LEARNING AND REASONING WITH CONSTRAINTS

Marco Gori
University of Siena (Italy)

OUTLINE

- Environment, machines, and constraints
- Bridging logic and real-valued constraints
- Constraint consistency and parsimony
- Data structure, constraints, and GNN
- Learning of constraints

ENVIRONMENTS, MACHINES, AND CONSTRAINTS

MSc DSAI 2021

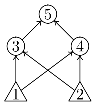
Supervised Learning

architectural and environmental constraints

environmental constraints

$$\mathcal{L} = \{((0,0),0),\ ((0,1),1),\ ((1,0),1),\ ((1,1),0)\} =$$

"soft"



1 2

$$\mathsf{y}((0,1),(1,0))$$

$$\neg\mathsf{y}((0,0),(1,1)).$$

4 3

targets $x_{15} = 1,\ x_{25} = 1,\ x_{35} = 0,\ x_{45} = 0$

architectural constraints

"hard"

$$\begin{aligned} x_{\kappa 3} - \sigma(w_{31}x_{\kappa 1} + w_{32}x_{\kappa 2} + b_3) &= 0 \\ x_{\kappa 4} - \sigma(w_{41}x_{\kappa 1} + w_{42}x_{\kappa 2} + b_4) &= 0 \qquad \kappa = 1,2,3,4 \\ x_{\kappa 5} - \sigma(w_{53}x_{\kappa 3} + w_{54}x_{\kappa 4} + b_4) &= 0 \end{aligned}$$

MSc DSAI 2021

Enforcing Consistencies

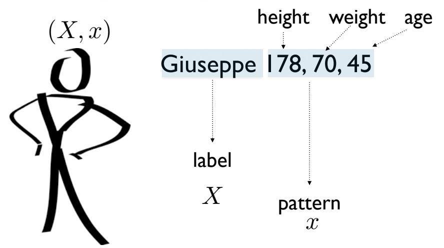
$$f_{\omega h} : \mathcal{W} \to \mathscr{H} : \ h \to \omega(h),$$
$$f_{ah} : \mathcal{W} \to \mathcal{A} : \ h \to a(h),$$
$$f_{\omega a} : \mathcal{A} \to \mathcal{W} : \ a \to \omega(a),$$

$$\begin{tikzcd} & a \arrow[d, "f_{\omega a}"] \\ h \arrow[ur, "f_{ah}"] \arrow[r, "f_{\omega h}"'] & \omega \end{tikzcd} \qquad f_{\omega h}(h) = f_{\omega a} \circ f_{ah}(h).$$

This functional equation is imposing the circulation of coherence. Since the functions are linear, this constraint can be converted to $w_{\omega h}h + b_{\omega h} = w_{\omega a}w_{ah}h + (w_{ah}b_{ah} + b_{\omega a})$. The equivalence $\forall h \in \mathbb{R}^+$ yields

$$w_{\omega a}w_{ah} - w_{\omega h} = 0,$$
$$w_{ah}b_{ah} + b_{\omega a} - b_{\omega h} = 0.$$

Patterns, Labels, and Individuals



What about learning and inference with individuals?

Inference in Formal Logic

only labels are involved!

```
Domain(label="People")
Individual(label="Marco", "People")
Individual(label="Giuseppe", "People")
Individual(label="Michelangelo", "People")
Individual(label="Francesco", "People")
Individual(label="Franco", "People")
Individual(label="Andrea", "People")
```

```
Predicate(label="fatherOf", ("People", "People"))
Predicate(label="grandFatherOf", ("People", "People"))
Predicate(label="eq", ("People", "People"), function=eq)
```

```
Constraint("fatherOf(Marco, Giuseppe)")
Constraint("fatherOf(Giuseppe, Michelangelo)")
Constraint("fatherOf(Giuseppe, Francesco)")
Constraint("fatherOf(Franco, Andrea)")
```

```
Constraint("forall x: not fatherOf(x,x)")
Constraint("forall x: not grandFatherOf(x,x)")
```


Inference in Formal Logic

```
Constraint("forall x: forall y: fatherOf(x,y) -> not fatherOf(y,x)")
Constraint("forall x: forall y: grandFatherOf(x,y)
-> not grandFatherOf(y,x)")
Constraint("forall x: forall y: fatherOf(x,y) -> not grandFatherOf(x,y)")
Constraint("forall x: forall y: grandFatherOf(x,y) -> not fatherOf(x,y)")
```

```
Constraint("forall x: forall y: forall z: fatherOf(x,z) and fatherOf(z,y) ->
    grandFatherOf(x,y)")
Constraint("forall x: forall y: forall z: (fatherOf(x,y) and not eq(x,z)) ->
    not fatherOf(z,y)")
```

Inference in Formal Logic

```
grandFatherOf("Marco", "Michelangelo")
grandFatherOf("Marco", "Francesco")
```

```
Constraint("forall x: forall y: forall z: grandFatherOf(x,z) and
    fatherOf(y,z) -> fatherOf(x,y)")
```


