Integration of symbolic knowledge into DL

HyAIAI : Hybrid Approaches for Interpretable AI

MULTISPEECH, ORPAILLEUR, TAU

Georgios Zervakis

supervised by

Amedeo Napoli, Emmanuel Vincent, Miguel Couceiro

About Me

Background

BSc in Mathematics

MSc in Machine Learning

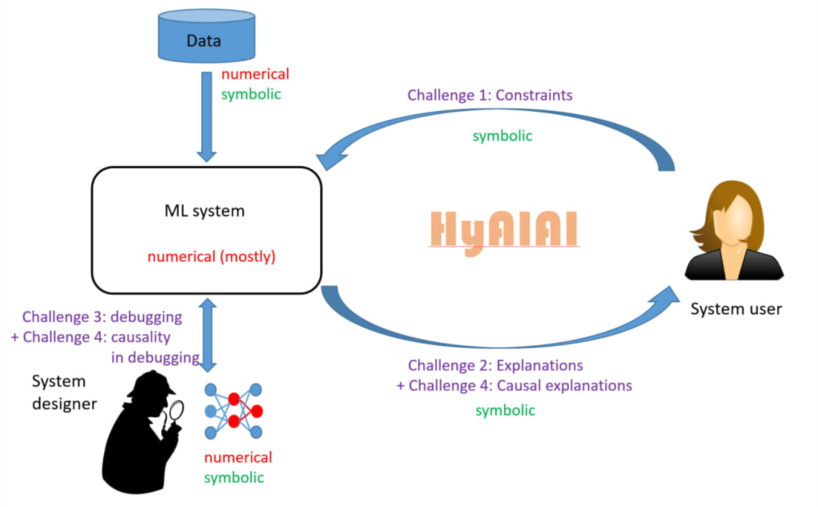
Master Thesis

Multivariate analysis of the parameters in a handwritten digit recognition LSTM system

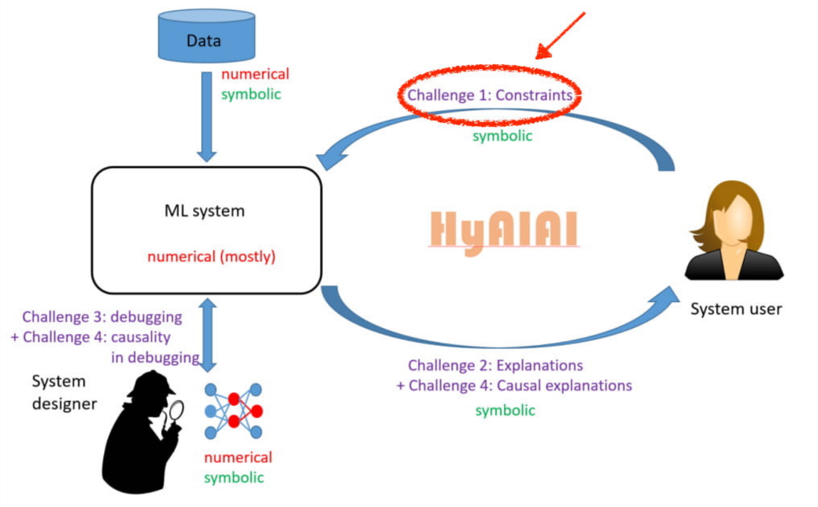
Interests

Artificial Neural Networks, Deep Learning, Explainable AI, Mathematics, Machine Learning, Music Technology

