A neuro-symbolic approach for real-world event recognition from weak supervision

Gianluca Apriceno[1,2], Andrea Passerini[2], Luciano Serafini[1]

[1]Fondazione Bruno Kessler, Trento, Italy

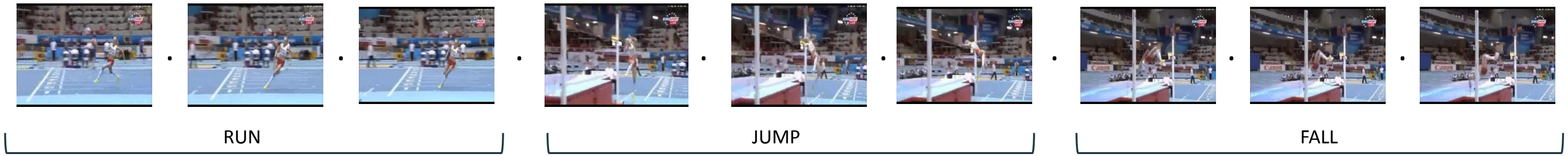
[2]University of Trento, Trento, Italy

29th International Symposium on Temporal Representation and Reasoning

Table of contents

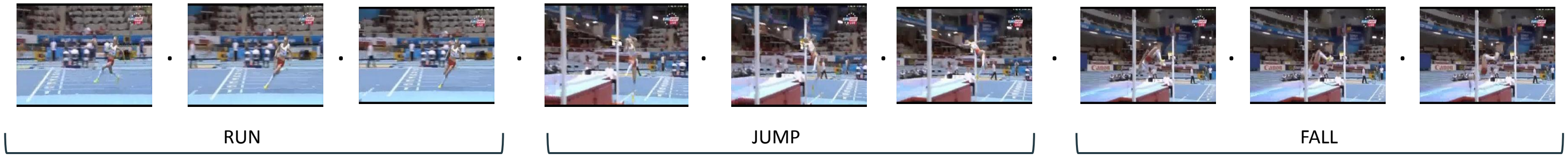
Introduction and motivation

- **Events**



Introduction and motivation

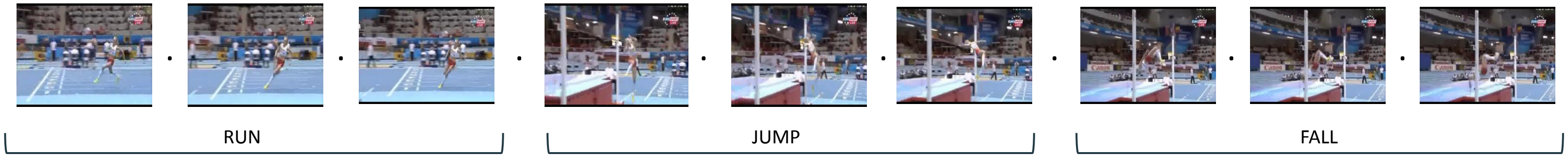
- **Events**



- Neural approaches:
  - Large amounts of annotated training data (errors in the annotations!)
  - Not guaranteed consistency of predictions

Introduction and motivation

- **Events**



- Neural approaches:
  - Large amounts of annotated training data (errors in the annotations!)
  - Not guaranteed consistency of predictions
- Neuro-symbolic approaches:
  - Low level processing with high level reasoning
  - Events - artificial scenarios -> issues (e.g. scalability)

Contributions

1. A Neuro-symbolic approach for event recognition in a real world scenario (sports) (MILP)
2. Experiment: Neural vs Neuro-symbolic

Problem definition

- Let $\mathcal{L}$ be a first order language:

Problem definition

- Let $\mathcal{L}$ be a first order language:
  - $\mathcal{E}$ event types:
    - $\mathcal{S}$ structured events
    - $\mathcal{A}$ atomic events
  - $\mathbb{N}$ time points
  - $happens(e, t_1, t_2)$

Problem definition

- Let $\mathcal{L}$ be a first order language:
  - $\mathcal{E}$ event types:
    - $\mathcal{S}$ structured events
    - $\mathcal{A}$ atomic events
  - $\mathbb{N}$ time points
  - $happens(e, t_1, t_2)$
- Axiom:

$$\forall xyz(happens(x, y, z) \Rightarrow y < z)$$

Problem definition

- Let $\mathcal{L}$ be a first order language:
  - $\mathcal{E}$ event types:
    - $\mathcal{S}$ structured events
    - $\mathcal{A}$ atomic events
  - $\mathbb{N}$ time points
  - $happens(e, t_1, t_2)$
- Axiom:

$$\forall xyz(happens(x, y, z) \Rightarrow y < z)$$

- Semantics:

$$\mathcal{H} = \{happens(e, t_1, t_2) \mid e \in \mathcal{E}, t_1 < t_2, t_1, t_2 \in \mathbb{N}\}$$

Problem definition (example)

- **Our aim:** Given a data sequence $X = \{x_i\}_{i=1}^{l}$ and background knowledge K we have to find an (Herbrand) interpretation I (i.e. description for X) such that $I \models K$

Problem definition (example)

- **Our aim:** Given a data sequence $X = \{x_i\}_{i=1}^{l}$ and background knowledge K we have to find an (Herbrand) interpretation I (i.e. description for X) such that $I \models K$
- **Example**
  - $X = \{x_i\}_{i=1}^{31}$ -> highjump
  - K :

$$\forall b_{hj} e_{ij} (happens(highjump, b_{hj}, e_{hj}) \leftrightarrow \exists\, b_r, e_r, b_j, e_j, b_f, e_f(\\ happens(run, b_r, e_r) \wedge happens(jump, b_j, e_j) \wedge happens(fall, b_f, e_f) \wedge \\ b_r = b_{hj} \wedge e_r = b_j \wedge e_j = b_f \wedge e_f = e_{hj}))$$

Problem definition (example cont.)

- Two examples of interpretations:
  - $I_1 = \{happens(highjump, 1, 31), happens(run, 1, 21), happens(jump, 21, 25), happens(fall, 25, 31)\}$
  - $I_2 = \{happens(highjump, 1, 31), happens(run, 1, 23), happens(jump, 23, 28), happens(fall, 28, 31)\}$
  - $\dots$

Problem definition (example cont.)

- Two examples of interpretations:
  - $I_1 = \{happens(highjump, 1, 31), happens(run, 1, 21), happens(jump, 21, 25), happens(fall, 25, 31)\}$
  - $I_2 = \{happens(highjump, 1, 31), happens(run, 1, 23), happens(jump, 23, 28), happens(fall, 28, 31)\}$
  - $\ldots$
- Cost function:

$$c : I \rightarrow R$$

Problem definition (example cont.)

- Two examples of interpretations:
  - $I_1 = \{happens(highjump, 1, 31), happens(run, 1, 21), happens(jump, 21, 25), happens(fall, 25, 31)\}$
  - $I_2 = \{happens(highjump, 1, 31), happens(run, 1, 23), happens(jump, 23, 28), happens(fall, 28, 31)\}$
  - $\ldots$
- Cost function:

$$c : I \rightarrow R$$

- Select:

$$I_c^* = \operatorname*{argmin}_{I_c \models K} c(I_c)$$

Problem definition (example cont.)