Full Inference on Individuals (X, x)

from formal logic

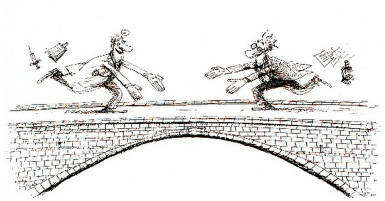
from neural nets

consistency constraints

$(\mathtt{age}_x, \mathtt{weight}_x, \mathtt{height}_x, \mathtt{age}_y, \mathtt{weight}_y, \mathtt{height}_y)$

Complexity issues: the inference in the environment avoids massive exploration of the Boolean hypercube

BRIDGING LOGIC AND REAL-VALUED CONSTRAINTS



learning   relations and logic

"There are finer fish in the sea that have ever been caught," Irish proverb

Logic by Real Numbers

$$\forall x \quad a(x) \wedge b(x) \Rightarrow c(x)$$

$$\neg(a(x) \wedge b(x)) \vee c(x))$$

$$\neg\neg(\neg(a(x) \wedge b(x)) \wedge c(x))$$

$$\neg(a(x) \wedge b(x) \wedge \neg c(x))$$

p-norm

$$1 - \boxed{f_a(x) \cdot f_b(x) \cdot (1 - f_c(x))} = 1$$

$$f_a(x) f_b(x)(1 - f_c(x)) = 0$$

general form $\forall x \quad \Phi(f(x)) = 0 \longrightarrow \Phi(x, f(x)) = 0$

MSc DSAI 2021

Logic by Real Numbers (con't)

$$\forall x \quad a(x) \wedge b(x) \Rightarrow c(x)$$

$$\neg(a(x) \wedge b(x) \wedge \neg c(x))$$

Gödel T-norm

$$1 - min\,\{f_a(x), f_b(x), 1 - f_c(x)\} = 1$$

$$min\,\{f_a(x), f_b(x), 1 - f_c(x)\} = 0$$

CONSTRAINT CONSISTENCY AND PARSIMONY

"the simplest solution" compatible with the constraints

We use the Lagrangian optimization framework

Constraint-Based Learning

$$\phi(f(x)) = 0$$

given task to be learned

everything revolves around this compositional structure
Gori et al, Neural Computation 2015

A New Communication Protocol

data + constraints

$\forall x \;\; \Phi(x, f(x)) = 0$ from constraints to

consistency check

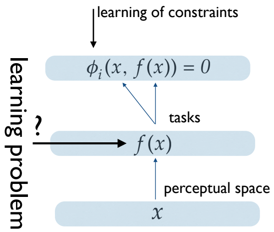
$$\sum_{\kappa \in U} \phi^2(x_\kappa, f(x_\kappa))$$ loss functions

unsupervised data!

A New Communication Protocol

data + constraints

- ~~Supervised~~
- ~~Unsupervised~~
- ~~Semi-supervised~~



Learning Under Constraint Consistency

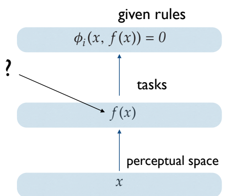
$$\mathtt{hair}(\mathtt{x}) \Rightarrow \mathtt{mammal}(\mathtt{x})$$
$$\mathtt{mammal}(\mathtt{x}) \wedge \mathtt{hoofs}(\mathtt{x}) \Rightarrow \mathtt{ungulate}(\mathtt{x})$$
$$\mathtt{ungulate}(\mathtt{x}) \wedge \mathtt{white}(\mathtt{x}) \wedge \mathtt{blackstripes}(\mathtt{x}) \Rightarrow \mathtt{zebra}(\mathtt{x}).$$

$$f_{\mathtt{hair}}(x)(1 - f_{\mathtt{mammal}}(x)) = 0$$
$$f_{\mathtt{mammal}}(x) f_{\mathtt{hoofs}}(x)(1 - f_{\mathtt{ungulate}}(x)) = 0$$
$$f_{\mathtt{ungulate}}(x) f_{\mathtt{white}}(x) f_{\mathtt{blackstripes}}(x)(1 - f_{\mathtt{zebra}}(x)) = 0.$$

penalty functions

perceptual space

x



The Marriage of Parsimony Principle and Constraints

Constraints turn out to be loss functions

keep these loss functions as small as possible

$$f_{\mathtt{hair}}(x)(1 - f_{\mathtt{mammal}}(x)) = 0$$
$$f_{\mathtt{mammal}}(x) f_{\mathtt{hoofs}}(x)(1 - f_{\mathtt{ungulate}}(x)) = 0$$
$$f_{\mathtt{ungulate}}(x) f_{\mathtt{white}}(x) f_{\mathtt{blackstripes}}(x)(1 - f_{\mathtt{zebra}}(x)) = 0.$$

penalty functions

perceptual space

x

Parsimony Principle

$\| f \|$

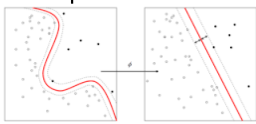
f: $f_{\mathtt{hair}}$, $f_{\mathtt{hoofs}}$, $f_{\mathtt{mammal}}$, $f_{\mathtt{ungulate}}$, $f_{\mathtt{white}}$, $f_{\mathtt{blackstripes}}$, $f_{\mathtt{zebra}}$

How to Represent the Tasks?

f ?

