Inference, Learning and Laws of Nature

Salvatore Frandina[1] Marco Gori[1] Marco Lippi[1]
Marco Maggini[1] Stefano Melacci[1]

[1]University of Siena, Italy

NeSy' 13
Ninth International Workshop on Neural-Symbolic Learning and Reasoning
IJCAI-13

Outline

Inference and Learning as cognitive processes

- Inference represents the deductive ability to derive the logical conclusion from a set of premises.
- Learning is the inductive ability to acquire, modify and reinforce knowledge from a set of observed data.
- Human decision mechanisms exploit both these abilities to take decisions.

We need to unify inference and learning

- Real–world problems are complex and uncertain.
- Complexity can be handled by logic theory.
- Uncertainty can be handled by probability theory.

Toward a unified framework

- Historically...
  - Inference is framed into logic formalism whereas the process of learning is addressed by statistical approaches.
- Nowadays...
  - Unification of inference and learning leads to the framework of probabilistic reasoning.
- For neural networks, the neural symbolic integration is well studied but it lacks of solid mathematical foundations like for probabilistic reasoning.

Toward a unified framework

- We replace the focus on probabilistic reasoning with cognitive laws.
- The human decision mechanisms may be better understood by means of the variational laws of Nature.
- There is a strong analogy between learning from constraints and analytic mechanics.

Example

An agent lives in the environment and behaves following laws like those governing a particle subject to a force field.

What are the cognitive laws?

- The formulation of the problem in terms of cognitive laws leads to a natural integration of inference and learning.
- An agent continuously interacts with the environment and receives stimuli expressed in terms of constraints among set of tasks.
- In our context, the reaction of an agent to the stimuli follows the laws emerging from stationary points of a cognitive action functional.
- In analytic mechanics, the motion of particles subject to a force field follows the minimization of an action functional.

How are Machine Learning and Analytic Mechanics related?

Machine Learning $\longleftrightarrow$ Analytic Mechanics		
variable	machine learning	analytic mechanics
w_i	weight	particle position
$\dot{w}_i$	weight variation	particle velocity
V	constraint penalty	potential energy
T	temporal smoothness	kinetic energy
$\mathcal{L}$	cognitive Lagrangian	mechanical Lagrangian
$\mathcal{S}$	cognitive action	mechanical action

Coupled inference and learning mechanism

- A newborn agent begins its life with a given potential energy and evolves by changing its parameters.
- The potential energy is partially transformed into kinetic energy and the rest is dissipated.
- The velocity of weights decreases until the agent ends into a stable configuration.
- The inference and learning process finishes when all the initial potential energy is dissipated.

Unified on–line formulation of inference and learning

- Consider a multitask problem with q interacting tasks.
- Each task i transforms the input $x \in X \subset R^n$ using weights $W \in R^m$ by means of a function $f : X \times W \to R$, e.g. a neural network.
- The learning process consists of finding

$$w^* = \arg\min_{w \in W} \mathcal{S}(w),$$

- where the cognitive action is defined as

$$\mathcal{S} = \int_0^{t_e} \underbrace{e^{\beta t}\mathcal{L}}_{\mathcal{L}_\beta \text{ Dissipative Lagrangian}} dt \quad \beta > 0,$$

$[0, t_e]$ is a temporal horizon.

Constraint penalty and temporal smoothness

- The Lagrangian is defined as $\mathcal{L}(w) = T(w) - V(w)$.
- The constraint penalty or potential energy is

$$\mathcal{V}(f) = \mathcal{V}(f(x, w)) = \int_0^{t_e} V(w(t))dt,$$

where $V(w(t))$ collects all the constraints i.e. supervisions, logic rules, etc.

- The temporal smoothness or cognitive kinetic energy is

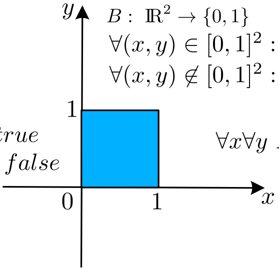
$$T = \frac{1}{2}\sum_{i=1}^{m} \mu_i \dot{w}_i^2(t),$$

where $\mu_i > 0$ is the cognitive mass associated with the particle i.

Logic information

- A toy example where the cognitive laws unify inference and learning into the same framework.

$$A: \mathbb{R} \to \{0,1\}$$
$$\forall x \in [0,1]: A(x) = true$$
$$\forall x \notin [0,1]: A(x) = false$$



$$B: \mathbb{R}^2 \to \{0,1\}$$
$$\forall (x,y) \in [0,1]^2: \ B(x,y) = true$$
$$\forall (x,y) \notin [0,1]^2: \ B(x,y) = false$$

$$\forall x \forall y \ A(x) \wedge A(y) \Rightarrow B(x,y)$$

- We have information about the functions and knowledge on their relationship.