- Two examples of interpretations:
  - $I_1 = \{happens(highjump, 1, 31), happens(run, 1, 21), happens(jump, 21, 25), happens(fall, 25, 31)\}$
  - $I_2 = \{happens(highjump, 1, 31), happens(run, 1, 23), happens(jump, 23, 28), happens(fall, 28, 31)\}$
  - $\dots$
- Cost function:

$$c : I \rightarrow R$$

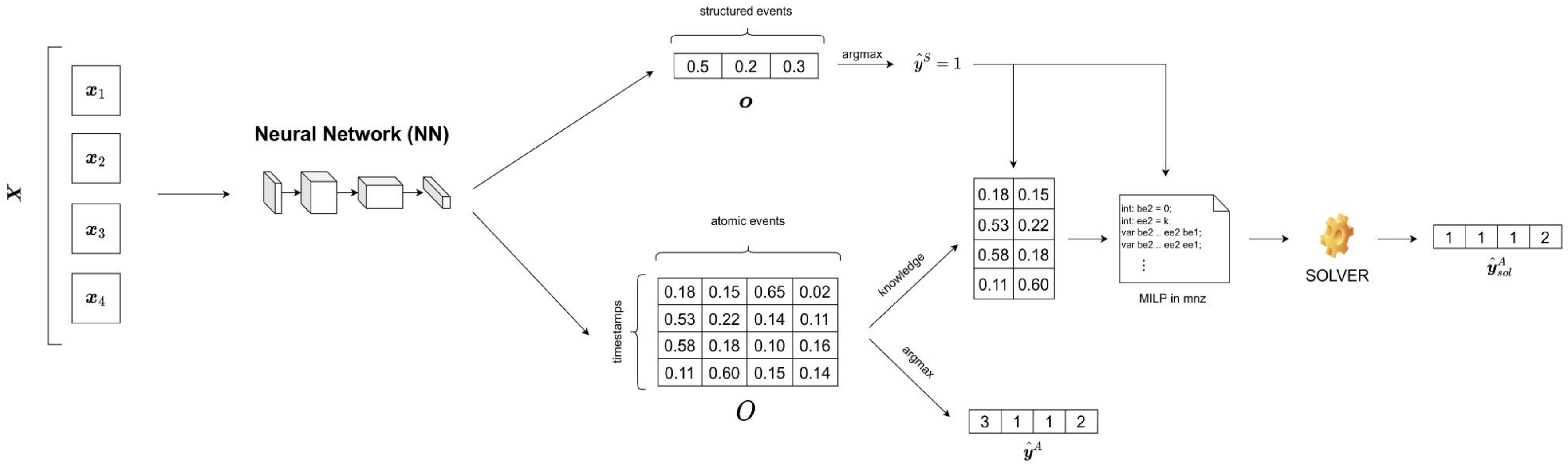
- Select:

$$I_c^* = \operatorname*{argmin}_{I_c \models K} c(I_c)$$

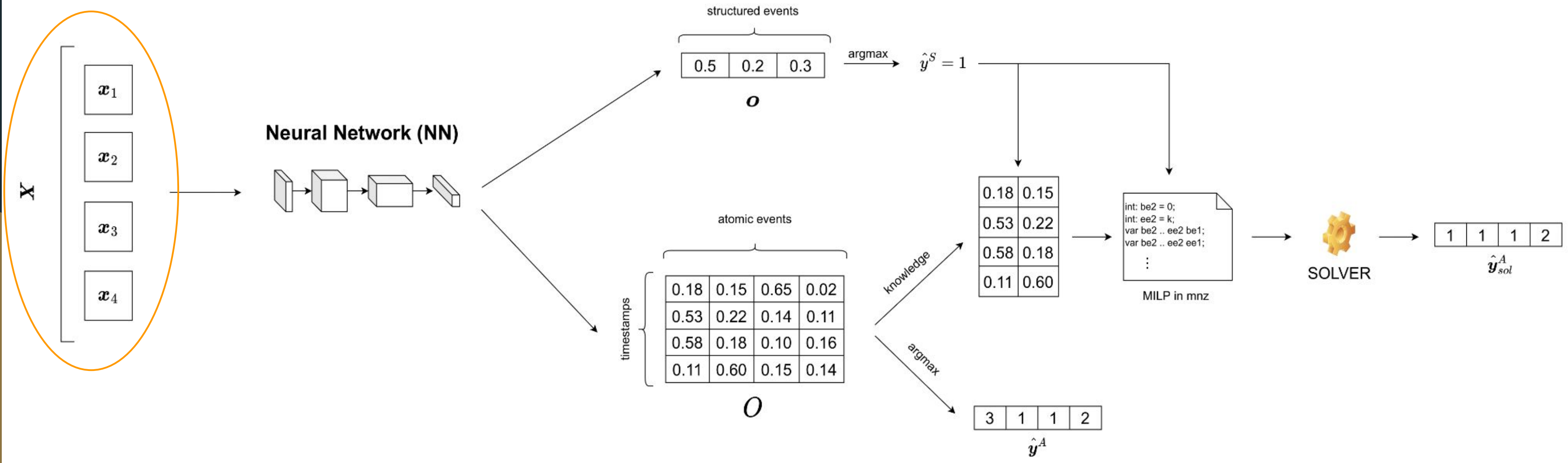
- Supervision:

