Logic Tensor Networks

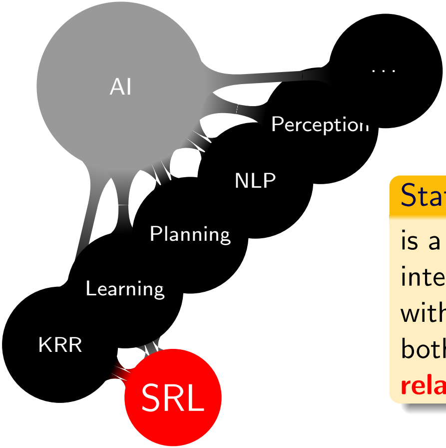
Luciano Serafini

Fondazione Bruno Kessler

AITP 2017

joint work with Artur d'Avila Garces - City Univ. London and Ivan Donadello, FBK

The SRL Mindmap



Statistical Relational Learning

is a subdiscipline of artificial intelligence that is concerned with domain models that exhibit both **uncertainty** and **complex relational structure**.

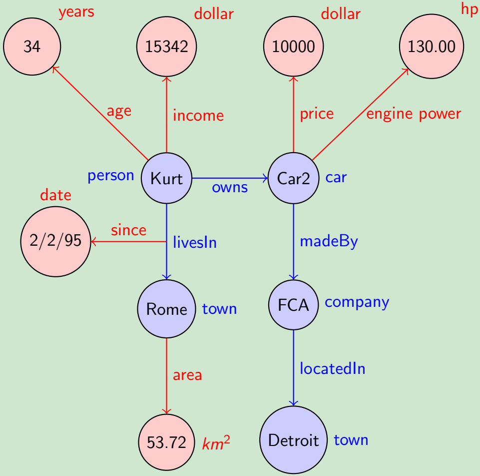
Hybrid domains

We are interested in Statistical Relational Learning over **hybrid domains**, i.e., domains that are characterized by the presence of

- structured data (categorical/semantic);
- continuous data (continuous features);

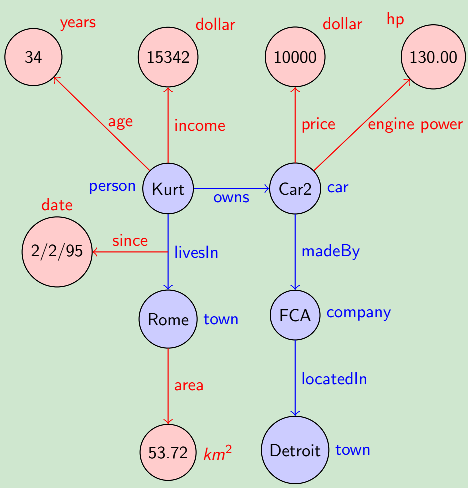
Hybrid domains

Example (SRL domain)



- Object Classification: Predicting the type of an object based on its relations and attributes;
- Reletion detenction: Predicting if two objects are connected by a relation, based on types and attributes of the participating objects;
- Regression: predicting the (distribution of) values of the attributies of an object, (a pair of related objects) based on the types and relations of the object(s) involved.

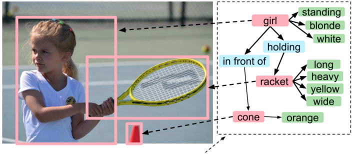
Example (SRL domain)



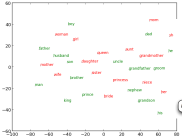
Real-world uncertain, structured and hybrid domains

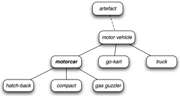
Robotics: a robot's location is a continuous values while the the types of the objects it encounters can be described by discrete set of classes

Semantic Image Interpretation: The visual features of a bounding box of a picture are continuous values, while the types of objects contained in a bounding box and the relations between them are taken from a discrete set



Natural Language Processing: The distributional semantics provide a vectorial (numerical) representation of the meaning of words, while WordNet associates to each word a set of synsets and a set of relations with other words which are finite and discrete





Language - to specify knowledge about models

Two sorted first order language: (abstract sort and numeric sort)

- Abstract constant symbols (*Ann*, *Bob*, *Cole*);
- Abstract function symbols (fatherOf(x));
- Abstract relation symbols (Person(x), Town(x), LivesIn(x,y);
- Numeric function symbols (age(x),area(y), livingInSince(x,y)
- Symbols for real numbers ($1, 0, \pi, \ldots$);
- Symbols for real functions $x + y$, $\sqrt{x}$, $\ldots$);
- Symbols for real relations ($x = y$, $x < y$).

