# Preferential Temporal Description Logics with Typicality and Weighted Knowledge Bases

Mario Alviano [1], Francesco Bartoli [2], Marco Botta [2], Roberto Esposito [2], Laura Giordano [3], Daniele Theseider Dupré [3]

1 Università della Calabria, Italy
2 Università di Torino, Italy,
3 Università del Piemonte Orientale, Italy

CILC 2024, Roma, June 25-28, 2024

# Introduction

- *Weighted Conditional Knowledge Bases* (with *typicality*) allow for defeasible reasoning in DLs:
  - definition of *prototypical properties* of individuals and classes;
  - *weights* describe the *plausibility/implausibility* of properties;
  - *multi-preferential* and *"concept-wise" semantics* for conditionals: *preferences* $<_{C_i}$ *associated to concepts*.
- the multi-preferential semantics has been exploited to provide a *preferential interpretation to some neural network models*, SOMs and MLPs, for post hoc explanation [JELIA 2021, JLC2022]
- *model checking and entailment approaches for explainability* in Multi Layer networks for emotion recognition (J. of Automat. Reas. 2024).

The paper aims at developing a *temporal extension of preferential DLs* and of weighted conditional knowledge bases.

# Motivations

- This line of research investigates the *relationships* between *logics of commonsense reasoning* and some *neural network models*.
- A contribution towards *explainable and trustworthy AI*: verification of knowledge learned by a neural network by *model checking* in a conditional logic.
- In perspective, also towards *neuro-symbolic integration*, given that knowledge learned by neural networks is interpreted/verified in a logical language: *a multilayer network* can be seen *as a weighted conditional KB in a simple DL*.
- Including a *temporal dimension* is important:
  - to capture the temporal evolution of a system;
  - to capture the temporal properties of concepts, when exceptions are admitted:
    *"normally, professors teach at least a course until they retire"*

# Motivations

## Motivations from the standpoint of XAI

As understanding the logic underlying neural networks is important for explainability.

## Motivations for multiple preferences

The *relative typicality* of domain individuals *depends on the aspects* we are considering for comparison (TPLP 2020):

$$bob <_{SportLover} jim \text{ and } jim <_{Swimmer} bob$$

## Desiderata for preferential entailment:

*to satisfy the KLM properties* of system **P**; to deal with *specificity* and *irrelevance*; to avoid *"blockage of property inheritance"* [Pearl, 90] or *"drowning problem"* [Benferhat et al. 1993];

## Relations to cognitively inspired common-sense reasoning:

strong relationships to *Conceptual Spaces* (Gardenfors 2000) (and to SOMs) and to cognitively inspired common-sense reasoning: *Prototype-based* view.

# Example

A weighted $\mathcal{ALC}$ knowledge base $K = \langle \mathcal{T}_{strict}, \mathcal{T}_{Horse}, \mathcal{A} \rangle$

- $\mathcal{A} = \{Horse(buddy), Horse(spirit), RunFast(buddy), \ldots\}$;
- $\mathcal{T}_{strict}$ contains the strict inclusions:
  $Horse \sqsubseteq Mammal \qquad Mammal \sqsubseteq Animal$;
- $\mathcal{T}_{Horse}$ contains the typicality inclusions:
  $(d_1)\ \mathbf{T}(Horse) \sqsubseteq Tall, \qquad 4.5$
  $(d_2)\ \mathbf{T}(Horse) \sqsubseteq RunFast, \qquad 4.2$
  $(d_3)\ \mathbf{T}(Horse) \sqsubseteq Has\_Tail, \qquad 9.7$
  $(d_4)\ \mathbf{T}(Horse) \sqsubseteq Has\_Stripes, \qquad -20$

$\mathcal{T}_{Horse}$ used to *define an ordering* comparing typicality of domain elements as horses:

$$spirit <_{Horse} buddy$$

- Spirit is tall, has tail, *no stripes and does not run fast*;
- Buddy is tall, has tail, runs fast and has stripes.