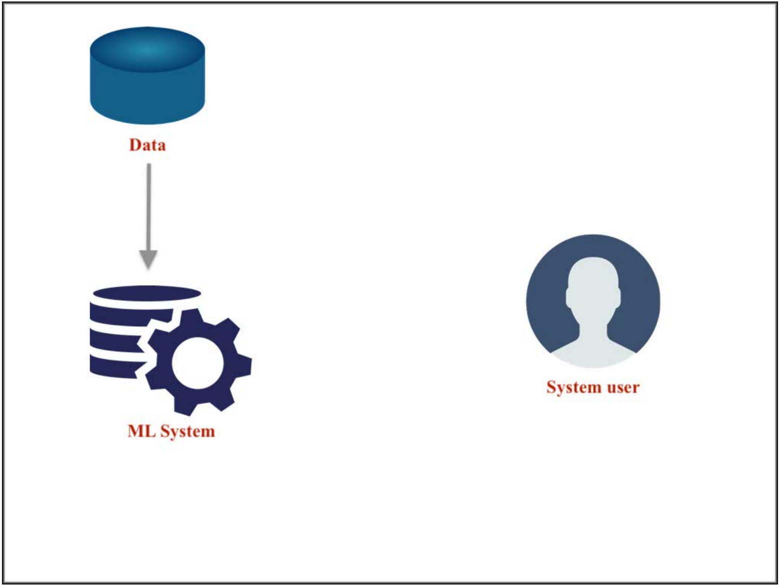
HyAIAI Challenges



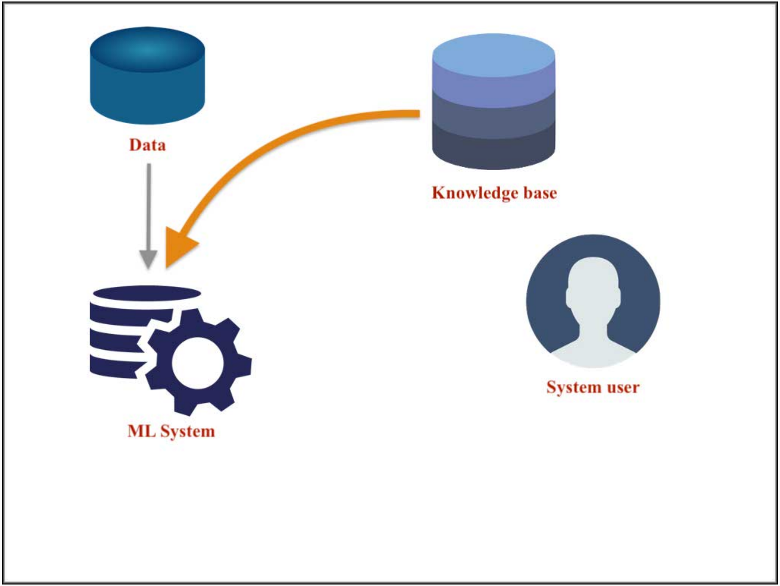
HyAIAI Challenges



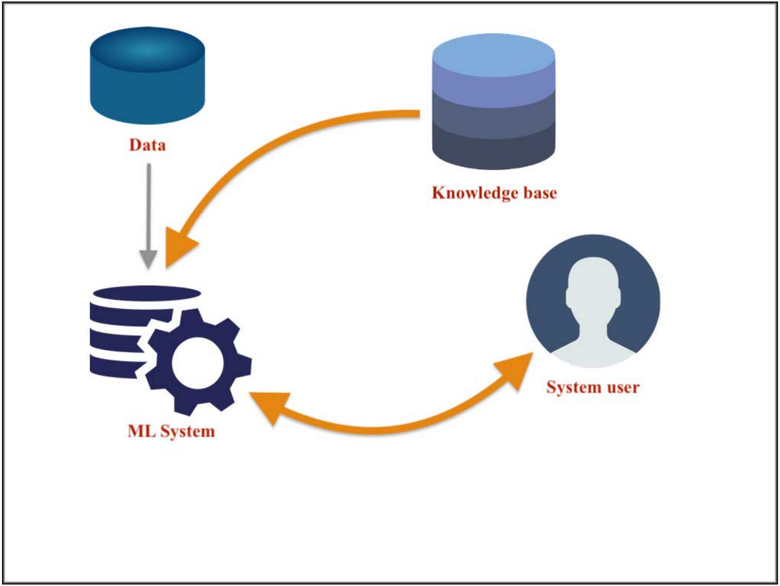
Challenge 1



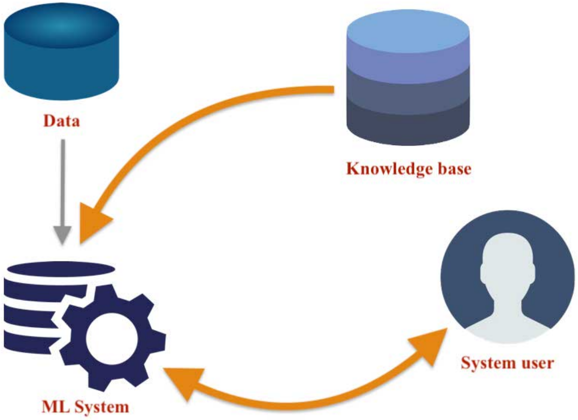
Challenge 1



Challenge 1



Challenge 1



$$\forall x \forall y \; Married(x,y) \Rightarrow (Republican(x) \Leftrightarrow Republican(y))$$

Constraint

Evaluation

The assessment of the system will consist of:

- verifying that the constraints are imposed
- analyzing the impact in terms of error and computational time
- testing the ability of the system to discover concepts and their relations

Evaluation

The assessment of the system will consist of:

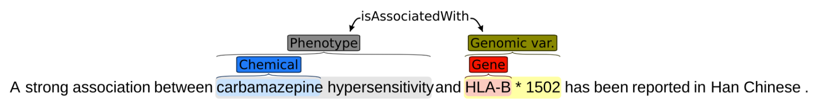
- verifying that the constraints are imposed
- analyzing the impact in terms of error and computational time
- testing the ability of the system to discover concepts and their relations

Experiments will possibly focus on:

- medical text data
- Pharmacogenomics (PGx)
- Pierre Monnin, Jo Legrand, Patrice Ringot, Andon Tchechmedjiev, Clément Jonquet, Amedeo Napoli and Adrien Coulet. PGxO and PGxLOD: a reconciliation of pharmacogenomic knowledge of various provenances, enabling further comparison. BMC Bioinformatics, 2019.