COLOR CODE:

- [blue] denotes objects and relations of the domain structure;
- [red] denotes attributes and relations between attributes of the numeric part of the domain.

Domain description and queries

Example (Domain descritpion:)

company(A), company(B),
worksFor(Alice,A), worksFor(Ann,A),
worksFor(Bob,B),worksFor(Bill,B);
friends(Alice,Ann), friends(Bob,Bill),
$\neg$ friends(Ann,Bill)

Domain description and queries

Example (Domain descritpion:)

company(A), company(B),
worksFor(Alice,A), worksFor(Ann,A),
worksFor(Bob,B),worksFor(Bill,B);
friends(Alice,Ann), friends(Bob,Bill),
$\neg$ friends(Ann,Bill)
$salary(Alice) = 10.000$,
$salary(Ann) \leq 12.000$,
$salary(Bob) = 30.000$,
$salary(Bill) \geq 27.000$,
$9.000 \leq Salary(Chris) \leq 11.000$

Domain description and queries

FONDAZIONE BRUNO KESSLER

Example (Domain descritpion:)

company(A), company(B),
worksFor(Alice,A), worksFor(Ann,A),
worksFor(Bob,B),worksFor(Bill,B);
friends(Alice,Ann), friends(Bob,Bill),
¬ friends(Ann,Bill)
$salary(Alice) = 10.000$,
$salary(Ann) \leq 12.000$,
$salary(Bob) = 30.000$,
$salary(Bill) \geq 27.000$,
$9.000 \leq Salary(Chris) \leq 11.000$
$\forall x.worksFor(x, A) \leftrightarrow \neg worksFor(x, B)$
$\forall xy.friends(x, y) \leftrightarrow friends(y, x)$
$\forall xy, worksFor(x, y) \rightarrow salary(x) > 3.000$
$\forall x \exists y.friends(x, y)$

Domain description and queries

FONDAZIONE BRUNO KESSLER

Example (Domain descritpion:)

company(A), company(B),
worksFor(Alice,A), worksFor(Ann,A),
worksFor(Bob,B),worksFor(Bill,B);
friends(Alice,Ann), friends(Bob,Bill),
¬ friends(Ann,Bill)
$salary(Alice) = 10.000$,
$salary(Ann) \leq 12.000$,
$salary(Bob) = 30.000$,
$salary(Bill) \geq 27.000$,
$9.000 \leq Salary(Chris) \leq 11.000$
$\forall x.worksFor(x, A) \leftrightarrow \neg worksFor(x, B)$
$\forall xy.friends(x, y) \leftrightarrow friends(y, x)$
$\forall xy, worksFor(x, y) \rightarrow salary(x) > 3.000$
$\forall x \exists y.friends(x, y)$

Example (Queries)

- ? $worksfor(Chris, B)$
- ? $?x{:}friends(Chris, ?x)$
- ? $?salary(Bill)$
- ? $?salary(x) : x = friendOf(Ann)$
- ? $?worksfor(x, z) \wedge worksfor(z, z) \rightarrow friends(x, y)$
- ? $?salary(x) > 15.000 \rightarrow worksfor(x, A)$

Fuzzy semantics for LTN

Let $\mathcal{L}$ contains the set $r_1, \ldots, r_n$ unary real functions (like age, salary, $\ldots$)

Fuzzy Semantics

An interpretation $\mathcal{G}$ of $\mathcal{L}$, called grounding, is a real function:

- $\mathcal{G}(c) \in \mathbb{R}^n$ for every constant c;
- $\mathcal{G}(f) \in \mathbb{R}^{n \cdot m} \longrightarrow \mathbb{R}^n$ for every m-ary abstract function f;
- $\mathcal{G}(P) \in \mathbb{R}^{n \cdot m} \longrightarrow [0, 1]$ for every m-ary abstract predic symbol P;