Primal space

Dual Space

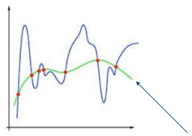


Kernel Machines

…

Parsimony Principle
smoothness of the tasks

$$\| f \|^2 := b_0 \int_{\mathcal{X}} f^2(x)dx + b_1 \int_{\mathcal{X}} \left(\frac{df}{dx}\right)^2 dx$$



Occam's razor, lex parsimoniae

Semi-norm in Sobolev Spaces

$$P = \sum_{|\alpha|<m} a_\alpha D_x^\alpha = \sum_{|\alpha|<m} a_\alpha \left(\frac{\partial}{\partial x_1} + \ldots + \frac{\partial}{\partial x_d}\right)^\alpha$$

∞ $\quad a_\alpha \in C^\infty$

under proper boundary conditions ...

$$P = \sum_{h=0}^{m} a_h \sum_{|\alpha|=h} \frac{h!}{\alpha!} \left(\frac{\partial}{\partial x}\right)^\alpha \qquad P^\star = \sum_{h=0}^{m} (-1)^h a_h \sum_{|\alpha|=h} \frac{h!}{\alpha!} \left(\frac{\partial}{\partial x}\right)^\alpha$$

Given P and $\gamma_i > 0, \ldots, i = 1, \ldots, n$

$$E(f) = \| f \|_{P,\gamma} = \sum_{j=1}^{n} \gamma_j < Pf_j, Pf_j > = \sum_{j=1}^{n} \gamma_j < f_j, P^\star P f_j > = \sum_{j=1}^{n} \gamma_j < f_j, L f_j >$$

Parsimony Principle

inference in the environment!

$\mathcal{F}_\phi$ admissible w.r.t the collection of constraints $\mathcal{C}_\phi$

$$f^\star = argmin_{f \in \mathcal{F}_\phi} \parallel f \parallel_{P,\gamma}$$

strictly (hard)

partially (soft)

check of a "new" constraint

$$\forall x \quad \phi(x, f^\star(x), Df^\star(x)) = 0 \;?$$

Representation of the Solution

hard constraints

$$\forall x \in \mathcal{X}_i \subset X : \phi_i(x, f(x)) = 0, \; i \in \mathbb{N}_m \qquad \frac{D(\phi_1, \ldots, \phi_m)}{D(f_1, \ldots, f_m)} \neq 0$$

$$\mathcal{L}(f) = \parallel f \parallel^2_{P,\gamma} + \sum_{i=1}^{m} \int_{\mathcal{X}} \lambda_i(x) \cdot \phi_i(x, f(x)) dx$$ Lagrangian approach

$$Lf(x) + \sum_{i=1}^{m} \lambda_i(x) \cdot \nabla_f \phi_i(x, f(x)) = 0$$ Euler-Lagrange equations

$Lg = \delta$ Green function

$$\omega_i(\cdot) = -\lambda_i(\cdot) \nabla_f \phi_i(\cdot, f^\star(\cdot))$$ reaction of the constraint

support constraints

Fredholm eq. (II kind)
"merging of two ideas ..."

$$f^\star(\cdot) = \sum_{i=1}^{m} g(\cdot) \otimes \omega_i(f^\star(\cdot))$$

Lagrange Multipliers and Probability Density

hard constraints

$$\forall x \in \mathcal{X}_i \subset X : \phi_i(x, f(x)) = 0, \; i \in I\!N_m$$

$$\mathcal{L}(f) = \| f \|^2_{P,\gamma} + \sum_{i=1}^{m} \int_{\mathcal{X}} \lambda_i(x) \check{\phi}_i(x, f(x)) dx$$

soft constraints

$$\mathcal{L}(f) = \| f \|^2_{P,\gamma} + C \sum_{i=1}^{m} \int_{\mathcal{X}} p_i(x) \check{\phi}_i(x, f(x)) dx$$

Case studies

- Supervised learning
- Unsupervised, semi-supervised learning
- Learning and inference in the environment
- Full inference (formal + environment+based)
- Pattern generation
- …

Supervised learning:
Lagrangian formulation

minimize $\frac{1}{2}\sum_{i\in O}\sum_{j\in H_o} w_{ij}^2 + \sum_{\kappa=1}^{\ell}\sum_{j\in H}\mu_{\kappa j}|x_{\kappa j}|$

subject to $x_{\kappa i} - \sigma\big(\sum_{j\in \mathrm{pa}(i)} w_{ij}x_{\kappa j}\big) = 0, \quad i \in H \cup O, \quad \kappa = 1,\dots,\ell,$
$1 - x_{\kappa i}y_{\kappa i} \le 0 \quad i \in O, \quad \kappa = 1,\dots,\ell$