$$\left\{ \boldsymbol{X}^{(i)}, G_a^{(i)} \right\}_{i=1}^{n}$$

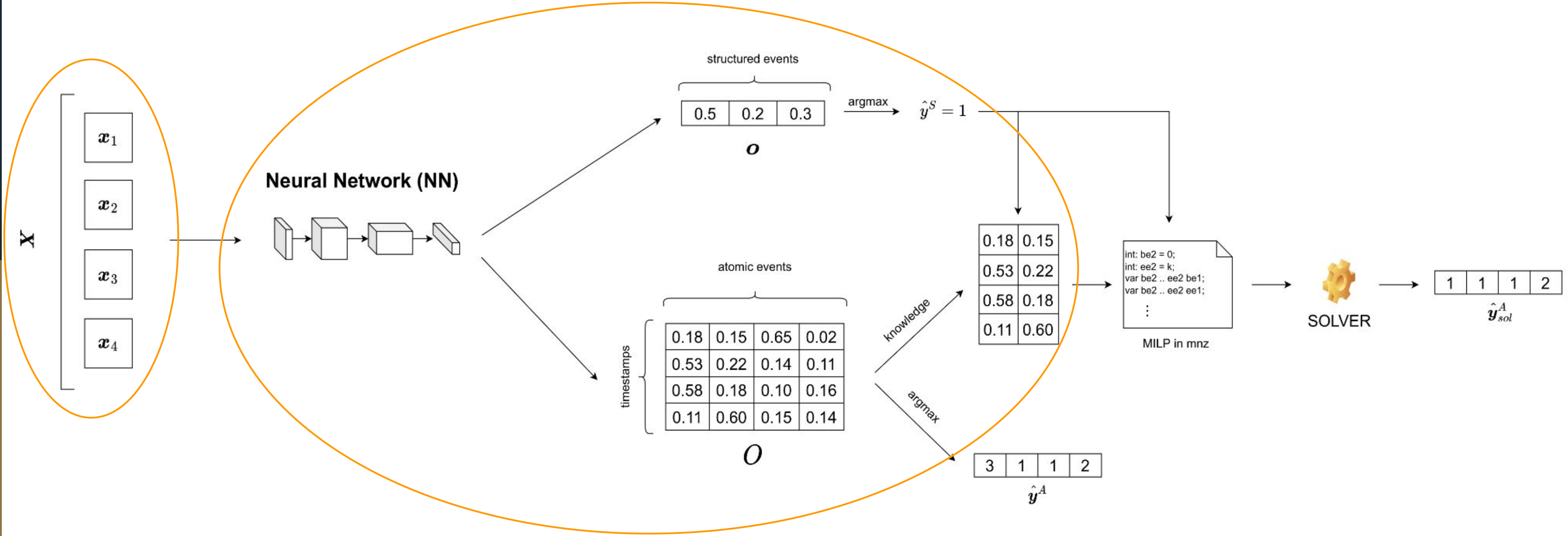
Proposed approach - inference



Proposed approach - inference

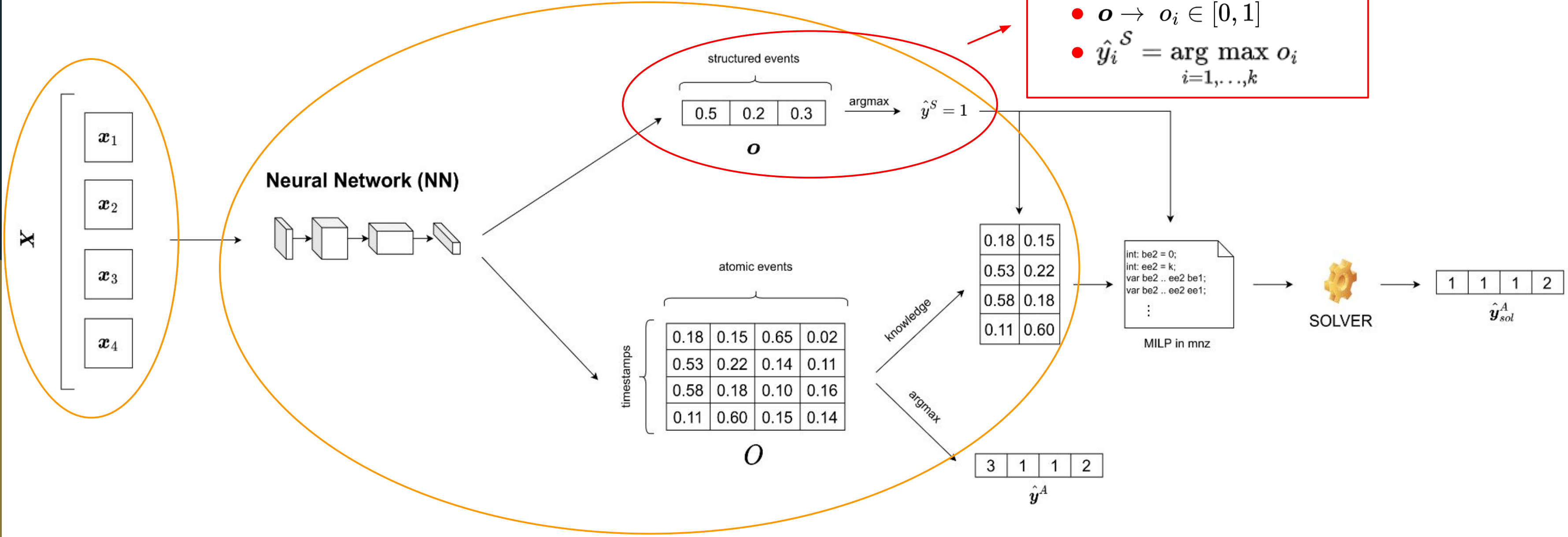


Proposed approach - inference

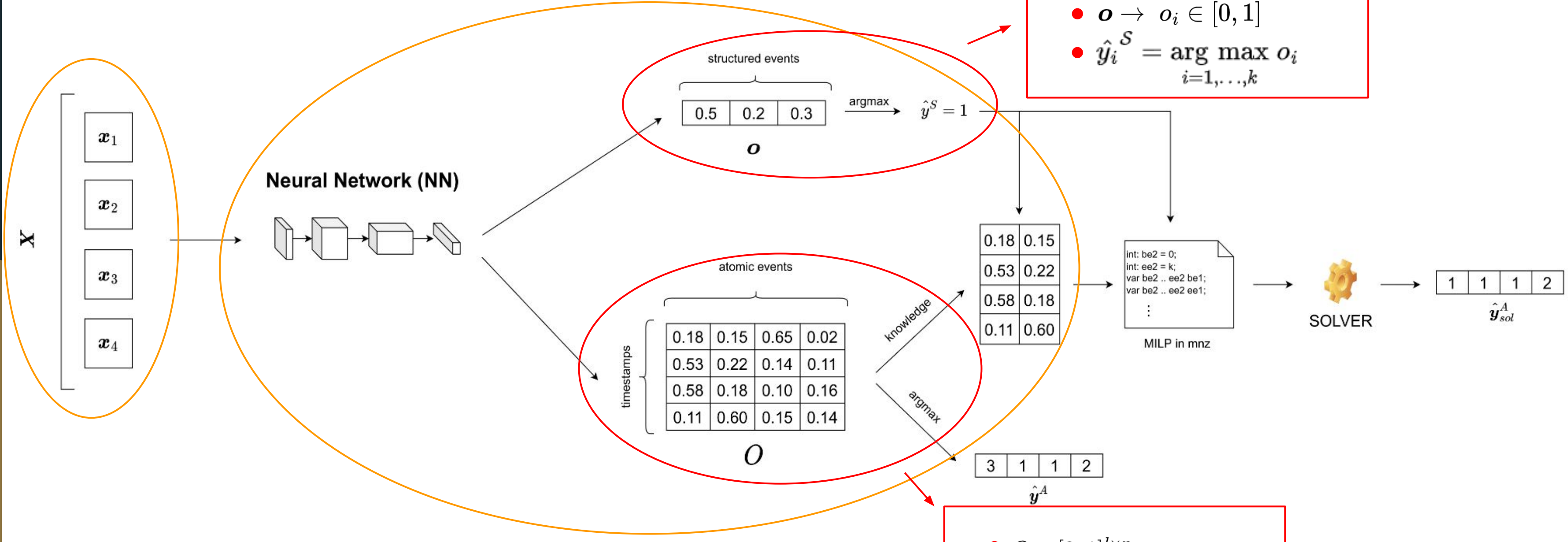


Proposed approach - inference

- $\boldsymbol{o} \rightarrow \; o_i \in [0,1]$
- $\hat{y_i}^{\mathcal{S}} = \arg\max_{i=1,\ldots,k} o_i$