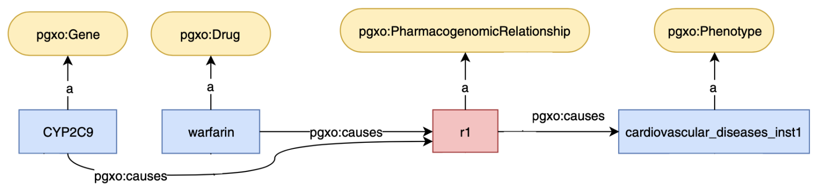
Evaluation

- PGx relationships in the form of triplets = (genomic variation, drug, phenotype)
- knowledge in PGx can be found in knowledge bases, scientific journals and clinical records



Evaluation

- triplet completion: predict a component in the triplet given the other two



Detecting Unseen Visual Relations Using Analogies

Julia Peyre, Ivan Laptev, Cordelia Schmid, and Josef Sivic. Detecting unseen visual relations using analogies. International Conference on Computer Vision (ICCV), 2019.

- Relations in images are represented as triplets $t = (\textit{subject}, \textit{predicate}, \textit{object})$
- Learning representations for such triplets and their individual components
- Moving from existing/known relations to unseens ones using analogies between similar triplets

Detecting Unseen Visual Relations Using Analogies

Julia Peyre, Ivan Laptev, Cordelia Schmid, and Josef Sivic. Detecting unseen visual relations using analogies. International Conference on Computer Vision (ICCV), 2019.

