

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC CẦN THƠ
KHOA CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG
BỘ MÔN KHOA HỌC MÁY TÍNH

-----❖-----



BÁO CÁO ĐỒ ÁN
KHAI KHOÁN DỮ LIỆU

TÌM HIỂU MÁY HỌC TỰ ĐỘNG

CÁN BỘ HƯỚNG DẪN
TS. Lưu Tiến Đạo

SV THỰC HIỆN:
Nguyễn Minh Nghiên
MSSV: B1304993

Tháng 07/2020

MỤC LỤC

	<i>Trang</i>
<i>MỤC LỤC</i>	<i>ii</i>
<i>TÓM TẮT ĐỀ TÀI</i>	<i>iii</i>
CHƯƠNG I	4
GIỚI THIỆU	4
1.1 ĐẶT VẤN ĐỀ	4
1.2 GIẢI QUYẾT VẤN ĐỀ	4
1.3 CÔNG NGHỆ SỬ DỤNG	4
CHƯƠNG II	5
MÁY HỌC TỰ ĐỘNG	5
2.1. Máy học tự động.....	5
2.2. Một số máy học tự động.....	5
2.2.1. Auto-sklearn	5
2.2.2 Auto-keras	5
2.2.3 Auto-pytorch	6
2.2.4 Máy học tự động H2o	6
2.3. Cài đặt.....	6
2.2.1. Auto-sklearn	6
2.2.2 Auto-keras	7
2.2.3 Auto-pytorch	7
2.2.4 Máy học tự động H2o	7
CHƯƠNG III	8
THỰC NGHIỆM	8
3.1. Chuẩn bị dữ liệu	8
3.2 KẾT QUẢ	8
CHƯƠNG IV	9
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	9
4.1. Kết luận.....	9
4.2. Hướng phát triển.....	9
LỜI CẢM ƠN	10

TÓM TẮT ĐỀ TÀI

Máy học là một thuật ngữ không còn xa lạ với những kỹ sư trong lĩnh vực công nghệ thông tin, nói chính xác hơn là kỹ sư ngành khoa học máy tính. Bằng cách tự học tập và phát triển bằng những kinh nghiệm học được để tìm kiếm, phát hiện một số thuộc tính đặc biệt hay phân loại dữ liệu. Máy học đã trở thành một công cụ không thể thiếu trong sự phát triển các phần mềm mới tự động hướng tới AI, trong học tập cũng như trong tất cả các lĩnh vực nghiên cứu khoa học.

Muốn sử dụng máy học thì cần kiến thức về khoa học máy tính, tốn quá trình học tập dài hạn, nên gây hạn chế cho phần lớn người có nhu cầu sử dụng machine learning nhưng không có kiến thức về khoa học máy tính. Vì thế, cấp thiết cần một giải pháp để có thể làm đơn giản quá trình sử dụng máy học nhưng không làm giảm hiệu quả.

Do đó, máy học tự động ra đời để đáp ứng sự phát triển của máy học. Máy học tự động được hiểu là tự động sử dụng máy học trong quá trình học tập và tối ưu các tham số để máy học hiệu quả được tốt hơn. Vì thế chúng tôi muốn tìm hiểu về máy học tự động để giới thiệu cho mọi người có thể dễ dàng sử dụng. Qua đó, chúng tôi giới thiệu một số AutoML như: Auto-sklearn, Auto-Pytorch, AutoML h2o, Auto-keras.

Cần Thơ, ngày tháng năm 2020

Sinh viên thực hiện

Nguyễn Minh Nghiên

CHƯƠNG I

GIỚI THIỆU

1.1 Đặt vấn đề

- Với sự phát triển của nhiều sản phẩm ứng dụng máy học
- Cần một số lượng lớn kỹ sư máy học
- Máy học được sử dụng rộng rãi cho tất cả các lĩnh vực, cần số lượng lớn kỹ sư máy học hiểu biết về các lĩnh vực cần máy học.
- Máy học tự động xuất hiện nhiều năm nhưng chưa được sử dụng rộng rãi, do chưa có nhiều người biết đến máy học tự động.

Vì vậy cần phải tìm hiểu về thư viện máy học tự động để nhiều người hiểu biết và có thể sử dụng các máy học tự động cho mục đích của mình.