λ

$$L(w,x,\alpha,\beta) = \frac{1}{2}\sum_{i\in O}\sum_{j\in H_o} w_{ij}^2 + \sum_{\kappa=1}^{\ell}\sum_{m}\Bigg(\lambda_{\kappa m}|x_{\kappa m}|\,[m\in H]$$
$$+\,\alpha_{\kappa m}\bigg(x_{\kappa m} - \sigma\Big(\sum_{r\in \mathrm{pa}(m)} w_{mr}x_{\kappa r}\Big)\bigg)[m\in H\cup O]$$
$$+\sum_{i\in O}\beta_{\kappa i}\big(1 - x_{\kappa i}y_{\kappa i}\big)_{+}\Bigg),$$

architectural constraints

"Saddle moves": gradient descent/ascent

A more biologically plausibile solution than Backpropagation

saddle points of the Lagrangian

$$w_{ij} \leftarrow w_{ij} - \eta_w \partial_{w_{ij}} L$$
$$x_{\kappa i} \leftarrow x_{\kappa i} - \eta_x \partial_{x_{\kappa i}} L$$

learning (gradient descent)

$$\lambda_{\kappa i} \leftarrow \lambda_{\kappa i} + \eta_\lambda \partial_{\lambda_{\kappa i}} L$$

focus of attention (gradient ascent)

related to BP delta error!

$$g_{\kappa i} = x_{\kappa i} - \sigma\left(\sum_{j\in pa(i)} w_{ij}x_{\kappa j}\right) = 0$$

saddle points of the Lagrangian

Lagrangian multipliers, straw and support neurons!

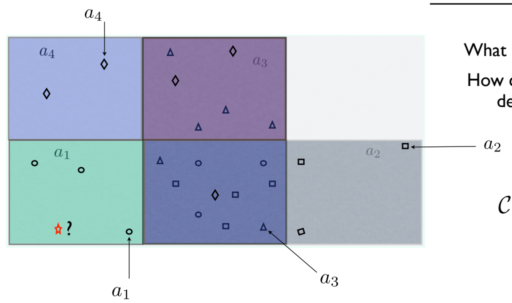
Network growing and constraint selection …

Learning and Inference in the Environment

$A = \{(x_1, x_2) \in R^2 : \; 0 \leq x_1 < 2, \; 0 \leq x_2 \leq 1\}$
$B = \{(x_1, x_2) \in R^2 : \; 1 \leq x_1 < 3, \; 0 \leq x_2 \leq 1\}$
$C = \{(x_1, x_2) \in R^2 : \; 1 \leq x_1 < 2, \; 0 \leq x_2 \leq 2\}$
$D = C \cup \{(x_1, x_2) \in R^2 : 0 \leq x_1 \leq 1, \; 1 \leq x_2 \leq 2\}$

"Knowledge Base"

$a_1(x) \wedge a_2(x) \Rightarrow a_3(x)$
$a_3(x) \Rightarrow a_4(x)$
$a_1(x) \vee a_2(x) \vee a_3(x)$

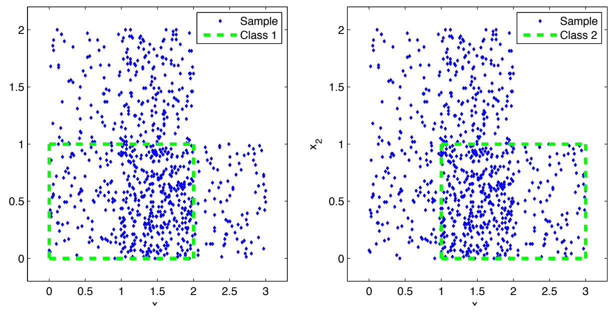


What can I deduce?

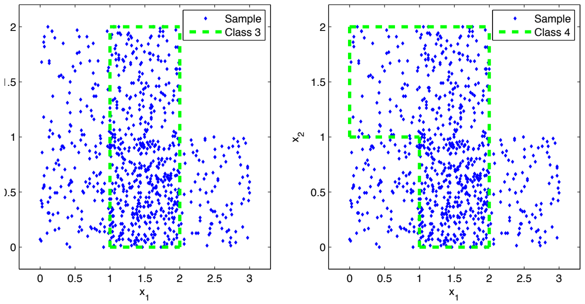
How can data help deduction?

$\mathcal{C} \models \phi$

Checking (Logic) Constraints



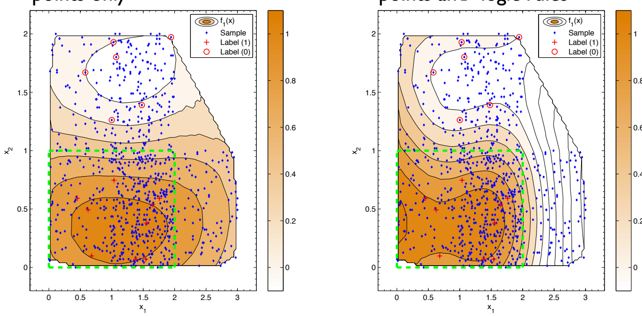
Checking (Logic) Constraints



$$a_1(x) \rightsquigarrow f_1(x)$$

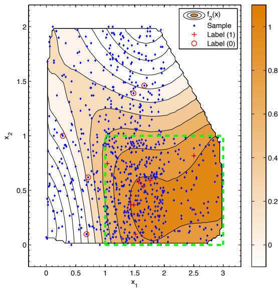
points only

points and “logic rules”