# Multipreference semantics of a conditional KB

A *closure construction* to build a canonical preferential model:

- For each $C_i \in \mathcal{C}$, define *total preorder* $\leq_{C_i}$,

  $x \leq_{C_i} y$ means *"x is at least as typical as y wrt $C_i$"*.
- Given inclusions $(\mathbf{T}(C_i) \sqsubseteq D_{i,h}\ , w_h^i)$, the *weight of x wrt $C_i$* :

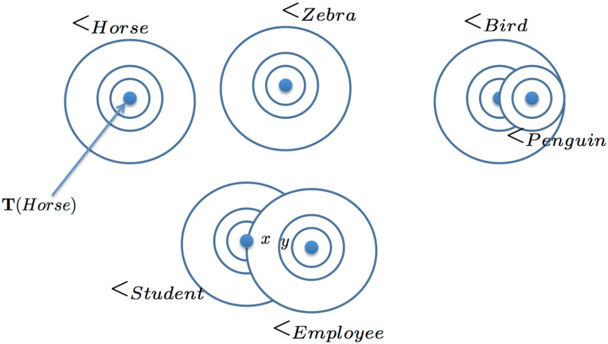
$$W_i(x) = \begin{cases} \sum_{h: x \in D^I_{i,h}} w_h^i & \text{if } x \in C_i^I \\ -\infty & \text{otherwise} \end{cases} \quad (1)$$

- *preference*: $x \leq_{C_i} y$ *iff* $W_i(x) \geq W_i(y)$

In the example: $W_{Horse}(spirit) = 14.2 > W_{Horse}(buddy) = -1.6$, hence $spirit <_{Horse} buddy$

- *Fuzzy approach* in [Jelia 2021, IJAR2024] for a logical characterization of multilayer networks
- *Finitely-valued* approach in [TPLP 2022, JELIA 2023] (with ASP).

# Multipreference interpretations



# Fuzzy interpretations and Typicality

- a semantic based on a *fuzzy interpretation of concepts* builds on fuzzy DLs [Straccia 2005, Stoilos 2005, Baader et al. 2008, Baader&Peñaloza 2011,...] and on *defeasible DLs* [Britz et al. 08, Giordano et al. 09, Casini&Straccia 10, ...]
- a domain element $x \in \Delta$ has a *degree of membership* $C^I(x) \in [0, 1]$ in a concept $C$.
- A fuzzy interpretation $I = \langle \Delta, \cdot^I \rangle$ *induces a preference relation* $<_C$ on $\Delta$ for all concepts $C$:

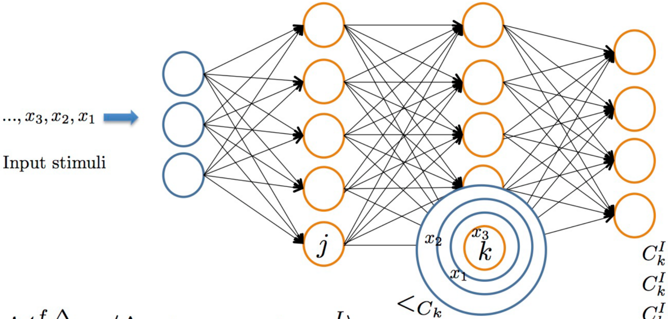
$$x <_{C_i} y \quad \text{iff} \quad C_i^I(x) > C_i^I(y)$$

- A notion of typicality can be defined: *typical C-elements* are the $<_C$-minimal $C$-elements:

$$(\mathbf{T}(C))^I(x) = \begin{cases} C^I(x) & \text{if there is no } y \text{ such that } y <_C x \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

- Satisfiability/entailment to verify fuzzy inclusion properties involving typicality concepts, such as $\langle \mathbf{T}(Penguin) \sqsubseteq Bird \geq 0.7 \rangle$

# Fuzzy interpretation of MLPs




$$C_k^I(x_1) = 0.5$$
$$C_k^I(x_2) = 0.3$$
$$C_k^I(x_3) = 0.8$$

$$\mathcal{M}_{\mathcal{N}}^{f,\Delta} = \langle \Delta, <_{C_1}, \ldots, <_{C_n}, \cdot^I \rangle$$
$$\Delta = \{x_1, \ldots, x_n\}$$

$$x_3 <_{C_k} x_1 <_{C_k} x_2$$

# A Temporal DL: $LTL_{\mathcal{ALC}}$

The concepts of the temporal description logic $LTL_{\mathcal{ALC}}$ can be formed from: standard constructors and the temporal operators $\bigcirc$ (next), $\mathcal{U}$ (until), $\diamond$ (eventually) and $\square$ (always) of linear time temporal logic (LTL).

