

38. 【X=1.3】随着目前数据的爆炸式增长，对于同一类任务有很多不同的算法被研究出来，（例如：分类、回归）但是对于具有不同特点的数据集而言，并不是所有算法最终的性能都很好，算法之间最后的性能差异有很大的浮动。所以，如何根据数据集和算法的特点，为每个数据集尽可能地选择最适合它的算法成为目前的研究热点。

构建一个模型（可以参考元学习的思想（meta-learning）），使得其训练后可以：

- A. 针对时间序列分类任务，为新任务自动推荐最适合它的算法
- B. 尽可能提升整个模型的自动化程度，并在实验报告中说明详细。（例如：如果采用 meta-vector 来代表每个历史任务，那么设计一种方法可以自动从一个 candidate meta-feature list 中选择效果最好的 meta-features 来构成 meta-vector）
- C. 选择一个自己熟悉的主流大数据框架，使用这个框架实现上述功能，深入思考哪个部分可以采用并行处理，并提升其效率。不要为了使用框架而使用框架，如果使用框架效率降低，将实验结果写明，详细分析原因，写入实验报告。

说明：时间序列分类算法的实现可以采用现有的算法库，例如：sk-time (python)，tsml (java) 等，待选择的算法不少于 15 种。实验数据集可以从 UCR 公开数据集下载，自己进行适当的数据集格式转换。对于算法自动选择来说，不要走入“一个模型可以找到所有最优算法”的误区。

我们设计的模型只能达到尽可能接近最优的算法。

除了工程文件和实验报告之外，你需要提交一份说明文档，描述你的算法设计、算法实现和实验结果。不必将思路局限于题目中提及的论文，可以考虑更合理的方式方法。

Reference:

- [1] Time Series Classification Website
- [2] Welcome to sktime — sktime documentation
- [3] TSML • Julia Packages
- [4] HOME • TSML Documentation (ibm.github.io)