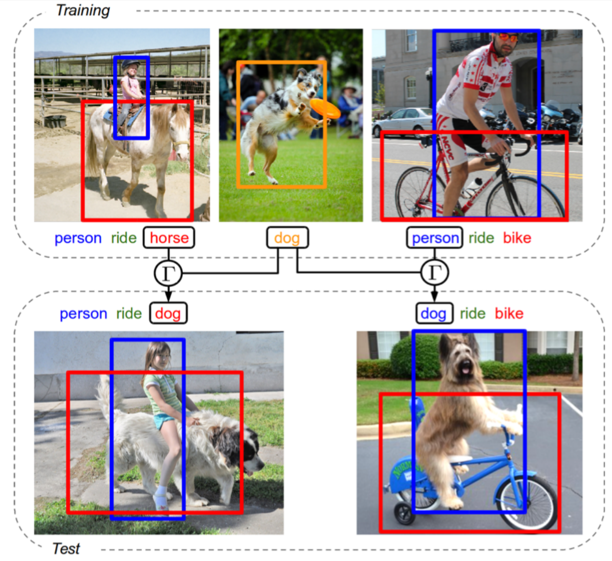
- Relations in images are represented as triplets $t = (subject, predicate, object)$
- Learning representations for these triplets and their individual components
- Moving from existing/known relations to unseens ones using analogies between similar triplets
- **Connection with our task** $\rightarrow$ Ontologies are represented by triplets + dealing with relations/reasoning

Objective

Task: Based on a query $t = (s, p, o)$ retrieve the image described by the triplet

Example: $t = (\textit{person}, \textit{ride}, \textit{dog})$
where training data of the individual components are available but the exact combination is unseen during training

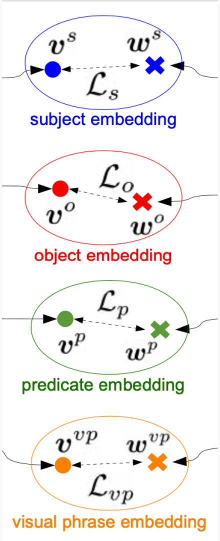
Objective


Training
person ride horse
dog
person ride bike
Γ
Γ
person ride dog
dog ride bike
Test

Learning representation of visual relations

Visual relations are represented in joint visual-semantic embedding spaces:

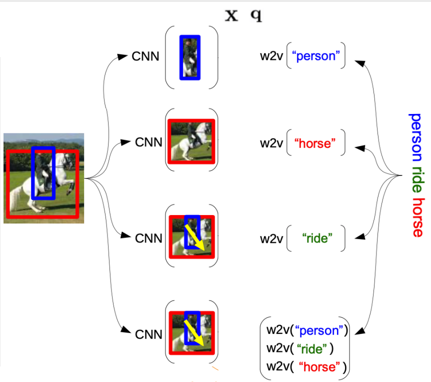
- unigram level:
  separate subject (s), object (o) and predicate (p) embeddings
- trigram level:
  using a visual phrase (vp) embedding of the whole triplet



Learning representation of visual relations

There are two kind of input features:

- visual representation ($\mathbf{x}$) – pre-computed appearance features from CNN object detector
- language representation ($\mathbf{q}$) – pre-trained Word2vec embeddings for each individual entity in $t = (s, p, o)$



Learning representation of visual relations

There are two kind of input features:

- visual representation ($\mathbf{x}$) – pre-computed appearance features from CNN object detector
- language representation ($\mathbf{q}$) – pre-trained Word2vec embeddings for each individual entity in $t = (s, p, o)$

For each input type $b \in \{s, o, p, vp\}$ $\mathbf{x}$ and $\mathbf{q}$ are projected into a common d-dimensional space using:

$$\mathbf{v}_i^b = f_v^b(\mathbf{x})$$
$$\mathbf{w}_t^b = f_w^b(\mathbf{q})$$

Learning representation of visual relations

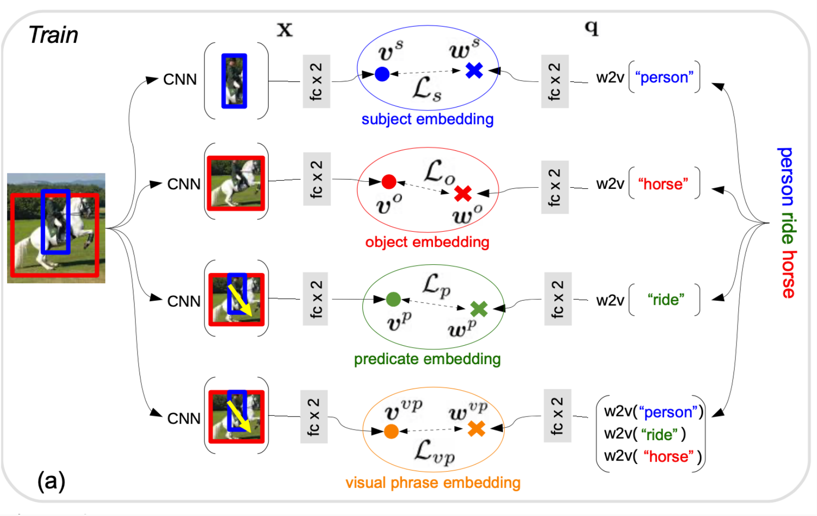
Training consists of optimizing the joint loss:

$$\mathcal{L}_{joint} = \mathcal{L}_s + \mathcal{L}_o + \mathcal{L}_p + \mathcal{L}_{vp}$$

where for $b \in \{s, o, p, vp\}$:

$$\mathcal{L}_b = \sum_{i=1}^{N} \sum_{t \in \mathcal{V}_b} \mathbb{1}_{y_t^i = 1} \log \left(\frac{1}{1 + e^{-\mathbf{w}_t^{b^T} \mathbf{v}_i^b}} \right) + \sum_{i=1}^{N} \sum_{t \in \mathcal{V}_b} \mathbb{1}_{y_t^i = 0} \log \left(\frac{1}{1 + e^{\mathbf{w}_t^{b^T} \mathbf{v}_i^b}} \right)$$

Training Overview



From seen to unseen triplets

Recognize a target triplet $t' = (s', p', o')$ given a source triplet $t = (s, p, o)$ using analogy transformation in the visual phrase embedding space

This is done in 2 steps:

- learning how to perform the transformation from vp_t to $vp_{t'}$
- selecting which visual phrases are suitable for analogy transfer

Transfer by analogy

Given a source triplet $t = (s, p, o)$ and a target triplet $t' = (s', p', o')$:

$$\mathbf{w}_{t'}^{vp} = \mathbf{w}_{t}^{vp} + \Gamma(t, t')$$

Similar to the idea of arithmetic operations with word embeddings:

"king" - *"man"* + *"woman"* = *"queen"*

Here: *"person ride horse"* - *"horse"* + *"cow"* = *"person ride cow"*

Transfer by analogy

Given a source triplet $t = (s, p, o)$ and a target triplet $t' = (s', p', o')$:

$$\mathbf{w}_{t'}^{vp} = \mathbf{w}_{t}^{vp} + \Gamma \begin{bmatrix} \mathbf{w}_{s'}^{vp} - \mathbf{w}_{s}^{vp} \\ \mathbf{w}_{p'}^{vp} - \mathbf{w}_{p}^{vp} \\ \mathbf{w}_{o'}^{vp} - \mathbf{w}_{o}^{vp} \end{bmatrix}$$

For example transforming $t = (\textit{person}, \textit{ride}, \textit{horse})$ to $t' = (\textit{person}, \textit{ride}, \textit{cow})$ will correspond to:

$$\mathbf{w}_{t'}^{vp} = \mathbf{w}_{t}^{vp} + \Gamma \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{w}_{cow}^{vp} - \mathbf{w}_{horse}^{vp} \end{bmatrix}$$

Selecting the right triplets

The selection is based on the cosine similarity of their corresponding subject/object/predicate representations in the embedding space

More specifically:

$$G(t, t') = \sum_{b \in \{s,p,o\}} \alpha_b \mathbf{w}_t^{b^T} \mathbf{w}_{t'}^b$$