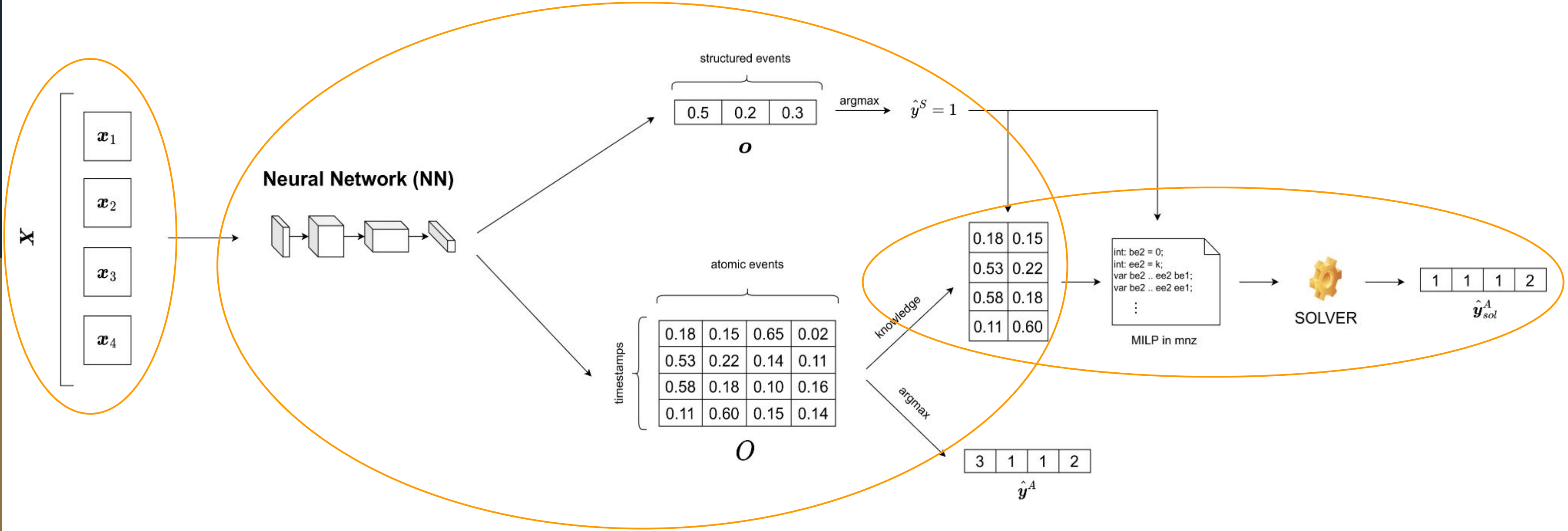
Proposed approach - inference



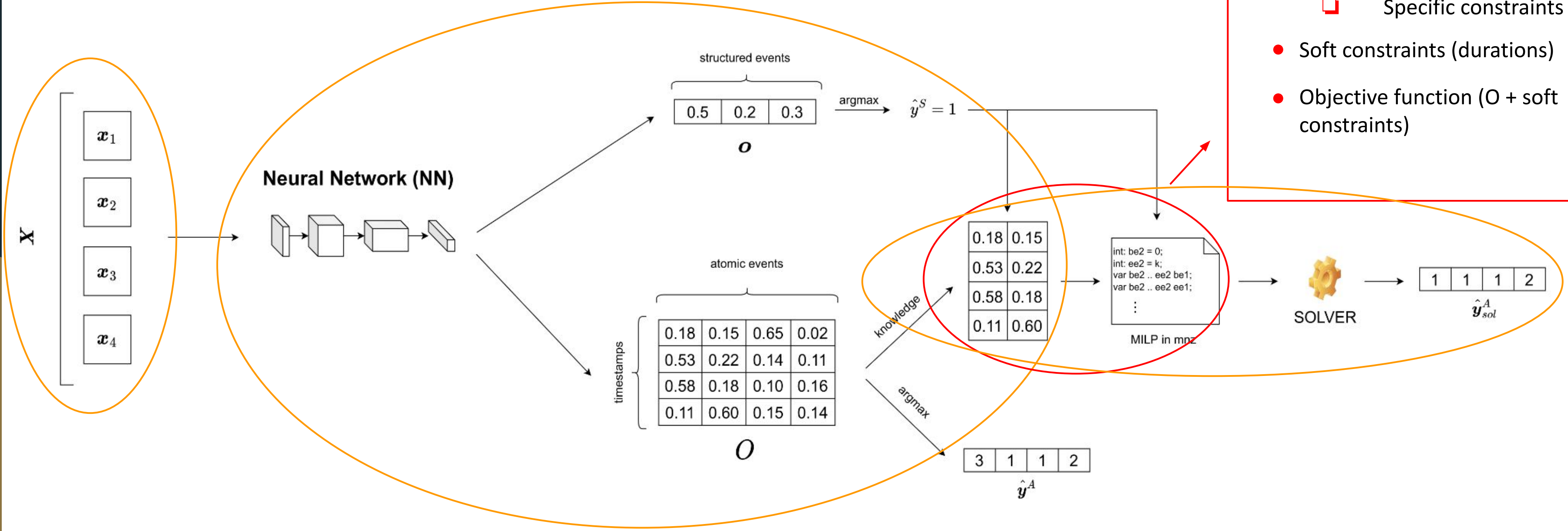
- $\boldsymbol{o} \rightarrow \ o_i \in [0,1]$
- $\hat{y_i}^{\mathcal{S}} = \underset{i=1,\ldots,k}{\arg\max}\ o_i$

- $O \in [0,1]^{l \times n}$
- $\hat{y_i}^{\mathcal{A}} = \underset{1 \leq j \leq n}{\arg\max}\ 0[i,j]$

Proposed approach - inference

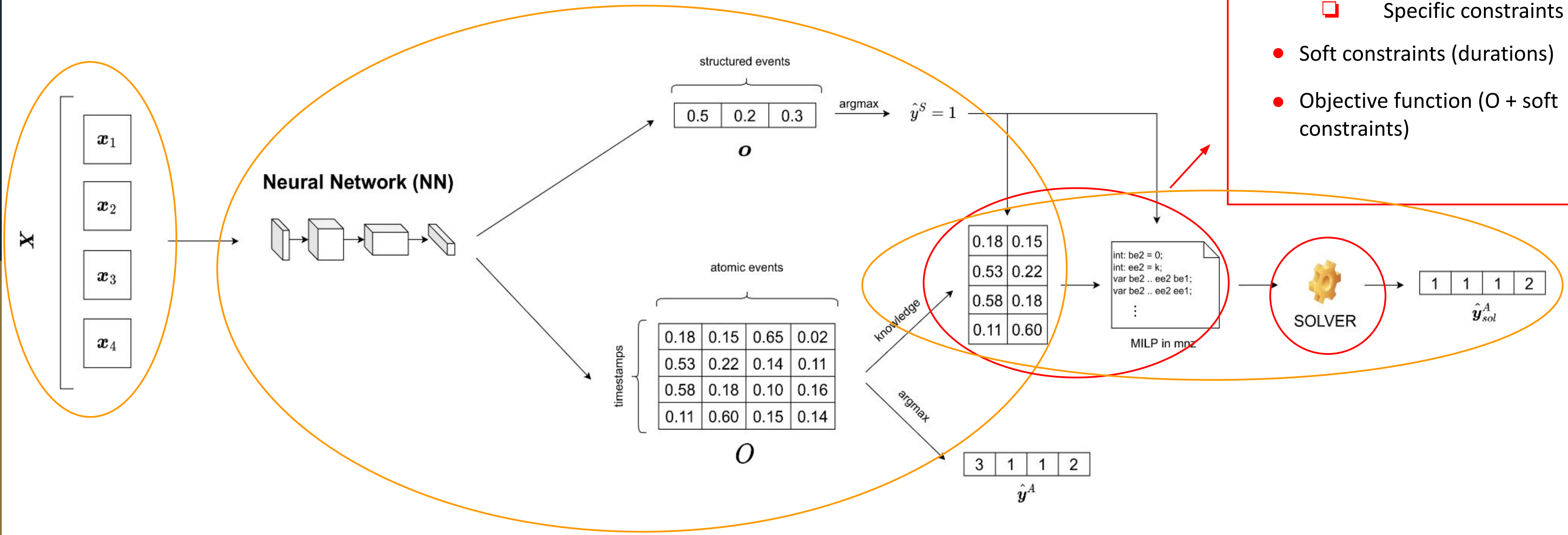


Proposed approach - inference



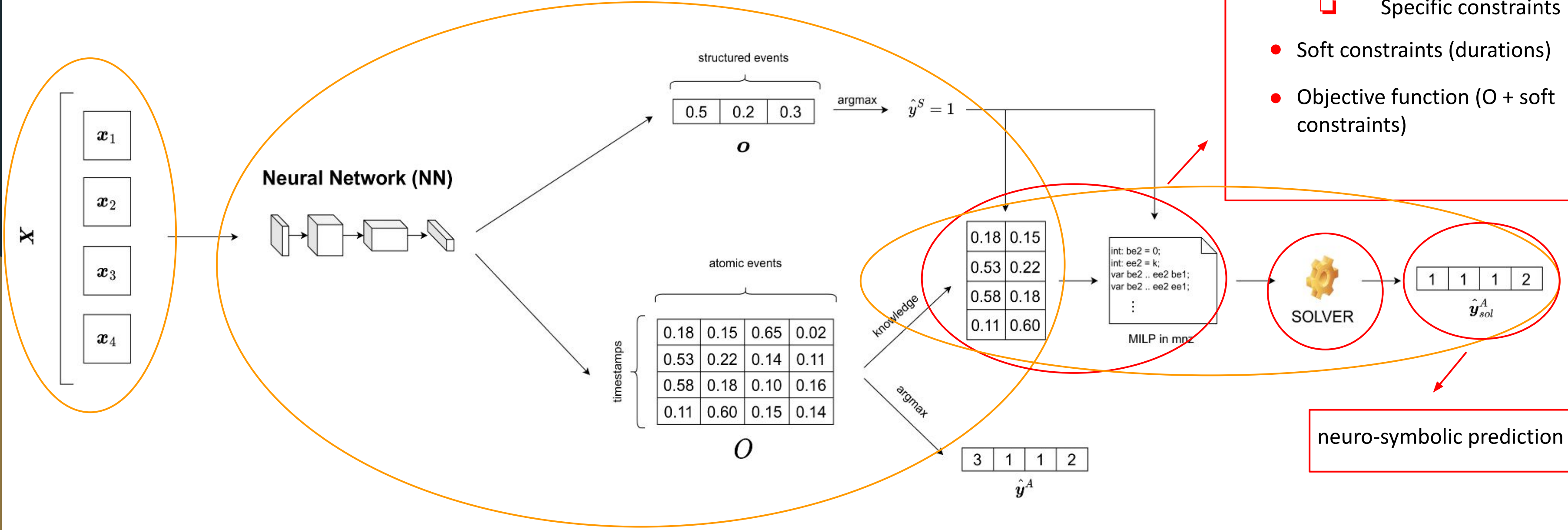
- Hard constraints:
  - Generic constraints
  - Specific constraints
- Soft constraints (durations)
- Objective function (O + soft constraints)

