concept $D$ is said to **subsume** a concept $C$ when, in every interpretation $\mathcal{I}$ it holds that

$$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}.$$

- We will consider this notion with respect to the **empty KB**.
- Differently from satisfiability, in $\mathcal{FL}^-$ it has **no trivial solution**, since the trivial model above is just one among all possible interpretations.

# Example

For example, concept

$$\texttt{Person}$$

is **not subsumed** by concept

$$\texttt{Person} \sqcap \texttt{Male}.$$

Consider the interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where:

- $\Delta^{\mathcal{I}} = \{v, w\}$,
- $\texttt{Person}^{\mathcal{I}} = \{v\}$,
- $\texttt{Male}^{\mathcal{I}} = \{w\}$,

Then, we have that $\texttt{Person}^{\mathcal{I}} = \{v\} \nsubseteq \emptyset = \texttt{Person}^{\mathcal{I}} \cap \texttt{Male}^{\mathcal{I}}$.

# Finite $t$-norms

We are considering **finite Łukasiewicz and Gödel $t$-norms** $Ł_n$ and $G_n$, that is, algebras with domain:

$$\{0, \frac{1}{n}, \ldots, \frac{n-1}{n}, 1\}$$

and **operations**:

| | Gödel | Łukasiewicz |
|---|---|---|
| $x * y$ | $\min(x, y)$ | $\max(0, x + y - 1)$ |
| $x \Rightarrow y$ | $\begin{cases} 1, & \text{if } x \leqslant y \\ y, & \text{otherwise} \end{cases}$ | $\min(1, 1 - x + y)$ |
| $\neg x$ | $\begin{cases} 1, & \text{if } x = 0 \\ 0, & \text{otherwise} \end{cases}$ | $1 - x$ |

# Interpretations

Being $\mathbf{T} \in \{Ł_n, G_n\}$, a **T-interpretation** is a pair $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ where:

- $\Delta^{\mathcal{I}}$ is a nonempty (crisp) set called **domain**,
- $\cdot^{\mathcal{I}}$ is a **fuzzy interpretation function** such that:
  - $A^{\mathcal{I}}\colon \Delta^{\mathcal{I}} \longrightarrow T$,
  - $R^{\mathcal{I}}\colon \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \longrightarrow T$,
  - $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$

The interpretation function can be **inductively extended** to complex concepts in the following way:

$$\begin{aligned}
(C \sqcap D)^{\mathcal{I}}(x) &:= C^{\mathcal{I}}(x) * D^{\mathcal{I}}(x) \\
(\forall R.C)^{\mathcal{I}}(x) &:= \inf_{y \in \Delta^{\mathcal{I}}}\{R^{\mathcal{I}}(x,y) \Rightarrow C^{\mathcal{I}}(y)\} \\
(\exists R.\top)^{\mathcal{I}}(x) &:= \sup_{y \in \Delta^{\mathcal{I}}}\{R^{\mathcal{I}}(x,y)\}
\end{aligned}$$

# Subsumption in $\mathbf{T}\text{-}\mathcal{FL}^-$

- As for $\mathcal{FL}^-$, also with finite-valued semantics, concepts and axioms are **trivially satisfiable**.
- A concept $D$ is said to **1-subsume** a concept $C$ when, in every interpretation $\mathcal{I}$ it holds that $C^\mathcal{I} \subseteq D^\mathcal{I}$.
- Though in the fuzzy case, a **graded notion of subsumption** can be defined, in this talk **we will restrict to 1-subsumption**.

# The structural subsumption algorithm for classical $\mathcal{FL}^-$

# Structural subsumption algorithm $SUBS?[D, C]$ from [Brachman and Levesque, 1984]

1: Flatten both $C$ and $D$ by removing all nested $\sqcap$ operators.

2: Collect all arguments to an $\forall R.$ for a given role $R$.

3: Assuming that $C := C_1 \sqcap \ldots \sqcap C_n$ and $D := D_1 \sqcap \ldots \sqcap D_m$, then return **true** iff for each $C_i$:

(a) if $D_i$ is an atom or a $\exists R.\top$, then one of $C_j$ is $D_i$.

(b) if $D_i$ is $\forall R.E$ then one of the $C_j$ is $\forall R.F$, where $SUBS?[F, E]$.

# Behavior

- From step 1 we have:

$$((C_1 \sqcap C_2) \sqcap C_3) \sqcap (C_4 \sqcap C_5) \quad \rightsquigarrow \quad C_1 \sqcap C_2 \sqcap C_3 \sqcap C_4 \sqcap C_5$$

which means that the conjunctions are treated as **sets of concepts**.