where α_b controls the contribution of subject/object/predicate similarities

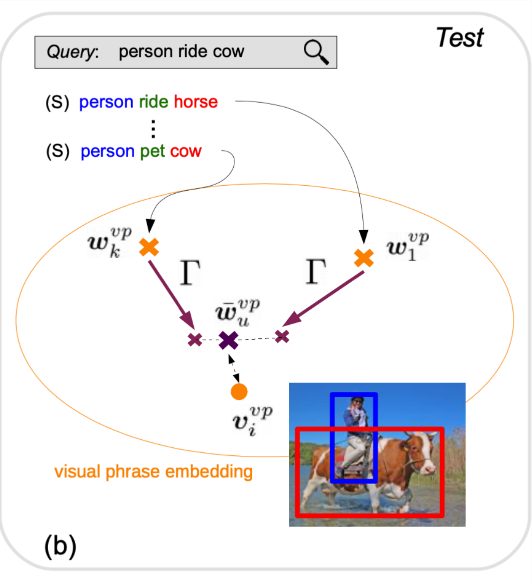
Aggregating embeddings

At test time the visual phrase embedding of an unseen triplet u, vp_u is computed by:

$$\widehat{\mathbf{w}}_u^{vp} = \sum_{t \in \mathcal{N}_u} G(t, u)(\mathbf{w}_t^{vp} + \Gamma(t, u))$$

where $\mathcal{N}_u$ is the set of the k-most similar source triplets according to G

Test Phase



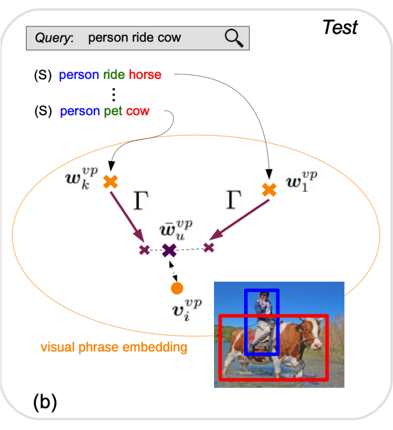
Test
Query: person ride cow
(S) person ride horse
⋮
(S) person pet cow
w_k^{vp}
w_1^{vp}
Γ
Γ
$\bar{w}_u^{vp}$
v_i^{vp}
visual phrase embedding
(b)

Test Phase

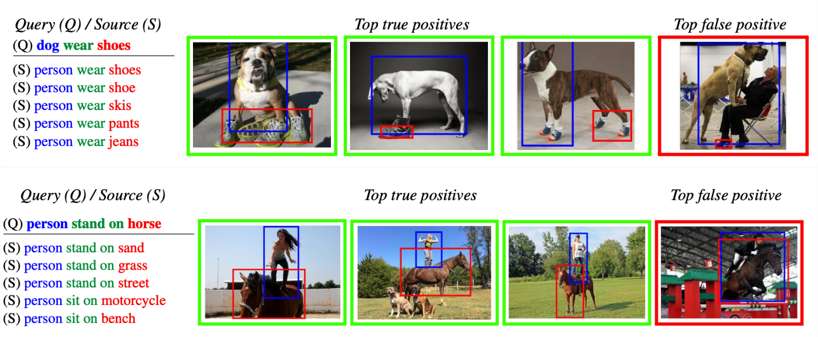
For every image in the test set we compute $\mathbf{v}_i^b$

Then we measure their similarity score with the unseen triplet u as:

$$S_{u,i} = \prod_{b \in \{s,p,o,vp\}} \frac{1}{1 + e^{-\mathbf{w}_u^{b\,T} \mathbf{v}_i^b}}$$



Results


Query (Q) / Source (S)
Top true positives
Top false positive
(Q) dog wear shoes
(S) person wear shoes
(S) person wear shoe
(S) person wear skis
(S) person wear pants
(S) person wear jeans
Query (Q) / Source (S)
Top true positives
Top false positive
(Q) person stand on horse
(S) person stand on sand
(S) person stand on grass
(S) person stand on street
(S) person sit on motorcycle
(S) person sit on bench

Possible extensions towards our task

- use triplets whose subjects or objects are themselves triplets
- use ontology in learning word/visual embeddings to ensure that they meet the rules e.g. the difference between man and woman and between king and queen is perfectly equal, not just approximated
- add a confidence value for every triplet: probability that a relationship is possible from existing relationships and similarities between subjects, objects or predicates (from existing ontology or from the data)