Method selection algorithm

Mathieu Giamberini, 26.03.2024

Introduction

There is a lot of AD method, and base on the available time it's not possible to make a review on all of them, and even if it was, the paper will be to long. The goal is then to find an algorithm to choose them.

1 Problem definition

Thanks to the work made the past few weeks there exist a taxonomy of 100 and so method ordered as a tree structure. Let consider the entity of thr tree is of interest to us (if not we just need to filter out the one witch's not), to maximized the usefulness, of the review we would need a list of method, which is general enough not go in un wanted detail and broad enough to be have a lot of different technic to choose from. Base of that, we can find an algorithm that rationalize this idea.

2 Definition and Core algorithm

Definition 1. *In our context a* ***loss function*** *will defied as a function which want to be minimized and which take a set of node and a node alone as input and a real number as output (e.i.* $L : \mathcal{P}(T) \times T \to \mathbb{R}$*)*

Base on l we can choose the next methods recessively by minimizing the loss function, in other words, with M_i the set of chosen method at the ith iteration:

$$M_i = \{\arg\min_{n \in T} \ L(M_{i-1},\, n)\} \cup M_{i-1}$$

3 Loss function

Here could be a good example of loss function :

$$(M,\, n) \mapsto \sum_{m \in\, M \cup \{\mathrm{AD}\}} \frac{1}{\delta(n,t)^2} + \lambda \log(\delta(n, \mathrm{AD}) + 1)$$

Where δ is the distance between to node in the tree, AD the root method (Anomaly detection) and λ a parameter.