

原油庫存量之增減方向預測

- 實驗報告 -

Alexis Lin

Partner: Alan Chen

Date Performed: March 11, 2019

實驗目的

- 取得美國政府提供的能源資料，以每週的原油輸入等各種數值為特徵值，隔週的原油存量為標注資料，以支持向量機等機器學習演算法，訓練一個預估模型。
- 若可以用能源歷史資料預估未來(隔週)的原油存量增減方向，即可藉此預估未來的油價漲跌方向，進而以此作為期貨交易的參考資訊。

實驗方法

- 資料來源 (Datasource)
 - U.S. Energy Information Administration
- 資料集 (Dataset)
 - Weekly Supply Estimates
 - July 31, 2009 ~ March 1, 2019
 - Total is 500 data points.
- 特徵值 (Features)
 - 實驗A：取 Data2-B, Data2-D, ..., 等多個欄位的 t 日資料作為特徵值。
 - 實驗B：取 Data2-B, Data2-D, ..., 等多個欄位，將 $t-1$ 日至 t 日之差額的正負作為特徵值，差額正值為1，負值為0。
- 標注值 (Label)
 - 取 Data6-E 資料，將 t 日至 $t+1$ 日之差額的正負作為標註值，差額正值為1，負值為0。
- 正規化 (Normalization)
 - 實驗A 需做 scaling to range (0,1)

實驗方法 (續)

- 訓練/測試資料集 (Training Set / Test Set)
 - 取最新的500筆資料作為本實驗的資料集
 - 以資料集的前80%的資料集作為訓練集
 - 以資料集的後20%的資料集作為測試集
- 演算法 (Algorithms)
 - 以 Support-Vector Machine(SVM) 演算法進行訓練
 - 以 sklearn 的 GridSearchCV 找到最佳參數
- 評量方式 (Evaluation)
 - 以正確率(Accuracy)做為評量模型效果優劣的方式

實驗方法 (續)

- 資料欄位列表

- D2-B : Weekly U.S. Refiner Net Input of Crude Oil (Thousand Barrels per Day)
- D2-D : Weekly U. S. Operable Crude Oil Distillation Capacity (Thousand Barrels per Calendar Day)
- D6-D : Weekly U.S. Ending Stocks of Crude Oil (Thousand Barrels)
- D6-E : Weekly U.S. Ending Stocks excluding SPR of Crude Oil (Thousand Barrels)
- D6-I : Weekly U.S. Ending Stocks of Total Gasoline (Thousand Barrels)
- D6-J : Weekly U.S. Ending Stocks of Finished Motor Gasoline (Thousand Barrels)
- D6-S : Weekly U.S. Ending Stocks of Gasoline Blending Components
- D6-AB : Weekly U.S. Ending Stocks of Distillate Fuel Oil
- D7-B : Weekly U.S. Days of Supply of Crude Oil excluding SPR (Number of Days)

實驗結果 A

Settings	Best Parameters	Score
D2-B, D2-D	{C: 8, gamma: 2}	0.65
D6-I, D6-J	{C: 2048, gamma: 0.5}	0.58
D6-S, D6-AB	{C: 0.03125, gamma: 3.0517578125e-05}	0.57
D2-B, D2-D, D6-I, D6-J	{C: 2048, gamma: 0.001953125}	0.68
D2-B, D2-D, D6-I, D6-J, D7-B	{C: 128, gamma: 0.03125}	0.67
D2-B, D2-D, D6-I, D6-J, D6-S	{C: 32768, gamma: 0.0001220703125}	0.68
D2-B, D2-D, D6-I, D6-J, D6-AB	{C: 32, gamma: 0.125}	0.66
D2-B, D2-D, D6-I, D6-J, D6-S, D6-AB, D7-B	{C: 8, gamma: 0.125}	0.67
D2-B, D2-D, D6-D, D6-E, ,D6-I, D6-J, D6-S, D6-AB, D7-B	{'C': 0.03125, 'gamma': 3.0517578125e-05}	0.57

實驗結果 B

Settings	Best Parameters	Score
D2-B, D2-D	{C: 0.03125, gamma: 3.0517578125e-05}	0.57
D6-I, D6-J	{C: 0.03125, gamma: 3.0517578125e-05}	0.57
D6-S, D6-AB	{C: 0.125, gamma: 2}	0.61
D2-B, D2-D, D6-I, D6-J	{C: 2, gamma: 8}	0.60
D2-B, D2-D, D6-I, D6-J, D7-B	{C: 8, gamma: 0.03125}	0.62
D2-B, D2-D, D6-I, D6-J, D6-S	{C: 512, gamma: 0.03125}	0.60
D2-B, D2-D, D6-I, D6-J, D6-AB	{C: 0.5, gamma: 0.125}	0.63
D2-B, D2-D, D6-I, D6-J, D6-S, D6-AB, D7-B	{C: 32, gamm: 0.03125}	0.65
D2-B, D2-D, D6-D, D6-E, ,D6-I, D6-J, D6-S, D6-AB, D7-B	{C: 128, gamma: 0.0078125}	0.66

結論

- 實驗A：以原始數值(有正規化)訓練模型，得到最高正確率68%，組合為 D2-B, D2-D, D6-I, D6-J。
- 實驗B：以增減方向(0和1)訓練模型，得到最高正確率66%，組合為 D2-B, D2-D, D6-D, D6-E, ,D6-I, D6-J, D6-S, D6-AB, D7-B。