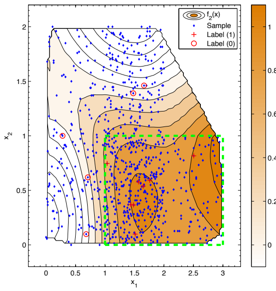


$$a_2(x) \rightsquigarrow f_2(x)$$

points only

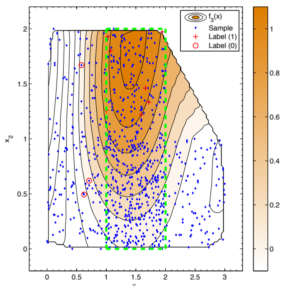


points and “logic rules”

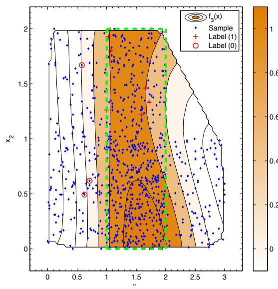


$$a_3(x) \rightsquigarrow f_3(x)$$

points only

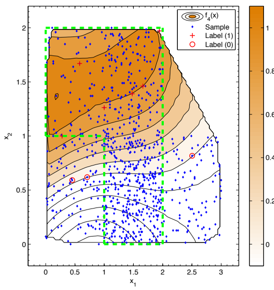


points and “logic rules”

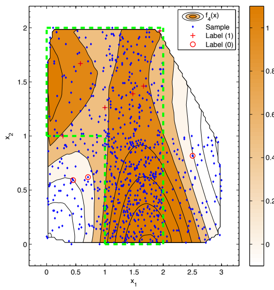


$$a_4(x) \rightsquigarrow f_4(x)$$

points only



points and "logic rules"



Checking Constraints

FOL clause	Category	Average Truth Value	
$a_1(x) \wedge a_2(x) \Rightarrow a_3(x)$	KB	98.26% (1.778)	
$a_3(x) \Rightarrow a_4(x)$	KB	98.11% (2.11)	
$a_1(x) \vee a_2(x) \vee a_3(x)$	KB	96.2% (3.34)	
$a_1(x) \wedge a_2(x) \Rightarrow a_4(x)$	LD	96.48% (3.76)	√ True
$a_1(x) \wedge a_3(x) \Rightarrow a_2(x)$	ENV	91.32% (5.67)	
$a_3(x) \wedge a_2(x) \Rightarrow a_1(x)$	ENV	91.7% (4.57)	
$a_2(x) \wedge a_3(x) \Rightarrow a_4(x)$	LD	96.58% (4.13)	√
$a_3(x) \Rightarrow a_1(x) \vee a_2(x) \vee a_4(x)$	LD	99.7% (0.54)	√
$a_1(x) \wedge a_4(x)$	ENV	45.26% (5.2)	False
$a_2(x) \vee a_3(x)$	ENV	78.26% (6.13)	
$a_1(x) \vee a_2(x) \Rightarrow a_3(x)$	ENV	68.28% (5.86)	
$a_1(x) \wedge a_2(x) \Rightarrow \neg a_4(x)$	ENV	3.51% (3.76)	
$a_1(x) \wedge \neg a_2(x) \Rightarrow a_3(x)$	ENV	27.74% (18.96)	
$a_2(x) \wedge \neg a_3(x) \Rightarrow a_1(x)$	ENV	5.71% (5.76)	

Search reduced to manifolds instead of the Boolean hypercube!

Checking Constraints in the Environment

$$a_1(x) \wedge a_2(x) \Rightarrow a_3(x)$$
$$a_3(x) \Rightarrow a_4(x)$$
$$a_1(x) \vee a_2(x) \vee a_3(x)$$

"rules"

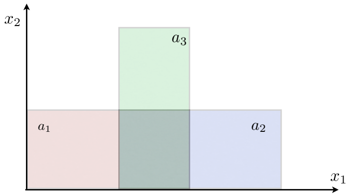
$\phi_i(x, f(x)) = 0$

tasks

$f(x)$

perceptual space

x



Formally false **?** but true in this environment!

$$a_1(x) \wedge a_3(x) \Rightarrow a_2(x) \quad a_1 = 1, \quad a_2 = 0 \quad a_3 = 1$$
$$a_3(x) \wedge a_2(x) \Rightarrow a_1(x) \quad a_1 = 0, \quad a_2 = 1 \quad a_3 = 1$$

Full Inference (formal+environment-based)

Patterns, labels and individuals

(X, x)

Giuseppe 178, 70, 45

label

X

pattern

x

What about learning and inference with individuals?

