MALHEPHYCA

MAchine Larning High Energy PHYsics Classification Algorithms

CdS - Preventivi 2019
Genova - 2 July 2018

Riccardo Torre

INFN Genova, CERN

Istituto Nazionale di Fisica Nucleare

General Information

1. Project name

Machine Learning High Energy Physics Classification Algorithms (MALHEPHYCA)

2. Struttura

INFN Sezione di Genova

3. CSN di riferimento

IV

4. Timing

The project will start August 1st 2018 and will end July 31st 2021

ML & HEP

- Machine Learning (in its different incarnations) is becoming ubiquitous in our lives
- This is becoming true also in High Energy Physics (HEP)
- *Experimental* applications range from triggers to reconstruction algorithms
- *Theory* applications go from jet clustering to constraining new physics
- Existing ML algorithms are inspired by solving "practical" problems (classification, regression, object detection/segmentation, data generation, auto-captioning, game playing, translation, question answering, etc.)
- Different algorithm optimized to solve different problems
- Our questions/problems are new to the ML world and offer new challenges
- Three possible approaches

1. Try to adapt existing ML techniques/algorithms to existing HEP questions

   Physicist approach

2. Develop new algorithms to solve HEP problems

   Data scientist approach

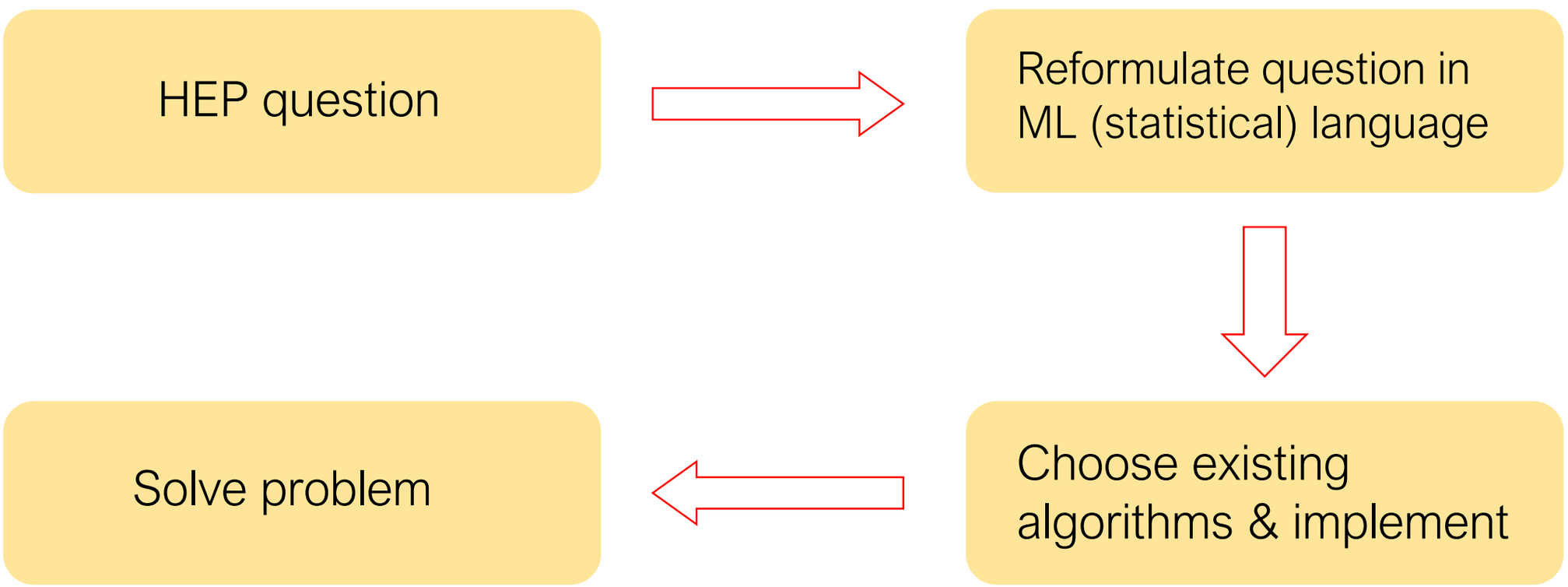
3. Optimize/revise HEP questions to make them similar to the problems ML is used to address (optimize the questions)

   More recent approach

Approaches

- In order to optimize the approach one should be able to take both the perspective of the physicist and the one of the data scientist
- Typically physicists learn ML starting from physics, and data scientists learn physics starting from ML, therefore interface is not perfect
- The scope of the project is to explore ML from the data science point of view and try to bring experience into our field