1.2 Giải quyết vấn đề

- Tìm hiểu các một số thư viện máy học tự động như: cách sử dụng, cài đặt.
- Sử dụng một số tập dữ liệu để kiểm tra hiệu quả của các thư viện
- So sánh sự khác nhau của máy học và máy học tự động

1.3 Công nghệ sử dụng

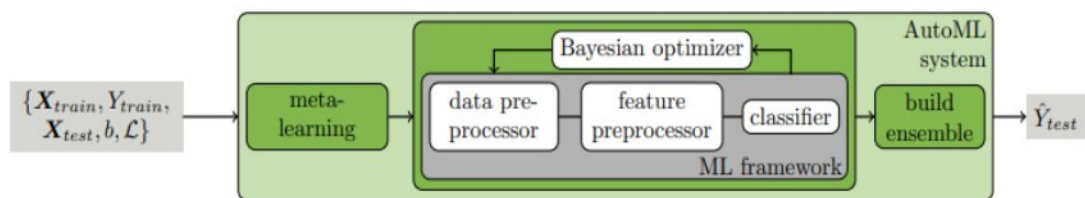
- Python: ngôn ngữ lập trình hướng đối tượng, cấp cao, dễ dàng sử dụng
- Google colab: dịch vụ đám mây miễn phí nhằm hỗ trợ cộng đồng nghiên cứu AI phát triển các ứng dụng deep learning bằng việc cung cấp GPU và TPU miễn phí

CHƯƠNG II

MÁY HỌC TỰ ĐỘNG

2.1. Máy học tự động

Máy học là quá trình hệ thống thực hiện một số công việc từ những kinh nghiệm đã học từ trước mà không cần lập trình cụ thể. Máy học tự động là sự tiếp nối của máy học trong quá trình hoàn thiện hệ thống hướng trí tuệ nhân tạo. Máy học tự động làm đơn giản hóa quá trình sử dụng máy học. Người sử dụng chỉ cần cung cấp dữ liệu và nạp vào máy học tự động thì máy học tự động sẽ cho ra kết quả cuối cùng là mô hình phù hợp.



Hình 2.1: Hình mô tả máy học tự động

2.2. Một số máy học tự động

Giới thiệu về một số đặt điểm của một vài máy học tự động

2.2.1. Auto-sklearn (Style “Cap 3”)

Auto-sklearn là sự thay thế của scikit-learn được tạo ra bởi thành quả của các nhà nghiên cứu tại Đại học Freiburg. Nó giải phóng người dùng máy học khỏi lựa chọn thuật toán và điều chỉnh siêu tham số. Nó tận dụng những lợi thế gần đây trong tối ưu hóa Bayes, siêu học và xây dựng toàn bộ.

Thay vì tập trung vào việc tìm kiếm các kiến trúc nơ-ron cho các hệ nơ-ron sâu, Auto-Scklearn sử dụng tối ưu hóa Bayesian cho siêu tham số, nhằm tương thích với các thuật toán machine learning “truyền thống” trong scikit-learn.

2.2.2 Auto-keras

Auto-keras là một thư viện mã nguồn mở được viết bằng ngôn ngữ python với việc sử dụng thư viện keras cho quá trình học sâu. Cung cấp việc tự động tìm kiếm các

kiến trúc nơ-ron và siêu tham số. Ngoài ra sử dụng tối ưu hóa Bayesian để tăng độ chính xác cho máy học.

Máy học tự động Auto-keras cung cấp nhiều tác vụ cho các bài toán phân loại và hồi quy, với nhiều dạng dữ liệu đa dạng như: văn bản, hình ảnh, dữ liệu có cấu trúc. Nó được phát triển bởi DATA LAB tại đại học Texas A&M và nhóm keras-team. Sự khác biệt thật sự của các phiên bản thể hiện rõ rệt nhất là phiên bản 1.0.2 vào tháng 2/2020 chỉ tương thích với tensorflow 2.1.0 so với phiên bản mới nhất hiện tại 1.0.3 vào tháng 6/2020 đã sửa lỗi không tương thích với tensorflow 2.2.0 cũng như với sự đột phá trong việc sử dụng các lớp tiền xử lý Keras cho ImageAug sắc tố.

Do đó việc sử dụng auto-keras đơn giản cho người không có kiến thức chuyên sâu về máy học. Với việc điều chỉnh kiến trúc mô hình phù hợp với từng kiểu dữ liệu và thời gian huấn luyện máy học tự động sẽ làm mọi chuyện còn lại để xuất ra mô hình tối ưu nhất.

2.2.3 Auto-pytorch

Auto-pytorch là một máy học tự động sử dụng thư viện pytorch của facebook với việc tự động lựa chọn kiến trúc và cài đặt siêu tham số, còn hơn thế auto-pytorch còn tối ưu tham số với 2 phương thức tối ưu hóa là đa chân thực và Bayesian (BOHB). Việc này đem lại một dự báo về một máy học tự động ưu việt hơn trong thời gian tới, bởi vì hiện tại auto-pytorch vẫn chưa ra bản chính thức mới chỉ là những bước đầu.