Inference in Formal Logic

only labels are involved!

```
Domain(label="People")
Individual(label="Marco", "People")
Individual(label="Giuseppe", "People")
Individual(label="Michelangelo", "People")
Individual(label="Francesco", "People")
Individual(label="Franco", "People")
Individual(label="Andrea", "People")
```

```
Predicate(label="fatherOf", ("People", "People"))
Predicate(label="grandFatherOf", ("People", "People"))
Predicate(label="eq", ("People", "People"), function=eq)
```

```
Constraint("fatherOf(Marco, Giuseppe)")
Constraint("fatherOf(Giuseppe, Michelangelo)")
Constraint("fatherOf(Giuseppe, Francesco)")
Constraint("fatherOf(Franco, Andrea)")
```

```
Constraint("forall x: not fatherOf(x,x)")
Constraint("forall x: not grandFatherOf(x,x)")
```


Inference in Formal Logic

```
Constraint("forall x: forall y: fatherOf(x,y) -> not fatherOf(y,x)")
Constraint("forall x: forall y: grandFatherOf(x,y)
-> not grandFatherOf(y,x)")
Constraint("forall x: forall y: fatherOf(x,y) -> not grandFatherOf(x,y)")
Constraint("forall x: forall y: grandFatherOf(x,y) -> not fatherOf(x,y)")
```

```
Constraint("forall x: forall y: forall z: fatherOf(x,z) and fatherOf(z,y) ->
    grandFatherOf(x,y)")
Constraint("forall x: forall y: forall z: (fatherOf(x,y) and not eq(x,z)) ->
    not fatherOf(z,y)")
```

Inference in Formal Logic

```
grandFatherOf("Marco", "Michelangelo")
grandFatherOf("Marco", "Francesco")
```

```
Constraint("forall x: forall y: forall z: grandFatherOf(x,z) and
    fatherOf(y,z) -> fatherOf(x,y)")
```


How does it work?

learnable

grounded pair	father	grandfather
(Marco, Giuseppe)	$w^f(Mar, Giu)$	$w^{gf}(Mar, Giu)$
(Marco, Francesco)	$w^f(Mar, Fra)$	$w^{gf}(Mar, Fra)$
...		

$w^f(Mar, Giu) = 1$ $w^f(Giu, Mic) = 1$ $w^f(Giu, Fra) = 1$ $w^f(Fra, And) = 1$

How does it work?

Łukasiewicz logic

$$T(x,y) = \max\{0, x+y-1\}$$
$$\Rightarrow \quad \min\{1, 1-x+y\}$$

$w^f(Mar, Giu) = 1 \quad w^f(Giu, Mic) = 1 \quad w^f(Giu, Fra) = 1 \quad w^f(Fra, And) = 1$

```
Constraint("forall x: forall y: forall z: fatherOf(x,z) and fatherOf(z,y) ->
     grandFatherOf(x,y)")
```

$$\sum_{X,Y,Z} \min\{1 - \max\{\underbrace{w^f(X,Z) + w^f(Z,Y)}_{\text{constant}} - 1, 0\} + \underbrace{w^{gf}(X,Y)}_{\text{learnable}}, 1\}$$

Full Inference on Individuals (X, x)

$w^f(X,Y), w^{gf}(X,Y)$ from formal logic

consistency constraints

$\omega^f(x,y), \omega^{gf}(x,y)$ from neural nets

$(\mathtt{age}_x, \mathtt{weight}_x, \mathtt{height}_x, \mathtt{age}_y, \mathtt{weight}_y, \mathtt{height}_y)$

Complexity issues: the inference in the environment avoids massive exploration of the Boolean hypercube

Pattern Generation

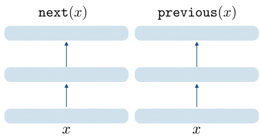
Provide a constrained-based description
of what you want to generate

Generating the next/previous char

$\forall x\ \texttt{IsZero}(x) \Rightarrow \texttt{zero}(x)$

$\forall x\ \texttt{IsOne}(x) \Rightarrow \texttt{one}(x)$

$\forall x\ \texttt{IsTwo}(x) \Rightarrow \texttt{two}(x)$



$$\forall x\ \texttt{IsZero}(x) \Rightarrow \texttt{one}(\texttt{next}(x)) \wedge \texttt{two}(\texttt{previous}(x))$$
$$\forall x\ \texttt{IsOne}(x) \Rightarrow \texttt{two}(\texttt{next}(x)) \wedge \texttt{zero}(\texttt{previous}(x))$$
$$\forall x\ \texttt{IsTwo}(x) \Rightarrow \texttt{zero}(\texttt{next}(x)) \wedge \texttt{one}(\texttt{previous}(x))$$
$$\forall x\ \texttt{next}(\texttt{previous}(x)) = x$$
$$\forall x\ \texttt{previous}(\texttt{next}(x)) = x$$

Generating the next char (con't)