Given a grounding $\mathcal{G}$ the semantics of closed terms and atomic formulas is defined as follows:

$$\mathcal{G}(f(t_1, \ldots, t_m)) = \mathcal{G}(f)(\mathcal{G}(t_1), \ldots, \mathcal{G}(t_m))$$
$$\mathcal{G}(P(t_1, \ldots, t_m)) = \mathcal{G}(P)(\mathcal{G}(t_1), \ldots, \mathcal{G}(t_m))$$

Grounding as parametrized neural network = Logic Tensor Network (LTN)

FONDAZIONE BRUNO KESSLER

- Grounding of constant symbols: Real vectors

$$\mathcal{G}(c) \in \mathbb{R}^n$$

For every i $\mathcal{G}_i(c) = r_i(c)$ if $r_i(c)$ is known, otherwise $\mathcal{G}_i(c)$ is a **parameter** of the LTN.

- Grounding of functional symbols: Two layer feed-forward neural network with $m \cdot n$ imput nodes and n output nodes.

$$\mathcal{G}(f)(\mathbf{v}) = M_f \sigma(N_f \mathbf{v})$$

$M_f \in R^{mn \times n}$ and $N_f \in R^{mn \times mn}$ are **parameters** of the LTN;

- Grounding of predicate symbols: Tensor quadratic network

$$\mathcal{G}(P)(\mathbf{v}) = \sigma\left(u_P^\intercal \tanh\left(\mathbf{v}^\intercal W_P^{[1:k]}\mathbf{v} + V_P\mathbf{v} + b_P\right)\right)$$

$w_P \in \mathbb{R}^{k \times mn \times mn}$, $V_P \in \mathbb{R}^{k \times mn}$, $b_P \in \mathbb{R}^k$, and $u_P \in \mathbb{R}^k$ are **parameters** of the LTN.

Grounding as parametrized neural network = Logic Tensor Network (LTN)

- Grounding of real functions are the real functions themselves. For instance:

$$\mathcal{G}(+)(\mathbf{v},\mathbf{u}) = \mathbf{v}+\mathbf{u}$$

- Grounding of real relations are the real relations themselves. For instance:

$$\mathcal{G}(=)(\mathbf{v},\mathbf{u}) = \begin{cases} 1 & \text{if } \mathbf{v}=\mathbf{u} \\ 0 & \text{Otherwise} \end{cases}$$

or some soft version

$$\mathcal{G}(=)(\mathbf{v},\mathbf{u}) = \frac{\mathbf{v}\cdot\mathbf{u}}{||\mathbf{v}||\ ||\mathbf{u}||}$$

Domain description and queries

FONDAZIONE BRUNO KESSLER

Example (Domain descritpion:)

company(A), company(B),
worksFor(Alice,A), worksFor(Ann,A),
worksFor(Bob,B),worksFor(Bill,B);
friends(Alice,Ann), friends(Bob,Bill),
¬ friends(Ann,Bill)
$salary(Alice) = 10.000$,
$salary(Ann) \leq 12.000$,
$salary(Bob) = 30.000$,
$salary(Bill) \geq 27.000$,

Example (Queries)

- ? $worksfor(Chris, B)$
- ? ?x:$friends(Chris, ?x)$
- ? $?salary(Bill)$
- ? $?salary(x) : x = friendOf(Ann)$

Domain description and queries

Example (Domain descritpion:)

company(A), company(B),
worksFor(Alice,A), worksFor(Ann,A),
worksFor(Bob,B),worksFor(Bill,B);
friends(Alice,Ann), friends(Bob,Bill),
¬ friends(Ann,Bill)
$salary(Alice) = 10.000$,
$salary(Ann) \leq 12.000$,
$salary(Bob) = 30.000$,
$salary(Bill) \geq 27.000$,
$9.000 \leq Salary(Chris) \leq 11.000$
$\forall x.worksFor(x, A) \leftrightarrow \neg worksFor(x, B)$
$\forall xy.friends(x, y) \leftrightarrow friends(y, x)$
$\forall xy, worksFor(x, y) \rightarrow salary(x) > 3.000$
$\forall x \exists y.friends(x, y)$

Example (Queries)

- ? $worksfor(Chris, B)$
- ? $?x{:}friends(Chris, ?x)$
- ? $?salary(Bill)$
- ? $?salary(x) : x = friendOf(Ann)$
- ? $?worksfor(x, z) \wedge worksfor(z, z) \rightarrow friends(x, y)$
- ? $?salary(x) > 15.000 \rightarrow worksfor(x, A)$