Proposed approach - inference



- Hard constraints:
  - Generic constraints
  - Specific constraints
- Soft constraints (durations)
- Objective function (O + soft constraints)

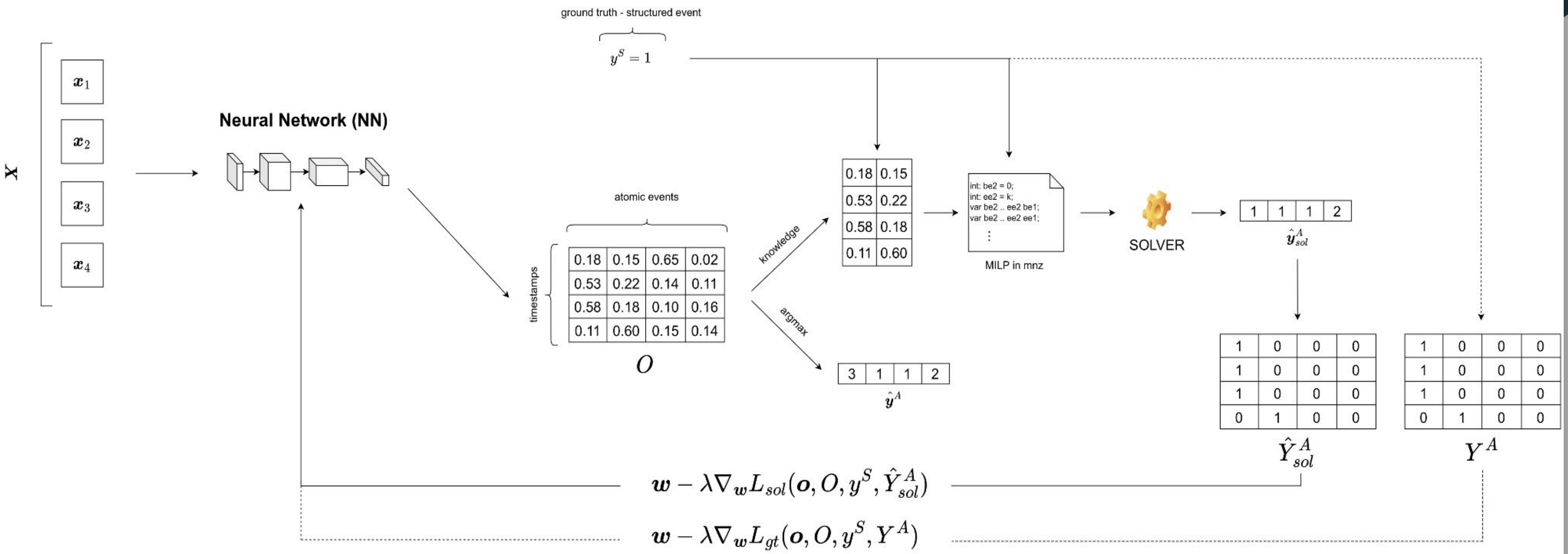
Proposed approach - inference



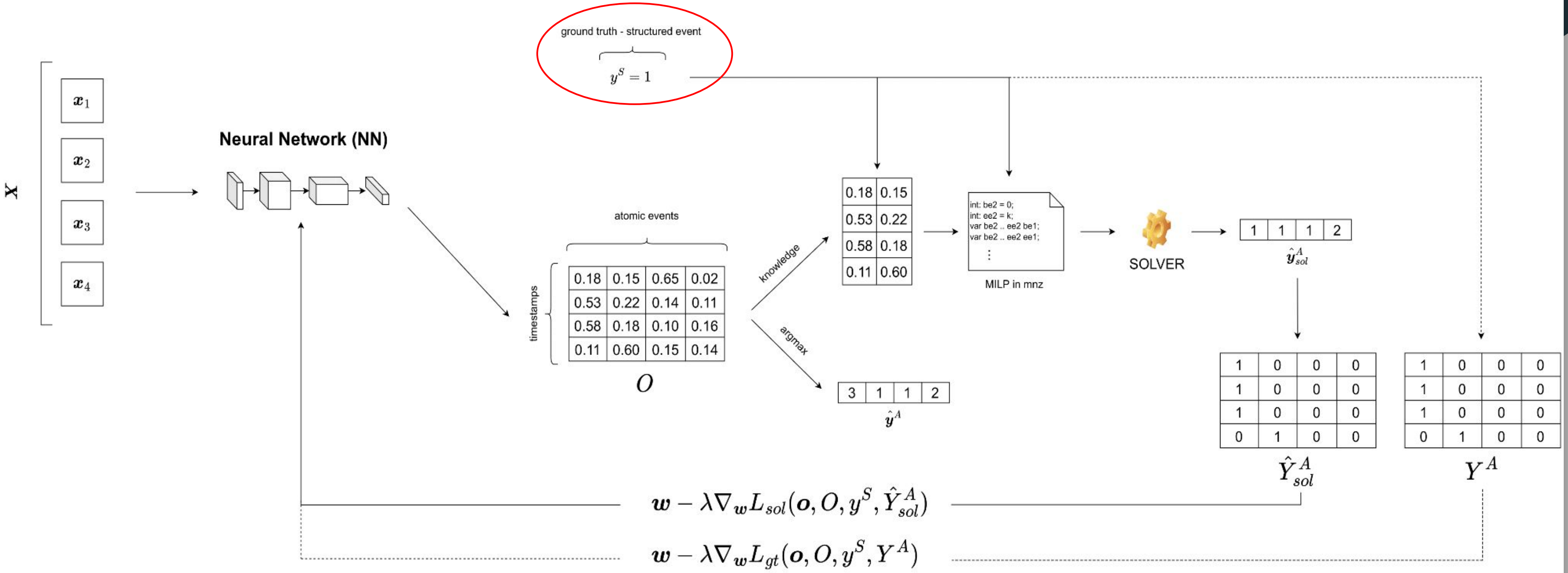
- Hard constraints:
  - Generic constraints
  - Specific constraints
- Soft constraints (durations)
- Objective function (O + soft constraints)

