資料探勘期末功課

R1231003 陳通, R1231001 高義順, B0831035 楊植翔

1. 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。

1. 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。

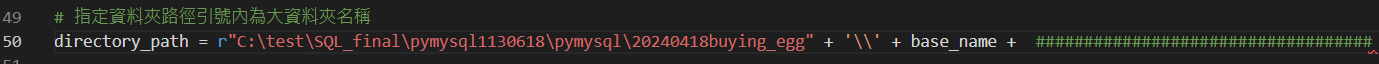
在資料庫端以不同日期的光譜實驗做資料庫，每個以日期命名的資料庫下有當日實驗的所有雞蛋編號，每個雞蛋表的首欄為波長，後5欄為波長所對應的不同照射面的反射率。

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

1. 程式使用說明

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



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]

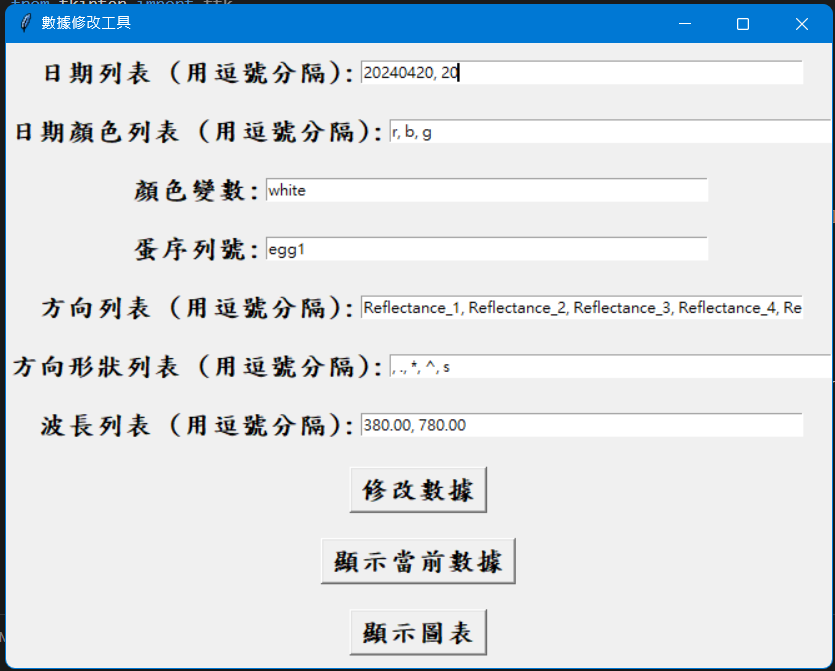
c.main.py:手動輸入查詢引索一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

1. 資料處理結果與說明一張含有 文字, 螢幕擷取畫面, 數字, 軟體 的圖片

   自動產生的描述

圖.資料庫及資料表



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。

一張含有 文字, 螢幕擷取畫面, 行, 繪圖 的圖片

自動產生的描述

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)

1. 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