

利用機器學習預測咖啡品質

邱振銓(S10210043)

吳桂光 老師 指導

*Email: S10210043@thu.edu.tw

摘要

本次研究計畫將會利用機器學習，處理大量由咖啡品質協會(Coffee Quality Institute, CQI)所提供之，三大咖啡產區之綜合咖啡杯測分數。並利用線性回歸方式，預測不同變因對於咖啡杯測分數之影響。

研究計畫內容

(一) 研究動機與研究問題

咖啡是世界上交易量相對較大農產品之一，有超過1.25億的人是依靠此來維持生計。然而，在目前的咖啡市場中，大型主要莊園佔據了咖啡貿易市場的一大部分，也因此中小型咖啡農民，將會難以賺取足夠收入。而咖啡品質協會(Coffee Quality Institute, CQI)，是一家非營利組織，該協會致力於提升咖啡品質，並改善生產國之咖啡農民的生活品質。為此，該協會成立了'Q計畫'。'Q計畫(Q Grader)'是CQI透過可靠且可驗證的系統，精準的辨識出高品質的咖啡，並將其推向市場。然而，評分方式往往對於農民來說，是困難且複雜的。因此我想利用CQI對於咖啡主要生產之三大區域，阿拉比卡、羅布斯塔、賴比瑞亞，針對不同地區、海拔、處理法等等，不同變因對於咖啡杯測分數之影響之數據，利用深度學習之-監督式學習之方式，精準預測該咖啡之杯測分數。並試著尋找出該地區能產出最高分數之咖啡豆，所需要的條件。

(二) 文獻回顧與探討

在咖啡的領域中結合機器學習，已經在市面上具有一定的討論度。如在挑選出咖啡瑕疵豆時，為了節省人力成本以及提升精確度，便有了利用機器學習進行影像辨識，將咖啡豆從外觀上劃分出瑕疵豆種類。例如文獻[1]的作者使用了CNN(Convolutional neural network)作為Training Model、使用ResNet-50作為影像分析之工具、並在CNN之捲基層內，使用AlexNet 作為第一層神經網路層使用。由於此次專題的目的是藉由迴歸方式，找出數據內的咖啡分數與其特徵值之相關性之外，我還希望能夠輸入一組新的數據之後，能夠對未知品項做預測。因此首先，我將考慮使用機器學習中「監督式學習」的方式來進行討論。為了精準地找出最有效率的模型。我參考了有關咖啡品質的論文[2]。在文獻內容中，已經清楚的將各式模型(如SVR(Support Vector Regression)、Random Forest、Decision Tree、KNN(K-Nearest Neighbors)等)，套用了統一的咖啡數值，找出各模型對於MSE、MAE以及RMAE等數值。由於Random Forest 在Tuned之後，MSE與RMAE中的數值呈現最接近0。因此本次專題討論我將會參考作者對於Random Forest模型，如何去調整並優化參數後，來探討有關咖啡品質數據的線性回歸問題。

(三) 研究方法及步驟

對於研究所制定的計劃，將使用監督式學習方式來對咖啡數據進行回歸並預測。監督式學習的方法主要是需要給定輸入數據和相應的輸出標籤，透過模型學習輸入與輸出之間的映射關係。並將整個過程分為兩階段：訓練模型與預測。由於在CQI所提供之數據中還參雜著像是咖啡豆擁有者、證書編號、收成年分等，對於我所需要預測之咖啡分數，顯然不會有太大的影響。因此首要步驟便是透過python語法中，將資料進行初步處理的panda dataframe，來提取我們在這資料中，真正需要的特徵值。下一步驟我將會利用TensorFlow，建構機器學習之模型，並導入keras以簡化建構和訓練神經網路模型之過程。接著未來將考慮使用CNN與KNN模型，針對不同咖啡烘焙度採集其圖像特徵、以及對於咖啡產區、處理法、品種等等.....對咖啡杯測分數做預測。

(四) 預期結果

目前方向最主要為整理數據，接著將會利用這些數據先找出最佳的一組函數，並利用這個函數，使loss值能夠降到最低。最後的目標是希望能夠將咖啡數據中，主要三大變因與咖啡杯測分數之關係找出來，並且給定義組隨機的新資料，輸入至模型中，預測該筆資料之咖啡杯測分數。期望能夠在最後對於咖啡杯測分數會有著8成的準確率。

(五) 參考文獻

- [1] Przybył, K.; Gawrysiak-Witulska, M.; Bielska, P.; Rusinek, R.; Gancarz, M.; Dobrzański, B., Jr.; Siger, A. Application of Machine Learning to Assess the Quality of Food Products—Case Study: Coffee Bean. Appl. Sci. 2023, 13, 10786. <https://doi.org/10.3390/app131910786>
- [2] Devan Anjelito, ML: Coffee Quality Regression, <https://www.kaggle.com/code/devananjelito/ml-coffee-quality-regression/notebook>