neuro-symbolic prediction

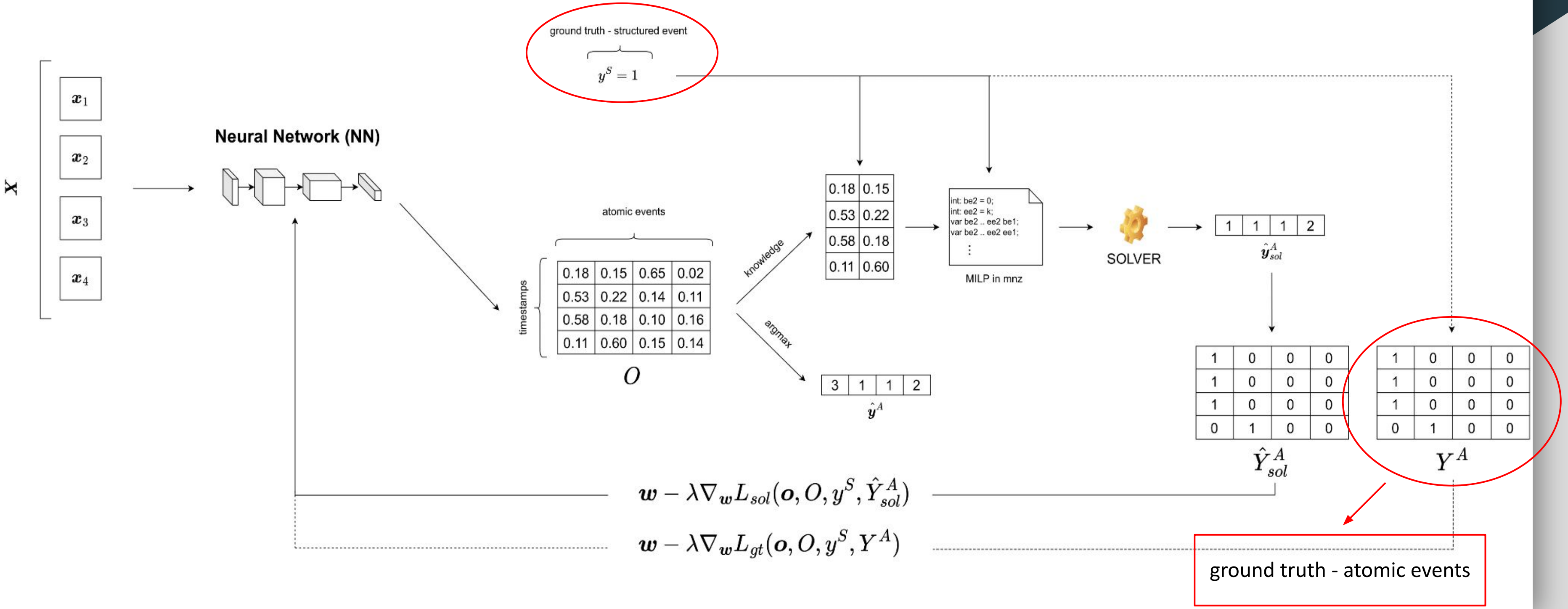
Proposed approach - train



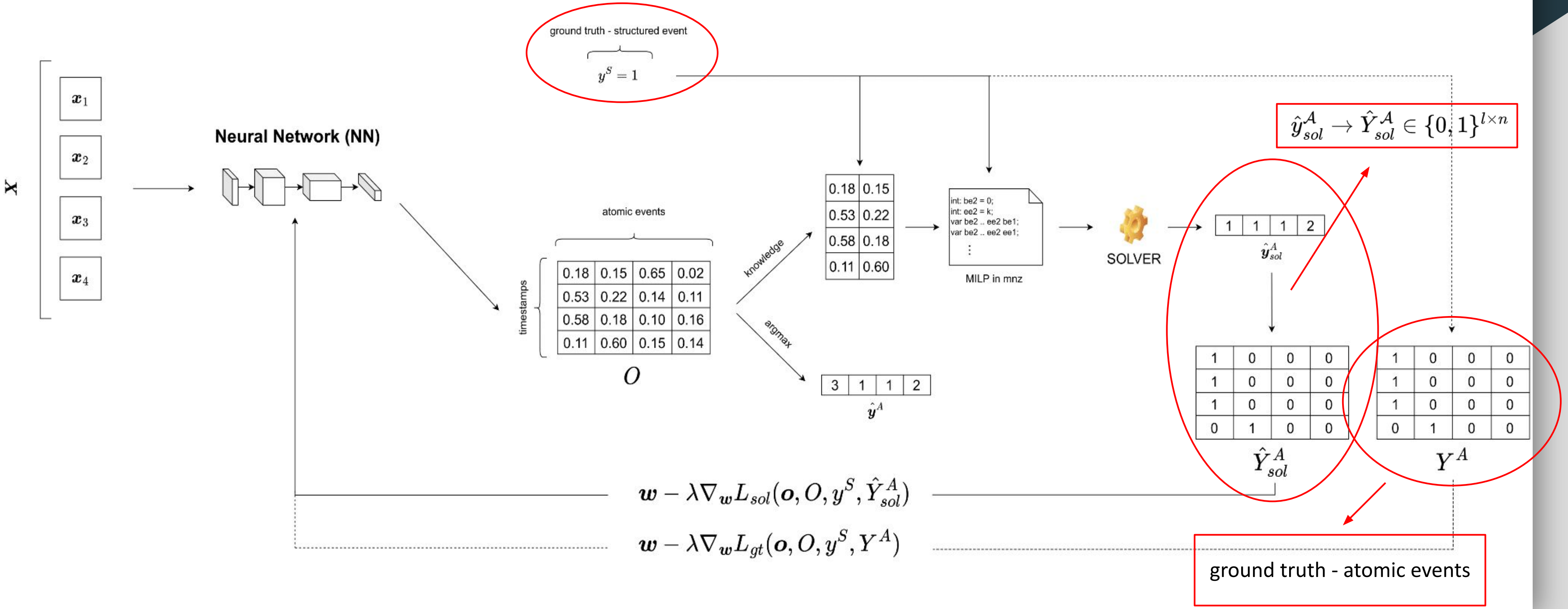
Proposed approach - train



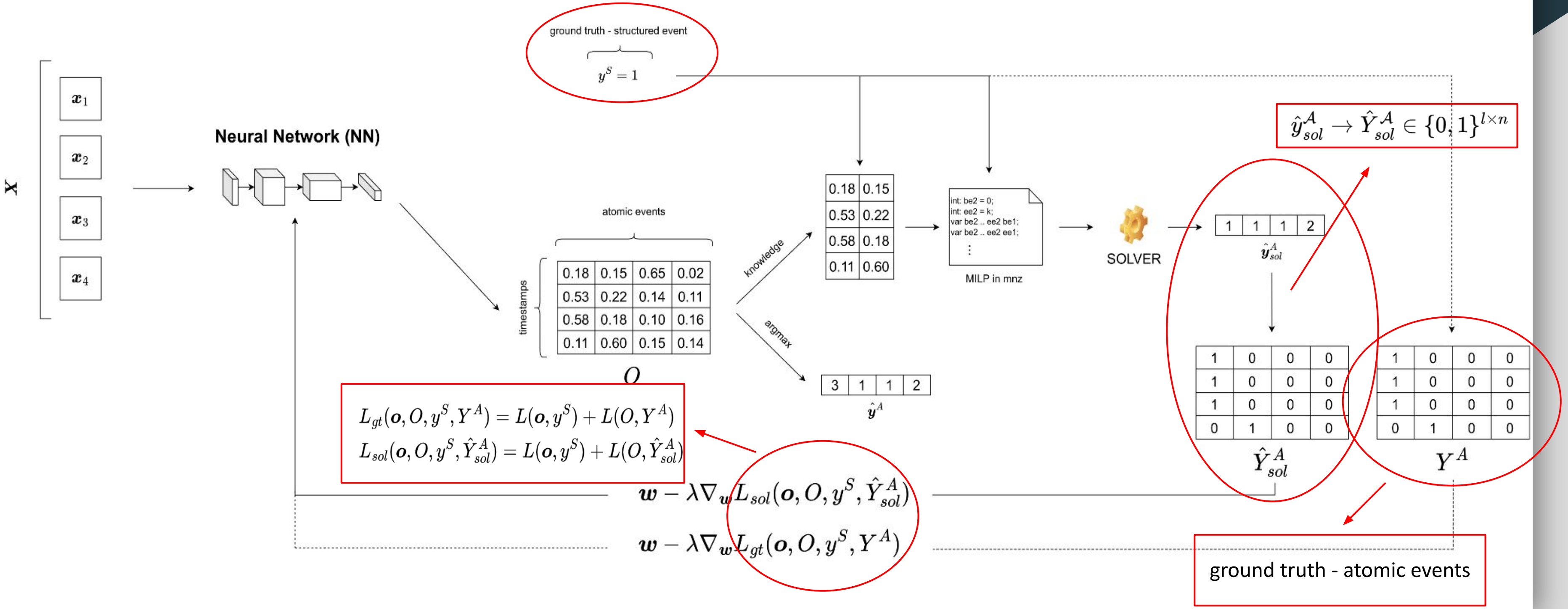
Proposed approach - train



Proposed approach - train



Proposed approach - train



Experimental setting

- **Research question:**

  Does our neuro-symbolic approach lead to an advantage with respect to a fully neural approach for

  (structured and atomic) event recognition using weak and limited supervision?

