

資料探勘期末功課

R1231003 陳通, R1231001 高義順, B0831035 楊植翔

A. 背景說明

資料來源於對實驗雞蛋進行非破壞檢測，依照實驗設計，每批次雞蛋會有 16 顆，需對每顆雞蛋的 5 個光源照射(尖端、鈍端、側面*3)面進行照射，並量測出其不同波長(wavelength)下的反射率(reflectance)，實驗共進行一個月，觀察不同影響因子下的反射率變化。

由於每照射結果會有 1937 組反射率對應特定波長段，一天內進行 16 顆雞蛋實驗，共執行一個月的實驗，期間可能會不同雞蛋處理方式，因此資料數量是極其龐大的，若以 excel 處存首先軟體處存空間會不足，另外後續資料調度及分析會變得很困難，sql 的關聯式資料庫就很好的解決了這些問題，關聯式資料庫以不同實驗因子建表，再用 python 聯結 sql 資料庫，對資料庫進行探勘，達成對後續數據分析的應用。

B. 資料類型與處理方法

1. 文字檔前處理+資料庫建立

光譜套裝軟體輸出的資料類型為.sps 檔, 將.sps 檔視為.txt 檔處理，在其輸出的文字檔裡首先排除檔案前 26 行的文字說明，並分別將波長和反射率以串列方式處存。

在資料庫端以不同日期的光譜實驗做資料庫，每個以日期命名的資料庫下有當日實驗的所有雞蛋編號，每個雞蛋表的首欄為波長，後 5 欄為波長所對應的不同照射面的反射率。

2. 資料分析+繪圖

依照要分析的項目做變數替換，選取感興趣的變數因子從資料表中選取，在 python 做數值計算及繪圖

C. 程式使用說明

a. readonefolderzip.py: 程式匯入 DB 資料庫，第 50 行改路徑即可執行

```
49 # 指定資料夾路徑引號內為大資料夾名稱
50 directory_path = r"C:\test\SQL_final\pymysql1130618\pymysql\20240418buying_egg" + '\\' + base_name + ###
```

b. 調用資料函式庫.py:

```
def query_1_2(date, color, egg_serial, direction, wavelength)
```

(date: 雞蛋實驗日期 list, color: 雞蛋顏色, egg_serial: 選擇雞蛋編號,
direction: 雞蛋照射方向 list, wavelength: 選擇波長區間)

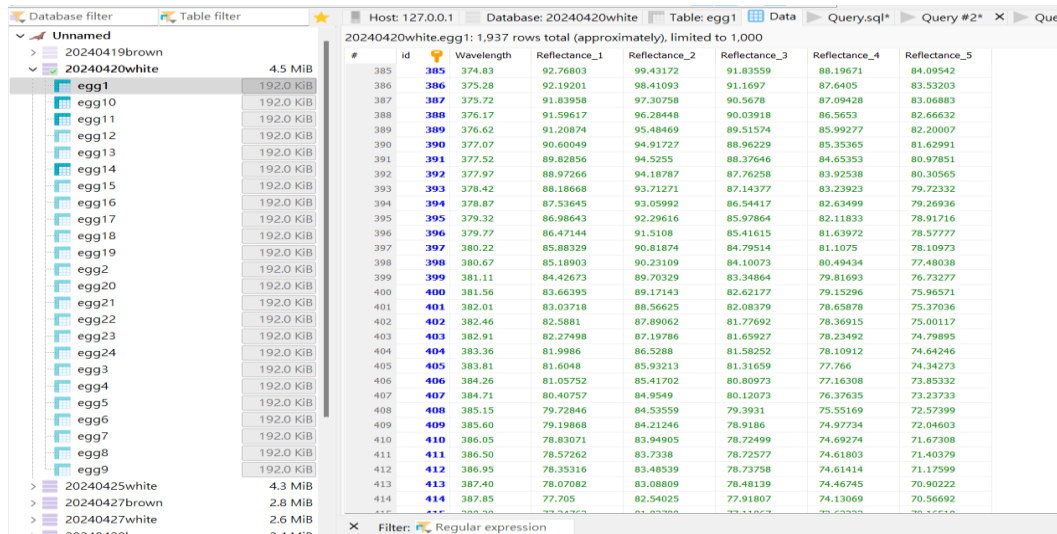
返回值: [retern1, retern2]

[retern1: 波長 list, retern2: 對應反射率強度 list]

c. main.py: 手動輸入查詢引索

```
6 date = ['20240420', '20240425', '20240427']
7 date_color = ['r', 'b', 'g']
8 color = 'white'
9 egg_serial = 'egg1'
10 direction = ['Reflectance_1', 'Reflectance_2', 'Reflectance_3', 'Reflectance_4', 'Reflectance_5']
11 direction_shape = ['', '.', '*', '^', 's']
12 wavelength = ['380.00', '780.00']
```

D. 資料處理結果與說明



Database filter Table filter Host: 127.0.0.1 Database: 20240420white Table: egg1 Data Query.sql* Query #2* X Que