- From step 2 we have:

$$\forall R.C_1 \sqcap \forall R.(C_2 \sqcap \forall R.C_3) \quad \rightsquigarrow \quad \forall R.(C_1 \sqcap C_2 \sqcap \forall R.C_3)$$

which is possible since with classical semantics the following equivalence **always holds**:

$$\forall R.C_1 \sqcap \forall R.C_2 \equiv \forall R.(C_1 \sqcap C_2)$$

- After steps 1 and 2 we obtain **normalized concepts** with:
  - sets of atomic and quantified concepts...
  - which are eventually inside the scope of universal quantifiers...
  - that appear only once every role and nesting degree.
- From step 3 the algorithm **inductively checks** whether every concept in the consequent appears in the antecedent:

$$\underline{C_1} \sqcap C_2 \sqcap \forall R.(C_3 \sqcap \underline{C_4}) \quad \overset{\checkmark}{\sqsubseteq} \quad \underline{C_1} \sqcap \forall R.\underline{C_4}$$

$$\underline{C_1} \sqcap C_2 \sqcap \forall R.(C_3 \sqcap C_4) \quad \overset{!}{\not\sqsubseteq} \quad \underline{C_1} \sqcap C_4 \sqcap \forall R.C_2$$

# Complexity

In order to define the complexity of algorithm $SUBS?[D, C]$, let $n$ be the **length of the longer argument**. Then:

- **Step 1** can be done in time linear in $n$ (just erase parenthesis).
- **Step 2** may require that the entire concepts $C$ and $D$ are checked out a number of times **equal to their length**. Hence it can be done in $\mathcal{O}(n^2)$ time.
- **Step 3** may require that each of the concepts $C$ and $D$ is checked out a number of times **equal to the length of the other**. Hence it can be done in $\mathcal{O}(n^2)$ time.

Hence, algorithm $SUBS?[D, C]$ **operates in** $\mathcal{O}(n^2)$ time.

# Generalizing $SUBS?[D, C]$ to finite-valued FDLs

# The case of $G_n$

- The structural subsumption algorithm $SUBS?[D, C]$ **can be consistently used** in order to decide **1-subsumption** for $\mathbf{G_n}$-$\mathcal{FL}^-$.
- This is due to the fact that the Gödel $t$-norm $\wedge$ **works well with its residuum** $\Rightarrow_{\mathbf{G_n}}$. That is, for every $x, y, z \in G_n$:

$$x \Rightarrow_{\mathbf{G_n}} (y \wedge z) = (x \Rightarrow_{\mathbf{G_n}} y) \wedge (x \Rightarrow_{\mathbf{G_n}} z) \ .$$

- Note that subsumption between two concepts in $\mathbf{G_n}$-$\mathcal{FL}^-$ always takes either value 0 or value 1. Therefore, speaking about $(\geqslant r)$-or $(= r)$-subsumption in $\mathbf{G_n}$-$\mathcal{FL}^-$ **does not make sense**.

# Lack of idempotence in Ł$_n$ (I)

- Classical concept conjunctions can be seen as **sets of concepts**.
- Since Łukasiewicz conjunction is not idempotent, complex concepts where just $\sqcap$ appears as concept constructor **can not be seen as sets of atomic concepts**.
- In this sense, an inclusion like

$$A \sqsubseteq A \sqcap A$$

  which is valid in classical $\mathcal{FL}^-$ or in $\mathbf{G_n}$-$\mathcal{FL}^-$, **is not a 1-subsumption** in $\mathbf{Ł_n}$-$\mathcal{FL}^-$.
- Nevertheless, complex concepts in $\mathbf{Ł_n}$-$\mathcal{FL}^-$ can be seen as **multisets** of simpler concepts, that is, different **occurrences** of atomic concepts are now seen as different elements of a given complex concept.

# Lack of idempotence in $Ł_n$ (II)

- Unfortunately, the same result **does not hold** for Łukasiewicz $t$-norm $*_{Ł_n}$ and its residuum $\Rightarrow_{Ł_n}$.
- That is, there are $x, y, z \in Ł_n$ such that

$$x \Rightarrow_{Ł_n} (y *_{Ł_n} z) \neq (x \Rightarrow_{Ł_n} y) *_{Ł_n} (x \Rightarrow_{Ł_n} z) \ .$$

- As an **example**, if we take $x = y = z = 0.8$, then we have that

$$x \Rightarrow_{Ł_n} (y *_{Ł_n} z) = 0.8 \neq 1 = (x \Rightarrow_{Ł_n} y) *_{Ł_n} (x \Rightarrow_{Ł_n} z).$$

- Since the residuum plays a fundamental role in the **semantics of quantified concepts** in FDL, then in $\mathbf{Ł_n}$-$\mathcal{FL}^-$, concepts

$$\forall R.(C \sqcap D) \qquad \text{and} \qquad \forall R.C \sqcap \forall R.D$$

are **not equivalent**.

- This is a great **source of nondeterminism**, since now a matching between the quantified concepts appearing in $C$ and in $D$ should be found.