2.2.4 Máy học tự động H2o

Máy học tự động h2o được viết bằng Java nên dễ dàng thực hiện các tác vụ đa luồng, phù hợp nhiều hệ điều hành khác nhau, tích hợp nhiều giao diện với các ngôn ngữ lập trình khác nhau. Với việc đặt giới hạn về thời gian hoặc hoạt động cụ thể, h2o sẽ trả về danh sách các mô hình tốt nhất

2.3. Cài đặt

Trước hết nên sử dụng các phần mềm tạo môi trường ảo để tránh việc đụng độ giữa các thư viện. vd: anaconda

2.2.1. Auto-sklearn

Yêu cầu cấu hình:

- Cài đặt python
-

- Không cài đặt được trên window

Cú pháp cài đặt: `pip install autosklearn`

Sử dụng colab:

`!apt-get install swig -y`

`!pip install Cython numpy`

`!pip install auto-sklearn`

2.2.2 Auto-keras

Yêu cầu cấu hình

- Python 3.5 trở lên

- Tensorflow 2.1.0 với autokeras 1.0.2

- Tensorflow 2.2.0 với autokeras 1.0.3

Cú pháp: `pip install autokeras`

2.2.3 Auto-pytorch

Cú pháp:

`pip install torch`

`pip install git+https://github.com/shukon/HpBandSter.git`

`pip install git+https://github.com/automl/Auto-PyTorch.git`

2.2.4 Máy học tự động H2o

Cú pháp: `pip install h2o`

CHƯƠNG III

THỰC NGHIỆM

3.1. Chuẩn bị dữ liệu

- Iris của sklearn với 150 hàng, 4 thuộc tính phân loại 3 loại hoa.
- Digits của sklearn 1797 hàng, 64 thuộc tính, phân loại 10 chữ số viết tay từ 0-9
- Nmist của keras gồm 60000 bức hình dùng để train, 10000 bức hình dùng để test

3.2 Kết quả

	Auto-keras	Auto-pytorch	Auto-sklearn	h2o
iris	0.98	1.0	0.97	0.98
digits	0.59	0.98	0.99	0.94
nmist	0.98	NaN	0.96	

CHƯƠNG IV

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

4.1. Kết luận

- Auto-sklearn dễ sử dụng, độ chính xác cao, xử lý dữ liệu 3D không được, chuyển sang 2D cho độ chính xác cao. nmist
- Auto-pytorch dữ liệu nhỏ cho độ chính xác cao, dữ liệu lớn không tìm được kết quả (nmist)
- Auto-keras dữ liệu hơi lớn digits cho độ chính xác thấp.
- Dễ sử dụng nhất là auto-sklearn
- Auto-keras hỗ trợ phân biệt nhiều kiểu dữ liệu khác nhau

4.2. Hướng phát triển

- Tìm hiểu thêm các AutoML mới như: cnvrg.io, AutoFolio, Robo, Flexfolio
 - Sử dụng AutoML để tạo phần mềm
-

LỜI CẢM ƠN

Tôi xin chân thành cảm ơn thầy Lưu Tiến Đạo đã luôn quan tâm chỉ dạy, theo dõi, giúp đỡ tận tình trong suốt khoảng thời gian em thực hiện đề tài.

Và hơn hết, Tôi xin bày tỏ lòng biết ơn trân trọng đến quý thầy cô trường Đại học Cần Thơ đã tận tình chỉ dạy, truyền đạt những kiến thức quý báu cho chúng em trong thời gian vừa qua để chúng em có đủ kiến thức, điều kiện để thực hiện đề tài này. Đồng thời chúng em cũng rất biết ơn các cán bộ trực ở thư viện khoa công nghệ, trung tâm học liệu, phòng máy... đã hỗ trợ giúp đỡ chúng em trong thời gian qua.

Đồng cảm ơn đến các tác giả trong các quyển sách báo, internet, anh chị đi trước đã tìm tòi, nghiên cứu đúc kết kinh nghiệm làm tài liệu để em có thể tham khảo trong quá trình thực hiện đề tài.

Sau cùng tôi xin cảm ơn các bạn cùng lớp Cơ khí chế tạo máy, khoa Công nghệ, trường Đại học Cần Thơ đã tận tình giúp đỡ, hỗ trợ cho tôi thực hiện tiểu luận này.

Cần Thơ, ngày tháng năm 2020

)

(Ký và ghi rõ họ tên)

Nguyễn Minh Nghiên