20240420white.egg1: 1,937 rows total (approximately), limited to 1,000

#	id	Wavelength	Reflectance_1	Reflectance_2	Reflectance_3	Reflectance_4	Reflectance_5
385	385	374.83	92.76803	99.43172	91.83559	88.19671	84.09542
386	386	375.28	92.19201	98.41093	91.1697	87.6405	83.53203
387	387	375.72	91.83958	97.30758	90.5678	87.09428	83.08883
388	388	376.17	91.59617	96.28448	90.03918	86.5653	82.66632
389	389	376.62	91.20874	95.48469	89.51574	85.99277	82.20007
390	390	377.07	90.60049	94.91727	88.96229	85.35365	81.62991
391	391	377.52	89.82856	94.53255	88.37646	84.65353	80.97851
392	392	377.97	88.97266	94.18787	87.76258	83.92538	80.30565
393	393	378.42	88.18668	93.71271	87.14377	83.23923	79.72332
394	394	378.87	87.53645	93.05992	86.54417	82.63499	79.26936
395	395	379.32	86.98643	92.29616	85.97864	82.11833	78.91716
396	396	379.77	86.47144	91.5108	85.41615	81.63972	78.57777
397	397	380.22	85.88329	90.81874	84.79514	81.1075	78.10973
398	398	380.67	85.18903	90.23109	84.10073	80.49434	77.48038
399	399	381.11	84.42673	89.70329	83.34864	79.81693	76.73277
400	400	381.56	83.66395	89.17143	82.62177	79.15296	75.96571
401	401	382.01	83.03718	88.56625	82.08379	78.65878	75.37036
402	402	382.46	82.5881	87.89062	81.77692	78.36915	75.00117
403	403	382.91	82.27498	87.19786	81.45927	78.23492	74.79895
404	404	383.36	81.9985	86.5388	81.58252	78.10912	74.64246
405	405	383.81	81.6048	85.93213	81.31659	77.766	74.34273
406	406	384.26	81.05752	85.41702	80.80973	77.16308	73.85332
407	407	384.71	80.40757	84.9549	80.12073	76.37635	73.23733
408	408	385.15	79.72846	84.53559	79.3931	75.55169	72.57399
409	409	385.60	79.19868	84.21246	78.9186	74.97734	72.04603
410	410	386.05	78.83071	83.94905	78.72499	74.69274	71.67308
411	411	386.50	78.57262	83.7338	78.72577	74.61803	71.40379
412	412	386.95	78.35316	83.48539	78.73758	74.61414	71.17599
413	413	387.40	78.07082	83.08809	78.48139	74.46745	70.90222
414	414	387.85	77.705	82.54025	77.91807	74.13069	70.56692

X Filter: Regular expression

圖. 資料庫及資料表



數據修改工具

日期列表（用逗號分隔）: 20240420, 2d

日期顏色列表（用逗號分隔）: r, b, g

顏色變數: white

蛋序列號: egg1

方向列表（用逗號分隔）: Reflectance_1, Reflectance_2, Reflectance_3, Reflectance_4, Re

方向形狀列表（用逗號分隔）: ., *, ^, s

波長列表（用逗號分隔）: 380.00, 780.00

修改數據

顯示當前數據

顯示圖表

使用 tkinter 套件做一個 UI 介面方便操作，有內建預設值可以直接顯示圖表。修改數據後在按下修改數據，可以顯示修改過後的顯示圖表。

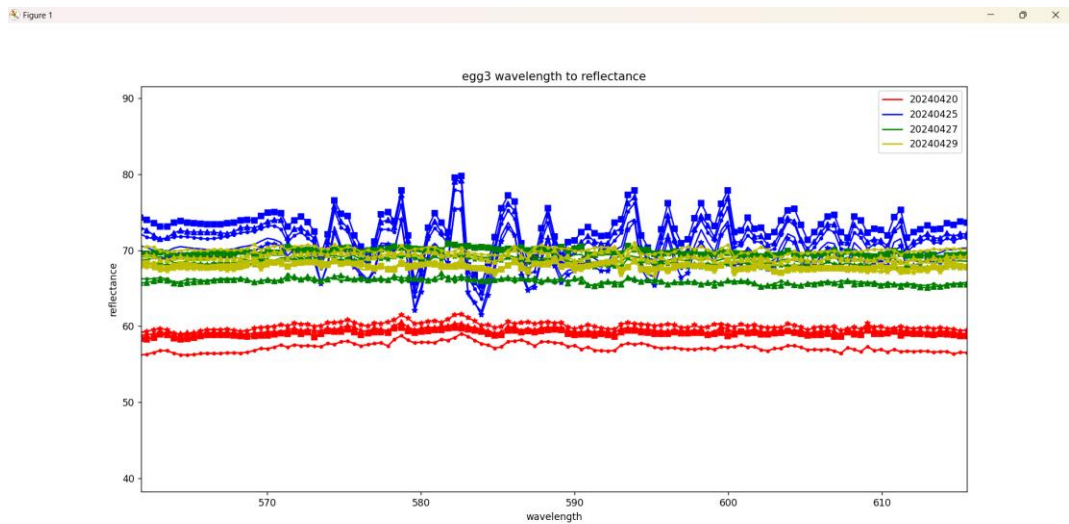


圖. 比較雞蛋在不同日期及不同照射面，感興趣波長段所對應的反射率(將不同天的實驗做顏色區隔;不同反射面做線段圖形做區隔;選取波長560~620(nm))

E. 資料處理後之銜接應用概述

根據想要分析的變因可做成一系列的分析圖表