Fuzzy semantics for propositional connectives

- In fuzzy semantics atoms are assigned with some truth value in real interval [0,1]
- connectives have functional semantics. e.g., a binary connective $\circ$ must be interpreted in a function $f_\circ : [0,1]^2 \rightarrow [0,1]$.
- Truth values are ordeblue, i.e., if $x > y$, then x is a stronger truth than y
- Generalization of classical propositional logic:

  0 corresponds to FALSE and
  1 corresponds to TRUE

T-norm

Definition (t-norm)

A t-norm is a binary operation $* : [0,1]^2 \to [0,1]$ satisfying the following conditions:

- **Commutativity:** $x * y = y * x$
- **Associativity:** $x * (y * z) = (x * y) * z$
- **Monotonicity:** $x \leq y \quad \to z * x \leq z * y$
- **Zero and One:** $0 * x = 0$ and $1 * x = x$

A t-norm $*$ is continuous if the function $* : [0,1]^2 \to [0,1]$ is a continuous function in the usual sense.

Fuzzy semantics for connectives

T-norm, T-conorm, residual, and precomplement

T-norm	$\wedge$	$a \otimes b$	$=$	Continuous T-norm
T-conorm	$\vee$	$a \oplus b$	$=$	$1 - \otimes(1-a, 1-b)$
residual	$\rightarrow$	$a \Rightarrow b$	$=$	$\begin{cases} \text{if } a > b & sup(\{z \mid z \otimes a \leq b\}) \\ \text{if } a \leq b & 1 \end{cases}$
precomplement	$\neg$	$\ominus a$	$=$	$a \Rightarrow 0 = \max(z \mid z \otimes a = 0\})$

Fuzzy semantics for connectives

FONDAZIONE BRUNO KESSLER

Lukasiewicz T-norm, T-conorm, residual, and precomplement

T-norm	$\wedge$	$a \otimes b$	$=$	$\max(0, a+b-1)$
T-conorm	$\vee$	$a \oplus b$	$=$	$\min(1, a+b)$
residual	$\rightarrow$	$a \Rightarrow b$	$=$	$\begin{cases} \text{if } a > b & 1-a+b \\ \text{if } a \leq b & 1 \end{cases}$
precomplement	$\neg$	$\ominus a$	$=$	$1-a$

Fuzzy semantics for connectives

Gödel T-norm, T-conorm, residual, and precomplement

T-norm	$\wedge$	$a \otimes b$	$=$	$\min(a, b)$
T-conorm	$\vee$	$a \oplus b$	$=$	$\max(a, b)$
residual	$\rightarrow$	$a \Rightarrow b$	$=$	$\begin{cases} \text{if } a > b & b \\ \text{if } a \leq b & 1 \end{cases}$
precomplement	$\neg$	$\ominus a$	$=$	$\begin{cases} \text{if } a = 0 & 1 \\ \text{if } a > 0 & 0 \end{cases}$

Fuzzy semantics for connectives

Product T-norm, T-conorm, residual, and precomplement

T-norm	$\wedge$	$a \otimes b$	$=$	$a \cdot b$ (scalar product)
T-conorm	$\vee$	$a \oplus b$	$=$	$a + b - a \cdot b$
residual	$\rightarrow$	$a \Rightarrow b$	$=$	$\begin{cases} \text{if } a > b & b/a \\ \text{if } a \leq b & 1 \end{cases}$
precomplement	$\neg$	$\ominus a$	$=$	$\begin{cases} \text{if } a = 0 & 1 \\ \text{if } a > 0 & 0 \end{cases}$

fuzzy semantics for quantifiers

$\forall x P(x)$ in fuzzy logic is consideblue as an infinite conjunction $P(a_1) \wedge P(a_2) \wedge P(a_3) \wedge \dots ,.$

Fuzzy semantics for $\forall$

$$\forall x a(x) = \min_{c \in C} a(c)$$

This semantics is not adeguate for our purpose.

Example

$Bird(tweety) = 1.0$ and $Fly(tweety) = 0.0$ implies that $\forall x(Bird(x) \rightarrow Fly(x)) = 0.0$.