The set of temporally extended concepts is as follows:

$$C ::= A \mid \top \mid \bot \mid C \sqcap D \mid C \sqcup D \mid \neg C \mid \forall r.C \mid \exists r.C \mid \bigcirc C \mid C \mathcal{U} D \mid \diamond C \mid \square C$$

where $A \in N_C$, and $C$ and $D$ are temporally extended concepts.

$LTL^{\mathbf{T}}_{\mathcal{ALC}}$

$$\mathbf{T}(Professor) \sqsubseteq (\exists teaches.Course) \mathcal{U} Retired$$

$$\exists lives\_in.Town \sqcap Young \sqsubseteq \mathbf{T}(\diamond \exists granted.Loan)$$

# Many-valued temporal interpretations for $LTL^{\mathbf{T}}_{\mathcal{ALC}}$

- We combine *fuzzy and many-valued DLs* [Straccia 2005, Stoilos 2005, Baader et al. 2008, ...] with a *temporal DL*.
- $\mathcal{S}$ a *truth degree set*, equipped with a preorder relation $\leq^{\mathcal{S}}$.
- An *$LTL^{\mathbf{T}}_{\mathcal{ALC}}$ interpretation* is a a pair $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where:
  - (i) $\Delta^{\mathcal{I}}$ is a non-empty *domain*;
  - (ii) $\cdot^{\mathcal{I}}$ is an *interpretation function* that maps:
    - each concept name $A \in N_C$ to a function $A^{\mathcal{I}} : \mathbb{N} \times \Delta^{\mathcal{I}} \rightarrow \mathcal{S}$
    - each role name $r \in N_R$ to a function $r^{\mathcal{I}} : \mathbb{N} \times \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \rightarrow \mathcal{S}$
    - and each individual name $a \in N_I$ to an element $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$
  - A notion of typicality can be defined:

$$x <^n_C y \text{ iff } C^I(n,x) > C^I(n,y)$$

    *typical $C$-elements* are the $<_C$-minimal elements:

$$(\mathbf{T}(C))^I(n,x) = \begin{cases} C^I(n,x) & \text{if there is no } y \text{ such that } y <^n_C x \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

- $C^I(n,x)$: the *degree of membership* of $x$ in $C$ at time point $n$

# Many-valued temporal interpretations for $LTL^{\mathbf{T}}_{\mathcal{ALC}}$

$$
\begin{aligned}
&\bot^{\mathcal{I}}(n,x) = 0, \;\; \top^{\mathcal{I}}(n,x) = 1\\
&(\neg C)^{\mathcal{I}}(n,x) = \ominus C^{\mathcal{I}}(n,x)\\
&(C \sqcap D)^{\mathcal{I}}(n,x) = C^{\mathcal{I}}(n,x) \otimes D^{\mathcal{I}}(n,x)\\
&(C \sqcup D)^{\mathcal{I}}(n,x) = C^{\mathcal{I}}(n,x) \oplus D^{\mathcal{I}}(n,x)\\
&(\exists r.C)^{\mathcal{I}}(n,x) = sup_{y \in \Delta}\; r^{\mathcal{I}}(n,x,y) \otimes C^{\mathcal{I}}(n,y)\\
&(\forall r.C)^{\mathcal{I}}(n,x) = inf_{y \in \Delta}\;\; r^{\mathcal{I}}(n,x,y) \rhd C^{\mathcal{I}}(n,y)\\
&(\bigcirc C)^{\mathcal{I}}(n,x) = C^{\mathcal{I}}(n+1,x)\\
&(\Diamond C)^{\mathcal{I}}(n,x) = \bigoplus_{m \geq n} C^{\mathcal{I}}(m,x)\\
&(\Box C)^{\mathcal{I}}(n,x) = \bigotimes_{m \geq n} C^{\mathcal{I}}(m,x) \; (C \mathcal{U} D)^{\mathcal{I}}(n,x)\\
&\qquad\qquad\qquad\;\; = \bigoplus_{m \geq n} (D^{\mathcal{I}}(m,x) \otimes \bigotimes_{k=n}^{m-1} C^{\mathcal{I}}(k,x))
\end{aligned}
$$

Following (Frigeri et al. 2014), one can introduce bounded versions for $\Diamond$, $\Box$ and $\mathcal{U}$

# Relation with FLTL

For the case $\mathcal{S} = [0, 1]$, the semantics above is an extension to $\mathcal{ALC}$ of the FLTL (Fuzzy Linear-time Temporal Logic) semantics by Lamine and Kabanza (2000).