Experimental setting

- **Research question:**

  Does our neuro-symbolic approach lead to an advantage with respect to a fully neural approach for (structured and atomic) event recognition using weak and limited supervision?

- Clips from Multi-THUMOS dataset

Experimental setting

- **Research question:**

  Does our neuro-symbolic approach lead to an advantage with respect to a fully neural approach for

  (structured and atomic) event recognition using weak and limited supervision?

- Clips from Multi-THUMOS dataset
- Scenario:
  - Clips of different length (only one structured event)
  - Learning -> Fully supervision in terms of structured event and limited (and noisy) labelling:

$$\{happens(highjump, 1, 50),\ happens(run, 1, 31),\ happens(jump, 31, 45),\\ happens(fall, 45, 50)\}$$
$$\{happens(hammerthrow, 1, 30),\ happens(windup, 1, 15),\ happens(spin, 10, 25),\\ happens(release, 25, 30)\}$$
$$\{happens(javelinthrow, 1, 30)\}$$

Experimental setting

- **Research question:**

  Does our neuro-symbolic approach lead to an advantage with respect to a fully neural approach for

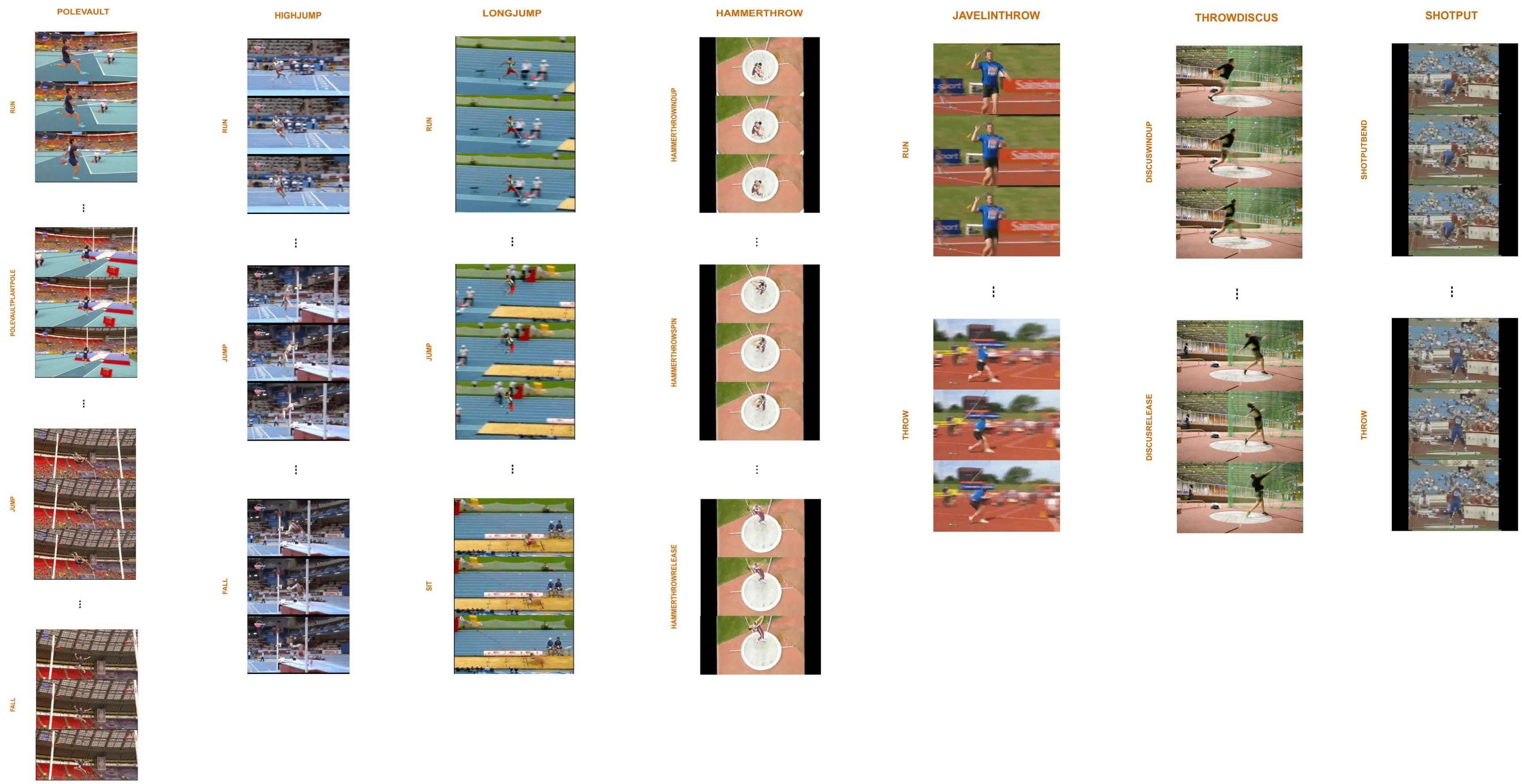
  (structured and atomic) event recognition using weak and limited supervision?

- Clips from Multi-THUMOS dataset
- Scenario:
  - Clips of different length (only one structured event)
  - Learning -> Fully supervision in terms of structured event and limited (and noisy) labelling:

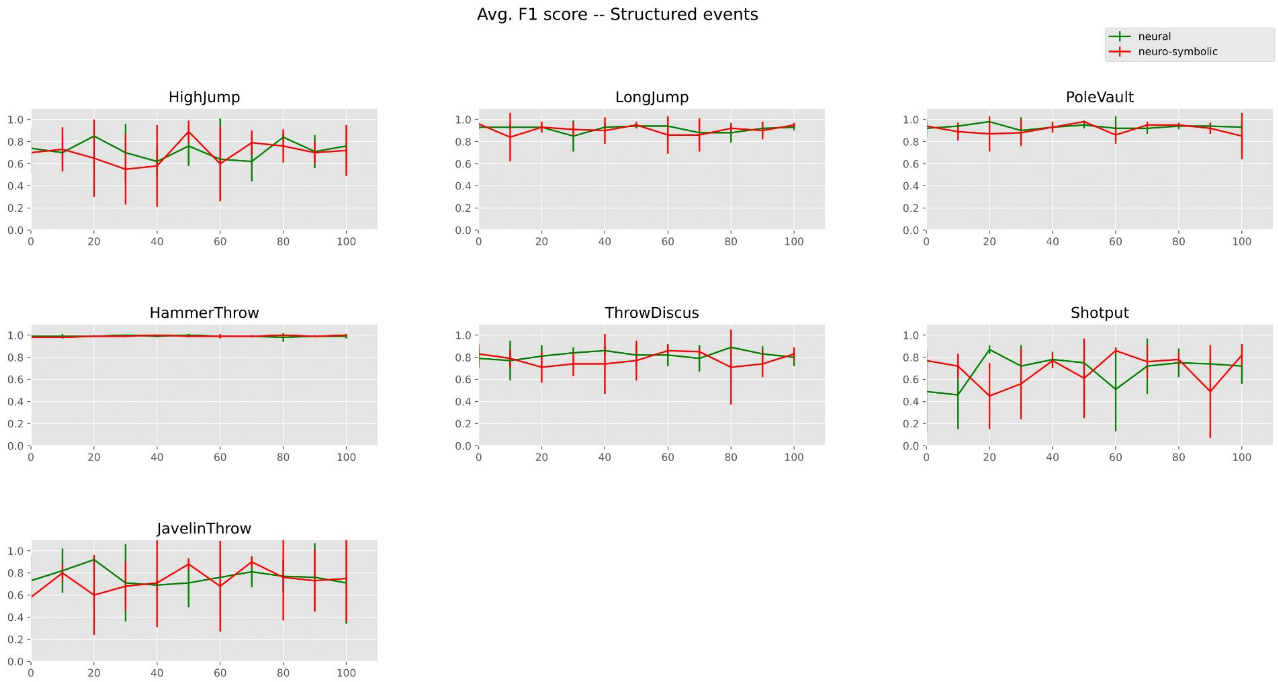
$$\{happens(highjump, 1, 50),\ happens(run, 1, 31),\ happens(jump, 31, 45),\\ happens(fall, 45, 50)\}$$
$$\{happens(hammerthrow, 1, 30),\ happens(windup, 1, 15),\ happens(spin, 10, 25),\\ happens(release, 25, 30)\}$$
$$\{happens(javelinthrow, 1, 30)\}$$