Instead we want to have something like, if the 90% of the birds fly then the truth value of $\forall x(Bird(x) \rightarrow Fly(x))$ should be 0.9.

Aggregational semantics for Quantifiers

Aggregation operator: $Agg : \bigcup_{n \geq 1} [0,1]^n \to [0,1]$

- **Bounded:**

$$\min(x_1, \ldots, x_n) \leq Agg(x_1, \ldots, x_n) \leq \max(x_1, \ldots, x_n)$$

- **Strict Monotonicity**

$$x < x' \;\Rightarrow\; Agg(\ldots, x, \ldots,) < Agg(\ldots, x', \ldots,)$$

- **Commutativity:**

$$Agg(\ldots, x, \ldots, y, \ldots,) = Agg(\ldots, y, \ldots, x, \ldots,)$$

- **Convergent:**

$$\lim_{n \to \infty} Agg(x_1, \ldots, x_n) \in [0,1]$$

Examples of aggregation operators

- **Min**

$$\min_{i=1}^{n}(x_i)$$

- **Aritmetic mean**

$$\frac{1}{n}\sum_{i=1}^{n} x_i$$

- **Geometric mean**

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}$$

- **Harmonic mean**

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i^{-1}\right)^{-1}$$

- **generalized mean** for $k \leq 1$

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i^k\right)^{\frac{1}{k}}$$

Constructive semantics for Existential quantifier

- LTN interprets existential quantifiers constructively via Skolemization.
- Every formula $\forall x_1, \ldots, x_n \exists y \phi(x_1, \ldots, x_n, y)$ is rewritten as $\forall x_1, \ldots, x_m \phi(x_1, \ldots, x_n, f(x_1, \ldots, x_m))$,
- by introducing a new m-ary function symbol f,