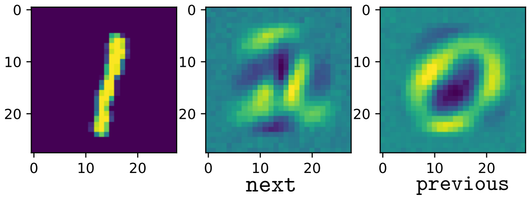
```
Domain("Images", data=X)
Predicate("zero",("Images"),function=Slice(NN, 0))
Predicate("one",("Images"),function=Slice(NN, 1))
Predicate("two",("Images"),function=Slice(NN, 2))
PointwiseConstraint(NN, y, X)
```

```
Predicate("eq",("Images", "Images"), function=eq)
Function("next",("Images"), function=NN_next)
Function("previous", ("Images"), function=NN_prev)

Constraint("forall x: zero(x) -> one(next(x))")
Constraint("forall x: one(x) -> two(next(x))")
Constraint("forall x: two(x) -> zero(next(x))")
Constraint("forall x: zero(x) -> two(previous(x))")
Constraint("forall x: one(x) -> zero(previous(x))")
Constraint("forall x: two(x) -> one(previous(x))")

Constraint("forall x: eq(previous(next(x)),x)")
Constraint("forall x: eq(next(previous(x)),x)")
```


Generating the next char … (con't)

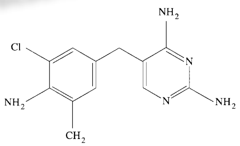


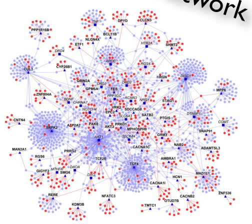
DATA STRUCTURES, CONSTRAINTS
AND GRAPHIC NEURAL NETWORKS

MSc DSAI 2021

physicochemical behavior

Protein Interaction Network



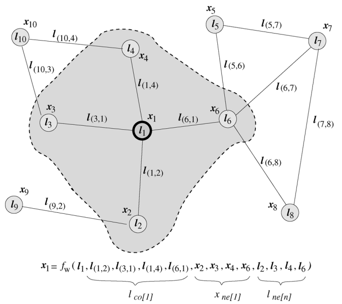


What are the features? The atoms, the bonds?

Graph Neural Networks

(Gori et al, TNN 2009)

neighbor-based computation equilibrium configuration!



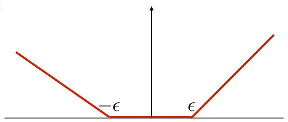
node label connection label neighbor state neighbor label

$$\boldsymbol{x}_n = f_{\boldsymbol{w}}(\boldsymbol{l}_n, \boldsymbol{l}_{\text{co}[n]}, \boldsymbol{x}_{\text{ne}[n]}, \boldsymbol{l}_{\text{ne}[n]})$$
$$\boldsymbol{o}_n = g_{\boldsymbol{w}}(\boldsymbol{x}_n, \boldsymbol{l}_n)$$

$$\boldsymbol{x} = F_{\boldsymbol{w}}(\boldsymbol{x}, \boldsymbol{l})$$
$$\boldsymbol{o} = G_{\boldsymbol{w}}(\boldsymbol{x}, \boldsymbol{l}_{\boldsymbol{N}})$$

Constrained-based expression of data structures

$$\mathcal{G}(x) = \max(||x||_1 - \epsilon, 0)$$



$$\forall v \in V,\ \mathcal{G}\left(x_v - f_a(x_{\text{ne}[v]}, l_{\text{ne}[v]}, l_{(v,\text{ch}[v])}, l_{(\text{pa}[v],v)}, x_v, l_v | \theta_{f_a})\right) = 0$$

$v \in S \subseteq V$

auxiliary variable

$$\min_{\theta_{f_a}, \theta_{f_r}, X} \sum_{v \in S} L(f_r(x_v|\theta_{f_r}), y_v)$$ any constraint can be incorporated on the vertexes

subject to $\mathcal{G}\left(x_v - f_a(x_{\text{ne}[v]}, l_{\text{ne}[v]}, l_{(v,\text{ch}[v])}, l_{(\text{pa}[v],v)}, x_v, l_v | \theta_{f_a})\right) = 0, \quad \forall v \in V$

… data diffusion

LEARNING OF
CONSTRAINTS

Learning of constraints?

rules

$\phi_i(x, f(x)) = 0$

tasks

$f(x)$

perceptual space

x

$$\phi(f(x)) = 0$$

given — task to be learned

What about a joint discovery of ϕ and f ?

$$\phi \circ f = 0$$

Formulation of Learning

$f : \mathcal{X} \subset \mathbf{R}^d \to \mathcal{Y} \subset \mathbf{R}^n$ and $\phi : \mathcal{Y} \subset \mathbf{R}^n \to \mathcal{Z} \subset \mathbf{R}^m$

knowledge map

$f(\mathcal{X}) \subset \mathcal{Y}$

perceptual space concept space

quantization over all the perceptual space

$$\forall x \in \mathcal{X} : \quad \bar{\phi}(f(x)) = \prod_{i=1}^{m} \phi_i(f(x)) = 0$$

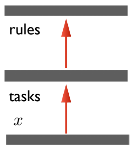
$$\forall x \in \mathcal{X}_i : \quad \phi_i(f(x)) = 0, \quad i = 1, \ldots, \bar{m}$$

$$1 \le \bar{m} \le m \quad \mathcal{Q}_x = \{\mathcal{X}_i, \quad i = 1, \ldots, \bar{m} \le m\}$$

