Đồ án môn học

Học kỳ 192

Đề tài: Phân loại rượu dựa trên số liệu đo đạt có sẵn bằng cách sử dụng tensorflow

Kế hoạch làm việc: chia làm 4 giai đoạn chính

1. Thu thập dữ liệu, xử lý dữ liệu thô (đưa tất cả về file .csv), ...
2. Lựu chọn phương pháp phân loại tối ưu: thử nghiệm vài loại
3. Tiến hành viết code, làm bản báo cáo giữa kỳ.
4. Xem xét tối ưu hóa, hoàn thành báo cáo.

Thời gian: từ ngày 11/3/2020 đến ngày 31/4/2020

* Học thêm về tensorflow xuyên suốt trong quá trình làm
* Bắt đầu thu thập dữ liệu từ ngày 11/3
* Xử lý, phân loại, … trong ngày 12/3
* Lựa chọn phương pháp (có thể hỏi trong nhóm ML) (trong 2 ngày 13/3 và 14/3)
* Tìm tài liệu: ngày 15/5
* Bắt đầu viết code ngày 17/3
* Làm báo cáo, lý thuyết, lời cảm ơn, …
* Cơ bản hoàn thành trước ngày 30/3

Vấn đề:

1. Phân loại rượu khi không pha tạp và khi có pha tạp

* Chỉ có hai loại là Volka và nếp
* Thông số: bước sóng quang phổ, nồng độ

1. Phân loại các loại nước uống:

* Có sẵn bài phân loại

1. Phân loại Coca Cola và Pepsi

* Có các thông số về Protein, energy, fat, …
* Thông số khi pha trộn 2 loại này (bước sóng)
* Thông số khi pha 2 loại này với nước (bước sóng)

Khó khăn:

* Chưa thành thạo code tensorflow
* Chưa đọc được file số liệu
* Cần chuyển toàn bộ về file csv
* Số liệu đã qua xử lý và chuẩn hóa

Cần làm:

* Tìm hiểu phương pháp đã sử dụng trong bài phân loại nước
* Đọc hiểu số liệu
* Tìm bài code tensorflow mẫu về phân loại

Ứng dụng của máy quang phổ UV-Vis

Một trong những ứng dụng phổ biến nhất của máy quang phổ là để xác định nồng độ mẫu lỏng.Trong thực tế, một loạt các mẫu chuẩn được chuẩn bị. Một mẫu chuẩn là một mẫu mà nồng độ phân tích được biết một cách chính xác. Phổ hấp thụ của các mẫu chuẩn đã đo lường và được dùng làm đường cong hiệu chỉnh, mà trong trường hợp này là một đồ thị của độ hấp thụ và nồng độ. Các điểm trên đường cong hiệu chỉnh nên cân chỉnh thẳng hàng.

Máy quang phổ UV-Vis được sử dụng trong phân tích lâm sàng, phòng thí nghiệm hóa học dầu mỏ, thí nghiệm hóa học và sinh học cũng như các phòng quản lý chất lượng như kiểm soát môi trường, kiểm soát nước, thực phẩm và nông nghiệp, kiểm soát chất lượng nghành công nghiệp đồ uống, y học…

Quang phổ của alcohol:

<https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Spectroscopy/Vibrational_Spectroscopy/Infrared_Spectroscopy/Identifying_the_Presence_of_Particular_Groups>

You will see that it contains the following bonds:

carbon-oxygen double, C=O: 1680 - 1750 cm-1

carbon-oxygen single, C-O: Phổ nằm trong khoảng 1000 đến 1300 cm-1

oxygen-hydrogen, O-H: 2500 - 3300 cm-1

carbon-hydrogen, C-H: 2853 - 2962 cm-1

carbon-carbon single, C-C

Phương pháp: lựa chọn 2 phương pháp CNN và RNN

***Đối với bài nhận dạng loại nước:***

* Sử dụng CNN với thuật toán SGDM, ADAM
* Sử dụng Deep Belief Network

Đã có:

* File csv của các mẫu coca cola và pepsi: cột ngang là bước sóng từ 180 đến 850 nm. Cột dọc là kết quả đo intensity theo thời gian
* File mat của các mẫu nước uống. Cột ngang là bước sóng từ 180 đến 400. Cột dọc là kết quả 100 lần đo.
* File csv của các mẫu rượu

Tài liệu tham khảo:

<http://reu.rnet.missouri.edu/REU07-18/REU16/Projects/Team5/Presentation.pdf>

<https://www.nature.com/articles/s41598-017-12853-y> (xác định chất lượng nước dựa trên quang phổ)

<https://onlinelibrary.wiley.com/doi/full/10.1002/advs.201801367> (ứng dụng quang phổ và deep learning)

<https://www.nature.com/articles/ncomms13890> (deep tensor neural network)

<https://www.hindawi.com/journals/cin/2019/3563761/> (ví dụ xác định lại đất qua quang phổ)

BÁO CÁO:

Phần giới thiệu:

Loài chim truyền cảm hứng cho chúng ta tạo ra máy bay,loài cá truyền cảm hứng cho chúng ta làm ra những con thuyền, thiên nhiên đã truyền cảm hứng cho rất nhiều phát minh khác. Với cùng một logic như vậy, cấu trúc bộ não của con người đã truyền cảm hứng cho việc xây dựng machine learning. Đó là tư tưởng cốt lõi của artificial neural networks (ANNs). Tuy máy bay được truyền cảm hứng từ loài chim nhưng chúng ta không có cánh để bay. Giống như vậy, ANNs dần trở nên khó khăn cho các nhà sinh vật học. Mặc dù nhiều nhà nghiên cứu cho rằng chúng ta nên giảm mức độ tương tự khi so với sinh vật học. E rằng chúng ta đã ngăn sự sáng tạo từ hệ thống sinh học có vẻ hợp lý.

ANNs là cốt lõi của Deep learning. Chúng đa dạng, mạnh mẽ, giúp chúng ta tạo ra những ý tưởng lớn, cực kỳ phức tạp về các ứng dụng Machine learning. Ví dụ như là phân loại hàng tỷ bức ảnh (Google Image), dịch vụ nhân dạng giọng nói đầy mạnh mẽ (Apple’s Siri), giới thiệu những videos tốt nhất cho hàng trăm triệu người dùng (Youtube), …

Từ sinh học đến Artificial Neurons

Tài liệu tham khảo:

1. Sách Hands-on Machine learning with Scikit-Learn & TensorFlow (Aurélien Géron)
2. Sách Machine Learning cơ bản (Vũ Hữu Tiệp)

* Ngày bắt đầu 11/3/2019: tìm kiếm tài liệu, quyết định đề tài, …
* Ngày 14/3/2019: xin file csv, tìm tài liệu
* Ngày 17/3/2019: biết cách import dữ liệu như dạng một array, sẽ chạy thử bằng thư viện sklearn, hiểu về DTNN
* Ngày 18/3: import dữ liệu, ghép thành một array tự động, thử nghiệp thành công bằng knn. sẽ tiếp tục thử với nhiều nồng độ khác nhau bằng sklearn. Sau đó sẽ chuyển sang MLP rồi DNN, cuối cùng DTNN