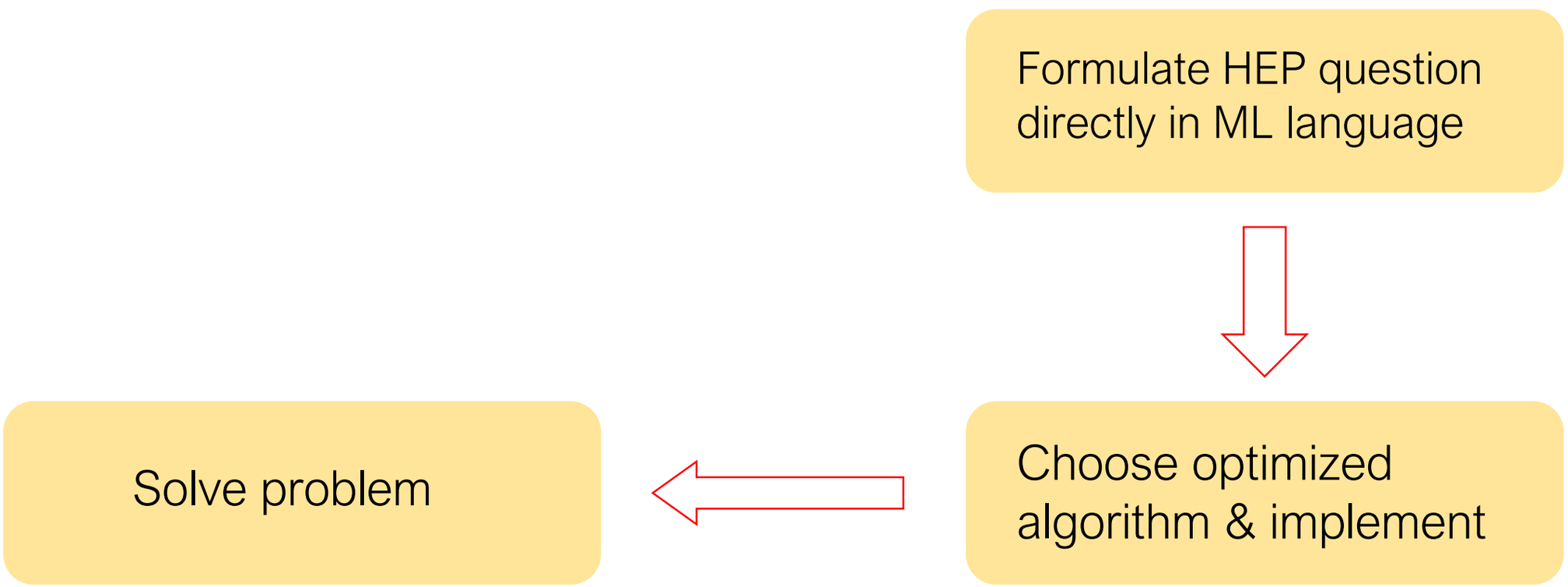
Basic approach



Approaches

- In order to optimize the approach one should be able to take both the perspective of the physicist and the one of the data scientist
- Typically physicists learn ML starting from physics, and data scientists learn physics starting from ML, therefore interface is not perfect
- The scope of the project is to explore ML from the data science point of view and try to bring experience into our field

Improved approach



Project objectives

1. Training

Acquire the competences necessary to the construction and training of complex neural networks from the data science point of view

2. Proof of concept

Optimize the HEP problems and apply acquired knowledge to a simple search of new physics, such as resonance searches

3. Application/optimization

Build and train a complex neural network optimized to search for new physics at the LHC

Evaluation parameters

1. **Comparison with state of the art**

Comparison with the state of the art of the applications of ML techniques to new physics research

2. **Impact (both research and training)**

Impact on the related community both in terms of research and training

3. **Efficiency of implementation**

Implementation of the project into an efficient program for new physics research at the LHC

Timeline

2018

2019

2020

2021

- Project starts:

  Training part

  Acquire the necessary knowledge on data science and machine learning, possibly collaborating with data scientists (DIBRIS?)

  Set up of the computing resources

- Start implementing simple neural networks for the treatment of massive datasets

  Training of the first simple NN and first applications to HEP

  Exploration of the different classes of questions vs machine learning techniques

- Start the build of a more complex neural network to address HEP questions

  Optimize the problem of searching for new physics

  Design and train the NN

  Interface with LHC experiments to apply the algorithms to the search of new physics with the full LHC dataset

THANK YOU!

Riccardo Torre

MALHEPHYCA