"Rules" and constraints

rule/neuron

$$P_{R=i|X=x}(\phi \circ f) = \frac{e^{-\phi_i(f(x))}}{\sum_{j=1}^{m} e^{-\phi_j(f(x))}}$$



probabilistic normalization

$$\sum_{i=1}^{m} \int_{\mathcal{X}} dx p_X(x) P_{R=i|X=x}(\phi \circ f) = \int_{\mathcal{X}} dx \left(\sum_{i=1}^{m} \frac{e^{-\phi_i(f(x))}}{\sum_{j=1}^{m} e^{-\phi_j(f(x))}} \right) p_X(x) = 1$$

Maximum Mutual Information Principle

$$f^\star(\bar\phi^\star) = \underset{(f,\phi)\in\mathcal{C}_{\mathcal{X}}(f,\bar\phi^\star)}{\arg\min} \quad \alpha\|f\|$$

$$\bar\phi^\star(f^\star) = \underset{(f,\phi)\in\mathcal{C}_{\mathcal{X}}(f^\star,\phi)}{\arg\min} \quad \big(\beta\|\bar\phi^\star\| - I_{X,R}(\bar\phi^\star\circ f)\big)$$

parsimony

information-based index

MMI Stage-based learning

$$\begin{aligned}
\phi_0, &\big[f_0^\star(\bar\phi_0), \phi_1^\star(f_0^\star)\big],\\
&\big[f_1^\star(\bar\phi_1^\star), \phi_2^\star(f_1^\star)\big],\\
&\big[f_2^\star(\bar\phi_2^\star), \phi_3^\star(f_2^\star)\big],\\
&\ldots\\
&\big[f_{\kappa-1}^\star(\bar\phi_{\kappa-1}^\star), \phi_\kappa^\star(f_{\kappa-1}^\star)\big],\\
&\big[f_\kappa^\star(\bar\phi_\kappa^\star), \phi_{\kappa+1}^\star(f_\kappa^\star)\big].
\end{aligned}$$

$$f^\star = \lim_{\kappa\to\infty} f_\kappa^\star(\bar\phi_\kappa^\star)$$

$$\phi^\star = \lim_{\kappa\to\infty} \phi_\kappa^\star(f_{\kappa-1}^\star)$$

Developmental function

$$D(f,\phi) = \|(f,\phi)\| - I_{X,R}(\phi \circ f)$$

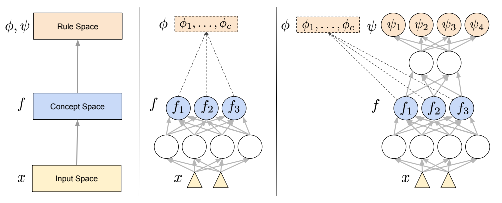
$$\begin{pmatrix} f^\star \\ \phi^\star \end{pmatrix} = \arg\min_{(f,\phi)\in\mathcal{C}_\mathcal{X}} D(f,\phi)$$

$$\bar\phi(f(x)) = \bar\phi_s(f(x)) \cdot \bar\phi_l(f(x))$$

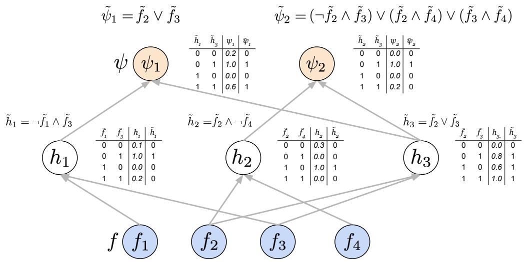
given constraints learned constraints

Lagrangian approach by using neural nets

Neural-Based Solution



Looking for explanations …



$\tilde{h}_1$	$\tilde{h}_3$	ψ_1	$\tilde{\psi}_1$
0	0	*0.2*	0
0	1	*1.0*	1
1	0	*0.0*	0
1	1	*0.6*	1

$\tilde{h}_2$	$\tilde{h}_3$	ψ_2	$\tilde{\psi}_2$
0	0	*0.0*	0
0	1	*1.0*	1
1	0	*0.0*	0
1	1	*0.2*	0

$\tilde{f}_1$	$\tilde{f}_3$	h_1	$\tilde{h}_1$
0	0	*0.1*	0
0	1	*1.0*	1
1	0	*0.0*	0
1	1	*0.2*	0

$\tilde{f}_2$	$\tilde{f}_4$	h_2	$\tilde{h}_2$
0	0	*0.3*	0
0	1	*0.0*	0
1	0	*1.0*	1
1	1	*0.0*	0

$\tilde{f}_2$	$\tilde{f}_3$	h_3	$\tilde{h}_3$
0	0	*0.0*	0
0	1	*0.8*	1
1	0	*0.6*	1
1	1	*1.0*	1

Conclusions

A unified framework for learning and reasoning

- Probability distributions and Lagrange multipliers, reactions of constraints … links with Backprop
- Bridging symbols and sub-symbols (logic representations & learning)
- Inference in the environment, full inference (searching in manifolds instead of the Boolean hypercube)
- GNN as a special case
- Learning of constraints and explanation
- Time and developmental issues (topics in Developmental Psychology), biological plausibility