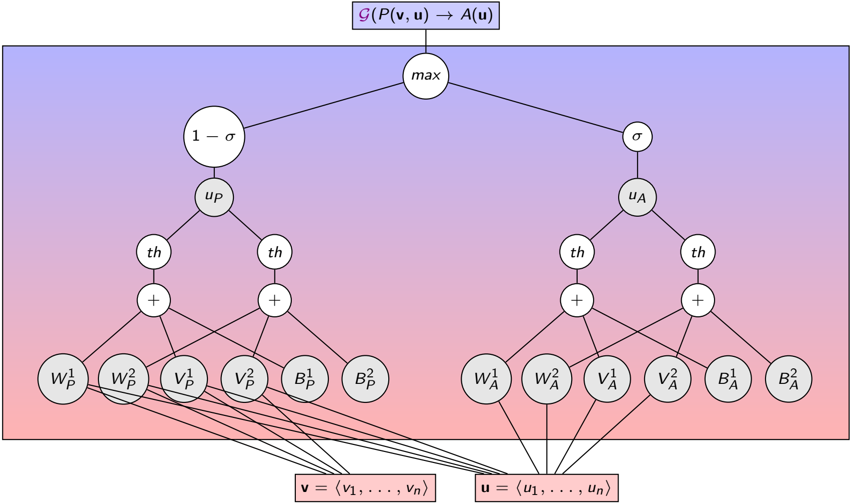
Example

$$\forall x.(cat(x) \rightarrow \exists y.partof(y, x) \wedge tail(y))$$

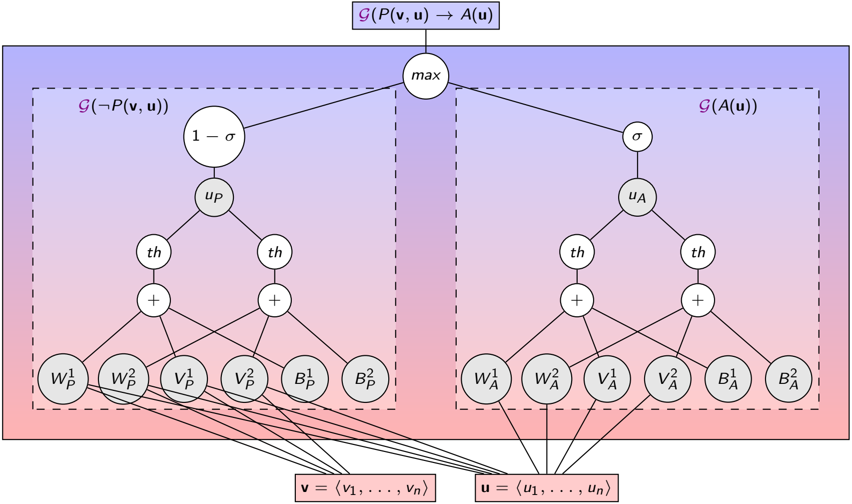
is transformed in

$$\forall x(cat(x) \rightarrow partOf(tailOf(x), x) \wedge tail(tailOf(x)))$$

Grounding = relation between logical symbols and data



Grounding = relation between logical symbols and data



Parameter learning = best satisfiability

Given a FOL theory $\mathtt{K}$ the **best satisfiability problem** as the problem of finding a grounding $\mathcal{G}^*$ for K that maximizes the truth values of the formulas entailed by $\mathtt{K}$, i.e.,

$$\mathcal{G}^* = \operatorname*{argmax}_{\mathcal{G}} \left(\min_{\mathtt{K} \models \phi} \mathcal{G}(\phi) \right)$$

Since $\mathcal{G}$ in LTN is defined by the set of parameters Θ of the LTN, then the problems become $\mathcal{G}^* = LTN(K, \Theta^*)$

$$\Theta^* = \operatorname*{argmax}_{\Theta} \left(\min_{\mathtt{K} \models \phi} LTN(K, \Theta)(\phi) \right)$$

Learning from model description and answering queries

K

company(A), company(B),
worksFor(Alice,A), worksFor(Ann,A),
worksFor(Bob,B),worksFor(Bill,B);
cfriends(Alice,Ann), friends(Bob,Bill),
¬ friends(Ann,Bill)
$salary(Alice) = 10.000$,
$salary(Ann) \leq 12.000$,
$salary(Bob) = 30.000$,
$salary(Bill) \geq 27.000$,
$9.000 \leq Salary(Chris) \leq 11.000$
$\forall x.worksFor(x, A) \leftrightarrow \neg worksFor(x, B)$
$\forall xy.friends(x, y) \leftrightarrow friends(y, x)$
$\forall xy, worksFor(x, y) \rightarrow salary(x) > 3.000$
c$\forall x \exists y.friends(x, y)$

Learning from model description and answering queries

$$\Theta^* = \operatorname{argmax}_\Theta \left(\min_{\mathtt{K} \models \phi} LTN(K, \Theta)(\phi) \right)$$

K

company(A), company(B),
worksFor(Alice,A), worksFor(Ann,A),
worksFor(Bob,B),worksFor(Bill,B);
cfriends(Alice,Ann), friends(Bob,Bill),
$\neg$ friends(Ann,Bill)
$salary(Alice) = 10.000$,
$salary(Ann) \leq 12.000$,
$salary(Bob) = 30.000$,
$salary(Bill) \geq 27.000$,
$9.000 \leq Salary(Chris) \leq 11.000$
$\forall x.worksFor(x, A) \leftrightarrow \neg worksFor(x, B)$
$\forall xy.friends(x, y) \leftrightarrow friends(y, x)$
$\forall xy, worksFor(x, y) \rightarrow salary(x) > 3.000$
c$\forall x \exists y.friends(x, y)$

Learning from model description and answering queries

$$\Theta^* = \mathrm{argmax}_\Theta \left(\mathsf{min}_{K \models \phi} LTN(K, \Theta)(\phi)\right)$$

K

company(A), company

worksFor(Alice,A), wor

worksFor(Bob,B),works

cfriends(Alice,Ann), fri

$\neg$ friends(Ann,Bill)