Perceptive information

- $a: \ R \to [0,1]$ and $b: \ R \times R \to [0,1]$ are real-valued functions associated with $A(\cdot)$ and $B(\cdot,\cdot)$.
- We have also supervised data

$$A(\cdot) \ : \ \{(x_\kappa, d^a_\kappa)\}_{\kappa=1}^{\ell_a} \text{ arriving at time } \{t^a_\kappa\}_{\kappa=1}^{\ell_a}$$

and

$$B(\cdot,\cdot) \ : \ \left\{((x_\kappa, y_\kappa), d^b_\kappa)\right\}_{\kappa=1}^{\ell_b} \text{ arriving at time } \left\{t^b_\kappa\right\}_{\kappa=1}^{\ell_b}.$$

Logical and perceptual potential energy

- The total potential energy is $\mathcal{V}(f) = \int_0^{t_e} V(w(t))dt$ where

$$V(w(t)) := c_1 \underbrace{a(x(t)) \cdot b(y(t))\,(1 - b(x(t), y(t)))}_{\text{logic part}} + \underbrace{c_2 \sum_{\kappa=1}^{\ell_a} h(a(x_\kappa), d_\kappa^a) \cdot \delta(t - t_\kappa^a)}_{\text{perception part}} + \underbrace{c_2 \sum_{\kappa=1}^{\ell_b} h(b(x_\kappa, y_\kappa), d_\kappa^b) \cdot \delta(t - t_\kappa^b)}_{\text{perception part}},$$

c_1 and c_2 are two constants and h is a loss function.

Lagrangian Cognitive Equation

- Each stationary point of the cognitive action satisfies the Euler-Lagrange equation $\frac{d}{dt}\frac{\partial \mathcal{L}_\beta}{\partial \dot{w}_i} - \frac{\partial \mathcal{L}_\beta}{\partial w_i} = 0$.
- Considering that $\mathcal{L}_\beta = e^{\beta t}\mathcal{L}$ we get

$$\beta e^{\beta t}\frac{\partial \mathcal{L}}{\partial \dot{w}_i} + e^{\beta t}\underbrace{\left(\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{w}_i} - \frac{\partial \mathcal{L}}{\partial w_i}\right)}_{\text{non dissipative term}} = 0.$$

- Rearranging the terms, we get the Lagrangian cognitive equation

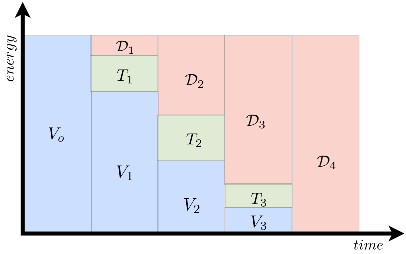
$$\ddot{w}_i + \beta \dot{w}_i + \mu_i^{-1}\ V'_{w_i} = 0, \quad i = 1, \ldots, m.$$

Evolution of the life of the agent

- The evolution of the agent is driven by the previous equation paired with Cauchy's conditions $w_i(0)$ and $\dot{w}_i(0)$.
- The Lagrangian cognitive equation leads to classical online Backpropagation when strong dissipation is enforced.
- For high values of β, the learning rate is $\eta_i = 1/(\beta\mu_i)$ and the solution of Lagrangian cognitive equation is $w_i^*|_k = w_i^*|_{k-1} - \eta_i * g_{i,k}$.

  Frandina S., Gori M., Lippi M., Maggini M., Melacci S. Variational Foundations of Online Backpropagation. 2013, September at ICANN, Sofia, Bulgaria.

The evolution of the energy balance



- At the begin the available energy is the potential energy i.e. the inference loss.
- As the time goes by the initial potential energy is continuously transformed into kinetic energy and dissipated energy.
- The inference and learning process ends when all the initial potential energy is dissipated.

Cognitive Energy

- The agent evolution is interpreted in terms of cognitive energy
$$\mathcal{E} = T + V + \mathcal{D},$$
where the term $\mathcal{D}(t) = \int_0^t D(w(\theta))\, d\theta$ is the dissipated energy over $[0, t_e]$.
- Multiplying the Lagrangian Cognitive equation by $\dot{w}_i$, we get
$$\dot{w}_i \cdot \ddot{w}_i + \beta_i \dot{w}_i^2 + \mu_i^{-1} V'_{w_i} \cdot \dot{w}_i = 0,$$
from which
$$\int_0^{t_e} \frac{d}{dt} \underbrace{\frac{1}{2} \sum_{i=1}^{m} \mu_i \dot{w}_i^2}_{T(w(t))} + \underbrace{\sum_{i=1}^{m} \mu_i \beta \dot{w}_i^2}_{D(w(t))} + \underbrace{\sum_{i=1}^{m} V'_{w_i} \dot{w}_i}_{\frac{dV(w(t))}{dt} - \frac{\partial V}{\partial t}} dt = 0.$$

Conservation of Cognitive Energy

- Rearranging the terms, we get the principle of conservation of cognitive energy

$$\int_0^{t_e} \underbrace{\left(\frac{dT(w(t))}{dt} + \frac{d\mathcal{D}(t)}{dt} + \frac{dV(w(t))}{dt} \right)}_{d\mathcal{E}/dt} dt = \int_0^{t_e} \frac{\partial V}{\partial t} dt.$$

- Cognitive energy is constant whenever there is not a new stimulus (i.e. constraints).
- In general, the agent could be provided with:
  - Fixed initial potential energy $\frac{\partial V}{\partial t} = 0$ i.e. fixed knowledge base.
  - Time variant potential energy $\frac{\partial V}{\partial t} \neq 0$ i.e. varying knowledge base.

Summary

- The variational laws of Nature define a unified on–line inference and learning scheme.
- The evolution of the life of the agent is expressed as principle of conservation of energy.

- Future work
  - Deal with the problem of local minima of potential energy.
  - The theory requires extended experimental evaluation.

Bridging Logic and Kernel Machine

- We have developed a theory that bridges logic and kernel machine.
- Semantic Based Regularization is a framework to learning from constraints that unifies inference and learning.
- You can perform this kind of inference and learning on batch-mode using the simulator at `https://sites.google.com/site/semanticbasedregularization/home/software`.

That's all!