**Proposition** For all concepts $C$ and $D$, and for all time points $n$, the following properties hold:

$$(\Diamond C)^{\mathcal{I}}(n, x) = C^{I}(n, x) \oplus (\Diamond C)^{\mathcal{I}}(n + 1, x)$$
$$(\Box C)^{\mathcal{I}}(n, x) = C^{I}(n, x) \otimes (\Box C)^{\mathcal{I}}(n + 1, x)$$
$$(C \mathcal{U} D)^{\mathcal{I}}(n, x) = D^{I}(n, x) \oplus (C^{I}(n, x) \otimes (C \mathcal{U} D)^{\mathcal{I}}(n + 1, x))$$

# Temporal Weighted KBs: Example

A weighted $\mathcal{LC}$ knowledge base
$K = \langle \mathcal{T}_{strict}, \mathcal{T}_{Student}, \mathcal{T}_{Emp}, \mathcal{A} \rangle$

- $\mathcal{A} = \{Student(bob), Professor(ann), teaches(ann, c1), ..\}$;
- $\mathcal{T}_{strict}$ contains the strict inclusions:
  $Emp \sqsubseteq Adult$ $\quad Adult \sqsubseteq \exists has\_SSN.\top$
  $PhdStudent \sqsubseteq Student$
- $\mathcal{T}_{Student}$ contains the typicality inclusions:
  $(d_1)$ $\mathbf{T}(Student) \sqsubseteq Young,$ $\quad 90$
  $(d_2)$ $\mathbf{T}(Student) \sqsubseteq \exists has\_classes.\top,$ $\quad 80$
  $(d_3)$ $\mathbf{T}(Student) \sqsubseteq \exists hasScholarship.\top,$ $\quad -30$
  $(d_4)$ $\mathbf{T}(Student) \sqsubseteq \Diamond(Promoted \sqcup Rejected),$ $\quad 100$
- $\mathcal{T}_{Emp}$ .....

The *prototype description* for concept *Student*, etc.
A (multi-preferential) *closure construction* is needed!

# Many-valued temporal $\varphi$-coherent semantics for weighted KBs

Given $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$, and inclusions $(\mathbf{T}(C_i) \sqsubseteq D_{i,h} \, , w_h^i) \in \mathcal{T}_{C_i}$.
The *weight of x wrt $C_i$ at time point n* in $\mathcal{I}$.

$$W_{i,n}^{\mathcal{I}}(x) = \sum_{(\mathbf{T}(C_i) \sqsubseteq D_j, w_{ij}) \in \mathcal{D}} w_{ij} \; D_j^{\mathcal{I}}(n, x).$$

$W_{i,n}^{\mathcal{I}}$ should agree with the fuzzy interpretation of concepts in $\mathcal{I}$:

$\mathcal{I}$ is *faithful at n* if, for all $x, y \in \Delta^{\mathcal{I}}$,

$$x \prec_{C_i}^{n} y \quad \Rightarrow \quad W_{i,n}^{\mathcal{I}}(x) > W_{i,n}^{\mathcal{I}}(y)$$

$\mathcal{I}$ is *coherent at n* if, for all $x, y \in \Delta^{\mathcal{I}}$,

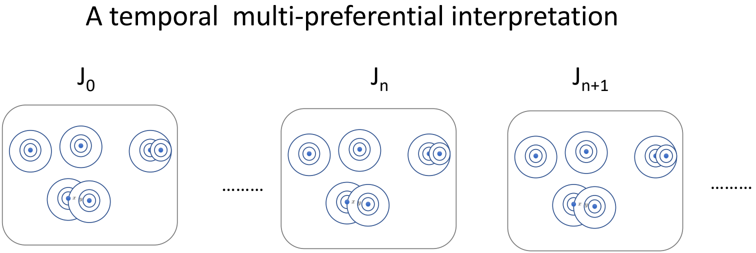
$$x \prec_{C_i}^{n} y \quad \text{iff} \quad W_{i,n}^{\mathcal{I}}(x) > W_{i,n}^{\mathcal{I}}(y)$$

Given a collection of monotonically non-decreasing functions $\varphi_i : \mathbb{R} \rightarrow \mathcal{S}$, one for each concept $C_i \in \mathcal{C}$:
- $\mathcal{I}$ is *$\varphi$-coherent at n* if, for all $x \in \Delta^{\mathcal{I}}$,

$$C_i^{\mathcal{I}}(n, x) = \varphi_i(W_{i,n}^{\mathcal{I}}(x))$$

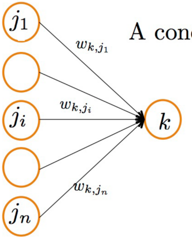
# $\varphi$-coherence for temporal weighted KBs

- We can see a many-valued temporal interpretation $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$ as a sequence $J^0, J^1, J^2, \ldots$ of many-valued preferential interpretations, as in [TPLP 2022, IJAR 2024].
- $J_n$ provides an interpretation of the KB at step $n$.