$salary(Alice) = 10.000$

$salary(Ann) \leq 12.000,$

$salary(Bob) = 30.000,$

$salary(Bill) \geq 27.000,$

$9.000 \leq Salary(Chris) \leq 11.000$

$\forall x.worksFor(x, A) \leftrightarrow \neg worksFor(x, B)$

$\forall xy.friends(x, y) \leftrightarrow friends(y, x)$

$\forall xy, worksFor(x, y) \rightarrow salary(x) > 3.000$

c$\forall x \exists y.friends(x, y)$

Q

$LTN_{K,\Theta^*}(worksfor(Chris, B))$

$LTN_{K,\Theta^*}(friends(Chris, x)|x = Alice, Ann, \ldots\}$

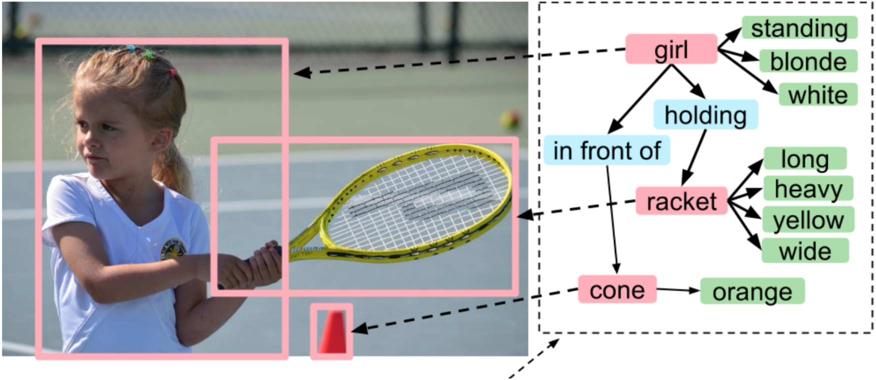
$LTN_{K,\Theta^*}(salary(Bill))$

$LTN_{K,\Theta^*}(salary(friendOf(Ann)))$

$LTN_{K,\Theta^*}(\forall xy.worksfor(x, z) \wedge worksfor(z, z) \rightarrow friends(x, y))$

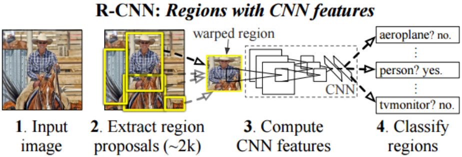
$LTN_{K,\Theta^*}(\forall x.salary(x) > 15.000 \rightarrow worksfor(x, A))$

Application of LTN to Semantic Image Interpretation



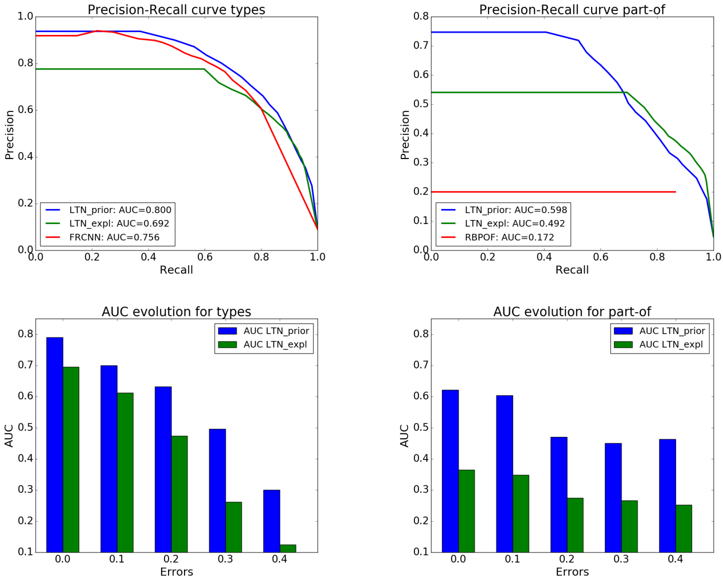
Semantic Image interpretation pipeline

- We apply te state-of-the-art object detector (Fast-RCNN) to extract bounding boxes around objects associated with semantic features.



- We train an LTN with the following theory
  - positive/negative examples for object classes (from training set) $weel(bb1),\ car(bb2),\ \neg horse(bb2), \neg person(bb4)$
  - positive/negative examples for relations (we focus on parthood relation). $partOf(bb1, bb2),\ \neg partOf(bb2, bb3), \dots,$
  - general axioms about parthood relation: $\forall x.car(x) \wedge partof(y, y) \rightarrow wheeel(y) \vee mirror(y) \vee door(y) \vee \dots,)$
  - Axioms for Fast-RCNN proposed classification of bounding boxes $rcnn_{car}(bb1) = .8,\ rcnn_{horse}(bb1) = .01,\ rcnn_{wheel}(bb2) = .75,\ \dots,$

LTN for SII results



Conclusions

Thanks

Thanks for your attention