- **How the prediction of structured and atomic events change when increase supervision of atomic events**

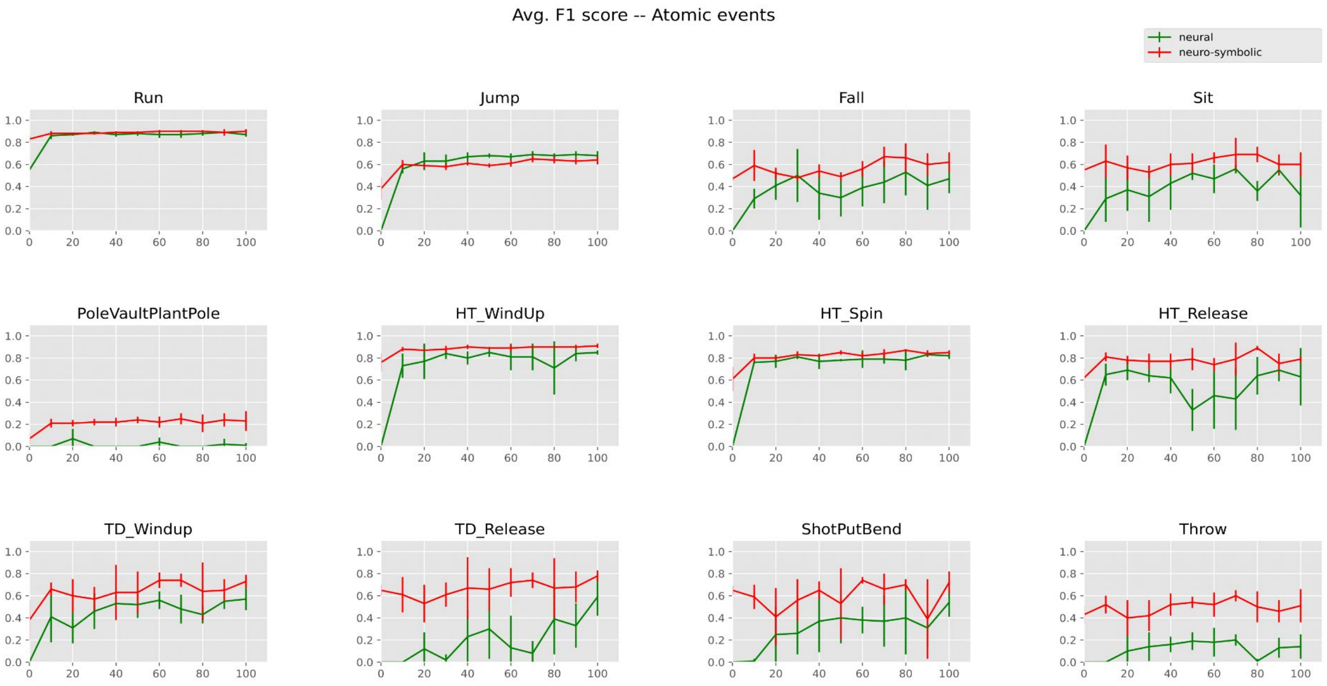
Structured events



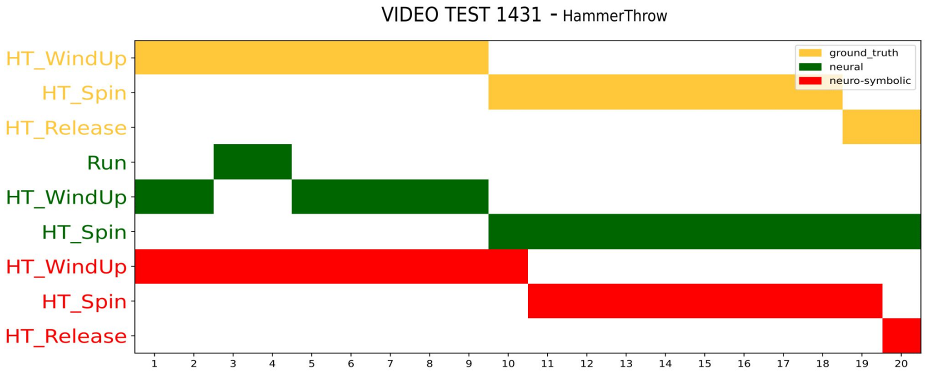
Results - structured events

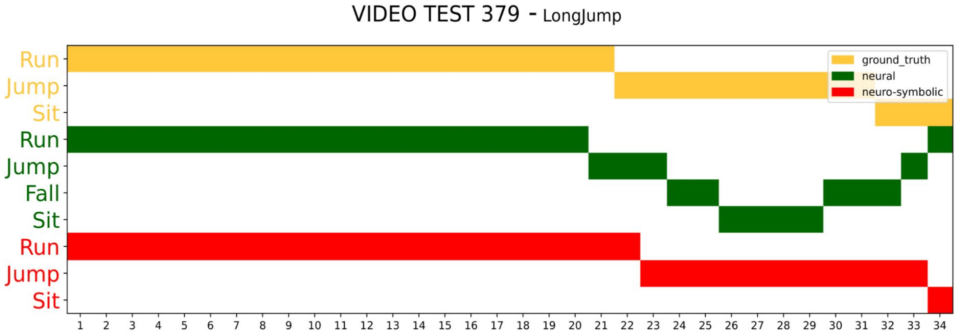


Results - atomic events



Results - predictions





Conclusion and future works

- **Summary:**
  - A Neuro-symbolic approach for (structured and atomic) event recognition exploiting knowledge
  - Real world scenario
  - Our approach outperforms neural baseline in terms of detection of atomic events

Conclusion and future works

- **Summary:**
  - A Neuro-symbolic approach for (structured and atomic) event recognition exploiting knowledge
  - Real world scenario
  - Our approach outperforms neural baseline in terms of detection of atomic events
- **Future works:**
  - Structured events events:
    - Multiple actors
    - More complex relationships (e.g. overlapping of events)

Thank you!

Hard constraints

Generic Constraints (assuming k atomic events)	
$e_i > b_i \quad \forall\, i$	Events should end after they began
$b_1 = 1 \wedge e_k = l$	Sequence of atomic events should span the whole clip
$e_i = b_{i+1} - 1 \quad \forall\, i \in 0 \ldots l-1$	No gap among consecutive events
Specific Constraints (for the *javelinthrow* structured event)	
$a_1 = run \wedge a_2 = throw$	*javelinthrow* is a *run* followed by a *throw*
$d_1 > d_2$	*run* should take longer than *throw*

Example of soft constraint

$$\min(|d_1 + d_2 - max_{run} - max_{jump}|, |d_1 + d_2 - min_{run} - min_{jump}|)$$

where:

$$d_1 = e_1 - b_1 + 1,\ d_2 = e_2 - b_2 + 1$$

$$a_1 = run,\ a_2 = jump$$

$$a_i \in \mathcal{E}$$

$$b_{a_i}, e_{a_i} = begin, end$$

$$d_{a_i} = duration$$

$$max_{a_i} min_{a_i} = max/min\ duration$$

$$(among\ all\ instances)$$

MILP problem

$$f(V,O) = \sum_{(a,b,e)\in V} \left(\sum_{i=b}^{e} O[i,a] - \sum_{j=1}^{b-1} O[j,a] - \sum_{j=e+1}^{l} O[j,a] \right)$$

NN output

$a \in \mathcal{A} \quad b, e \in \mathbb{N}$

$$\underset{V}{\text{minimize}} \quad -f(V,O) + \sum_{j=1}^{m_s} \xi_t c_j(V)$$

$$\text{subject to} \quad h_i(V) \quad \forall\, i = 1, \ldots, m_h$$

Soft constraints

Hard constraints