# MLPs as weighted KBs with typicality

- A Multi Layer neural network $\mathcal{N}$ can be represented by a weighted KB $K^{\mathcal{N}}$ (a set of weighted typicality inclusions):




A concept $C_i$ is associated to each neuron $I$:

$$\mathcal{T}_{C_k} \quad \mathbf{T}(\mathrm{C}_k) \sqsubseteq C_{j_1} \text{ with } w_{k,j_1}$$

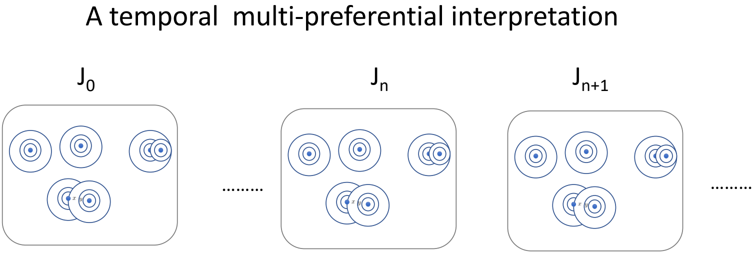
$$\cdots$$

$$\mathbf{T}(\mathrm{C}_k) \sqsubseteq C_{j_n} \text{ with } w_{k,j_n}$$

Typicality inclusions for unit k

- If the interpretation is *$\varphi$-coherent at n*, $J_n$ represents a *stationary state* of the network $\mathcal{N}$.

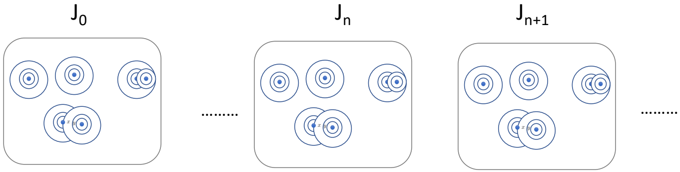
# Transient $\varphi$-coherence for weighted KBs




- a notion *transient $\varphi$-coherent of $\mathcal{I}$ at $n$* if, for all $x \in \Delta^{\mathcal{I}}$,
$$C_i^{\mathcal{I}}(n+1, x) = \varphi_i(W_{i,n}^{\mathcal{I}}(x))$$
- *transient $\varphi$-coherence* at all $n$ enforces the interpretations $J^0, J^1, J^2, \ldots$ to describe the dynamic evolution of the activity of units in the network
- An alternative: use temporal modalities in weighted typicality inclusions, e.g., to capture time delayed feedback connections.

# Property verification by model checking

A temporal multi-preferential interpretation




We may check, for instance:

$$\exists lives\_in.Town \sqcap Young \sqsubseteq \mathbf{T}(\Diamond \exists granted.Loan) \geq 0.8$$

# Conclusions and further work

- We have presented a temporal many-valued preferential extension of ALC, which provides a semantics for describing the evolution of the state of the world.
- The formalism allows capturing the *trajectories of the state* of a neural network;
- Future work includes:

  - extending *ASP encodings* to deal with model checking in temporal preferential interpretations (for the finitely-valued case);

  - studying the *decidability and complexity* of fragments of the logic

  - exploiting the formalism for *explainability*

# Related work

- On a different route, a preferential LTL with *defeasible temporal operators* has been studied in Chafik et al. [2020] and in Chafik's PhD Thesis [2022].
- The *decidability* of meaningful fragments of the logic has been proven, and *tableaux based proof methods* for such fragments have been developed [ChafikACV21] and in Anasse Chafik's PhD Thesis (2022).
    - *Our approach does not consider* defeasible temporal operators (nor *preferences over time points*). We have preferences are over the domain elements.
- A different approach for combining defeasibility in temporal DL formalism has been proposed in [LPNMR'22], by *combining a temporal action logic* based on *temporal answer sets* and *an $\mathcal{EL}$ ontology*.

Thank you!!!