# Algorithm $Ł_{\mathbf{n}}$-$SUBS(1, D, C)$

**if** there is an occurrence of an atomic or existential conjunct $A$ of $D$ that is not in $C$ where concept $A$ appears in $C$ strictly less $n - 1$ times return **false**
**else**
    $\mathbf{E}_{C,D} := \varnothing$
    **for all** value restriction $\forall R.F$ which is a conjunct of $D$
        **for all** value restriction $\forall R.E$ which is a conjunct of $C$
            $\mathbf{E}_{C,D}(\forall R.F, \forall R.E) := Ł_{\mathbf{n}}$-$SUBS(1, F, E)$
        **end for**
    **end for**
    **if** there is a maximal bipartite matching for $\mathbf{E}_{C,D}$ including $C$
    return **true**
    **else**
        return **false**
    **end if**
**end if**

# Behavior

- **Step 1** handles subsumption between conjunctions of atomic concepts considering the lack of idempotence:

$$\underline{A} \sqcap B \sqcap \underline{A} \sqcap C \quad \overset{\checkmark}{\sqsubseteq} \quad \underline{A} \sqcap \underline{A}$$

$$\underline{A} \sqcap \underline{B} \sqcap B \sqcap C \quad \overset{!}{\not\sqsubseteq} \quad \underline{A} \sqcap \underline{B} \sqcap A$$

  Moreover takes into consideration that, if a subconcept $A$ of $C$ **occurs in a conjunction more than** $n-1$ **times**, being $n$ the cardinality of $Ł_n$, then its value is in $\{0,1\}$.

- **In steps 2 to 8** a set of **bipartite graphs** or **matrices** are inductively built, taking into consideration the quantified concepts that appear in the scope of a quantification $\forall R.$ with the same $R$.

# Construction of the matrices

Consider the subsumption:

$$\forall R.(\forall P.(A \sqcap B) \sqcap \forall P.C) \sqcap \forall R.(C \sqcap D) \sqsubseteq \forall R.(\forall P.B \sqcap \forall P.C) \sqcap \forall R.C$$

Then the following matrices are built up:

- For concepts

$$\forall R.(\forall P.(A \sqcap B \sqcap C) \sqcap \forall P.C) \sqcap \forall R.(C \sqcap D) \sqsubseteq \forall R.(\forall P.B \sqcap \forall P.C) \sqcap \forall R.C$$

| | $\forall P.B$ | $\forall P.C$ |
|---|---|---|
| $\forall P.(A \sqcap B \sqcap C)$ | $\times$ | $\times$ |
| $\forall P.C$ | | $\times$ |

- For concepts

$$\forall R.(\forall P.(A \sqcap B \sqcap C) \sqcap \forall P.C) \sqcap \forall R.(C \sqcap D) \sqsubseteq \forall R.(\forall P.B \sqcap \forall P.C) \sqcap \forall R.C$$

| | $\forall R.(\forall P.B \sqcap \forall P.C)$ | $\forall R.C$ |
|---|---|---|
| $\forall R.(\forall P.(A \sqcap B \sqcap C) \sqcap \forall P.C)$ | $\times$ | |
| $\forall R.(C \sqcap D)$ | | $\times$ |

- the matching between concepts that **do not contain further quantifiers** e.g. $\forall R.(C \sqcap D)$ and $\forall R.C$ is due to the fact that $\text{Ł}_n$-$SUBS(1, C, C \sqcap D)$ returns **true**.
- the matching between concepts that **contain further quantifiers** e.g. $\forall R.(\forall P.(A \sqcap B \sqcap C)$ and $\forall R.(\forall P.B \sqcap \forall P.C)$ is due to the fact that **there is a maximal matching** between the quantified concept in their scope.

- Finally, in **step 9** a subroutine for solving the **maximal matching problem for bipartite graphs** is called.

- In particular, this problem is known to be solved in polynomial time from the 1973 paper:

Hopkroft, J.E., Karp, R.M.: **An $n^{\frac{5}{2}}$ algorithm for maximum matchings in bipartite graphs**

- Note that the call to a subroutine for the bipartite matching problem **manages the nondeterminism** arising from the fact that concepts

$$\forall R.(C \sqcap D) \qquad \text{and} \qquad \forall R.C \sqcap \forall R.D$$

are **not equivalent**.

# Complexity

- **Steps 1 and 6** can be performed in **linear time**;
- each matrix $\mathbf{E}_{E,F}$ is **at most quadratic** on the size of the largest concept between $C$ and $D$;
- there are at most $|C| \times |D|$ **different matrices** $\mathbf{E}_{E,F}$;
- **the only non-deterministic problem** can be managed in polynomial time by a suitable procedure for the bipartite matching problem.

Hence algorithm $\text{Ł}_{\mathbf{n}}$-$SUBS(1, D, C)$ runs in $\mathcal{O}(n^4)$, where $n$ is the **largest between** $C$ **and** $D$.

Thank you for the attention !