

原油庫存量之增減方向預測

- 實驗報告 -

Alexis Lin

Partner: Alan Chen

Date Performed: March 8, 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。

Experiment 1

- Datasource :
 - [U.S. Energy Information Administration](#)
- Dataset :
 - Weekly Supply Estimates
 - [Jan 5, 1990 ~ March 1, 2019.](#)
 - Total is 1216 data points.
- Features : (day t-1 -> day t)
 - 2B : 2Weekly U.S. Refiner Net Input of Crude Oil (Thousand Barrels per Day)
 - 2D : Weekly U. S. Operable Crude Oil Distillation Capacity (Thousand Barrels per Calendar Day)
 - 6I : Weekly U.S. Ending Stocks of Total Gasoline (Thousand Barrels)
 - 6J : Weekly U.S. Ending Stocks of Finished Motor Gasoline (Thousand Barrels)
 - 6M : Weekly U.S. Ending Stocks of Reformulated Motor Gasoline, Non-Oxygentated (Thousand Barrels)
 - 7B : Weekly U.S. Days of Supply of Crude Oil excluding SPR (Number of Days)
- Label : (day t -> day t+1)
 - [6C : Weekly U.S. Ending Stocks excluding SPR of Crude Oil and Petroleum Products \(Thousand Barrels\)](#)

Result :

Training a GaussianNB using a training set size of 1216. . .

Accuracy for training set: 0.5502.

Accuracy for test set: 0.4951.

Training a DecisionTreeClassifier using a training set size of 1216. . .

Accuracy for training set: 0.5970.

Accuracy for test set: 0.5049.

Training a SVC using a training set size of 1216. . .

Accuracy for training set: 0.5674.

Accuracy for test set: 0.5574.

Training a AdaBoostClassifier using a training set size of 1216. . .

Accuracy for training set: 0.5650.

Accuracy for test set: 0.5213.

Training a RandomForestClassifier using a training set size of 1216. . .

Accuracy for training set: 0.5970.

Accuracy for test set: 0.4885.

Experiment 2

- Datasource :
 - [U.S. Energy Information Administration](#)
- Dataset :
 - Weekly Supply Estimates
 - [Aug 20,1982 ~ March 1, 2019.](#)
 - Total is 1520 data points.
- Features : (day t-1 -> day t)
 - 2B : 2Weekly U.S. Refiner Net Input of Crude Oil (Thousand Barrels per Day)
 - 2D : Weekly U. S. Operable Crude Oil Distillation Capacity (Thousand Barrels per Calendar Day)
 - 6I : Weekly U.S. Ending Stocks of Total Gasoline (Thousand Barrels)
 - 6J : Weekly U.S. Ending Stocks of Finished Motor Gasoline (Thousand Barrels)
 - 6M : Weekly U.S. Ending Stocks of Reformulated Motor Gasoline, Non-Oxygentated (Thousand Barrels)
 - 7B : Weekly U.S. Days of Supply of Crude Oil excluding SPR (Number of Days)
- Label : (day t -> day t+1)
 - [6D : Weekly U.S. Ending Stocks of Crude Oil \(Thousand Barrels\)](#)

Result :

Training a GaussianNB using a training set size of 1520. . .

Accuracy for training set: 0.5270.

Accuracy for test set: 0.5105.

Training a DecisionTreeClassifier using a training set size of 1520. . .

Accuracy for training set: 0.5632.

Accuracy for test set: 0.5500.

Training a SVC using a training set size of 1520. . .

Accuracy for training set: 0.5388.

Accuracy for test set: 0.5237.

Training a AdaBoostClassifier using a training set size of 1520. . .

Accuracy for training set: 0.5408.

Accuracy for test set: 0.5368.

Training a RandomForestClassifier using a training set size of 1520. . .

Accuracy for training set: 0.5632.

Accuracy for test set: 0.5526.

Experiment 3

- Datasource :
 - [U.S. Energy Information Administration](#)
- Dataset :
 - Weekly Supply Estimates
 - [Jun 4, 2010 ~ March 1, 2019.](#)
 - Total is 365 data points.
- Features : (day t-1 -> day t)
 - 2B : 2Weekly U.S. Refiner Net Input of Crude Oil (Thousand Barrels per Day)
 - 2D : Weekly U. S. Operable Crude Oil Distillation Capacity (Thousand Barrels per Calendar Day)
 - 6I : Weekly U.S. Ending Stocks of Total Gasoline (Thousand Barrels)
 - 6J : Weekly U.S. Ending Stocks of Finished Motor Gasoline (Thousand Barrels)
 - ~~6M : Weekly U.S. Ending Stocks of Reformulated Motor Gasoline, Non-Oxygenated (Thousand Barrels)~~
 - 7B : Weekly U.S. Days of Supply of Crude Oil excluding SPR (Number of Days)
- Label : (day t -> day t+1)
 - [6D : Weekly U.S. Ending Stocks of Crude Oil \(Thousand Barrels\)](#)

Result :

Training a GaussianNB using a training set size of 365. . .

Accuracy for training set: 0.6466.

Accuracy for test set: 0.6304.

Training a DecisionTreeClassifier using a training set size of 365. . .

Accuracy for training set: 0.6767.

Accuracy for test set: 0.6739.

Training a SVC using a training set size of 365. . .

Accuracy for training set: 0.6384.

Accuracy for test set: 0.6957.

Training a AdaBoostClassifier using a training set size of 365. . .

Accuracy for training set: 0.6575.

Accuracy for test set: 0.6848.

Training a RandomForestClassifier using a training set size of 365. . .

Accuracy for training set: 0.6767.

Accuracy for test set: 0.6630.

Experiment 4

- Datasource :
 - [U.S. Energy Information Administration](#)
- Dataset :
 - Weekly Supply Estimates
 - [Jun 4, 2010 ~ March 1, 2019.](#)
 - Total is 365 data points.
- Features : (day t-1 -> day t)
 - 2B : 2Weekly U.S. Refiner Net Input of Crude Oil (Thousand Barrels per Day)
 - 2D : Weekly U. S. Operable Crude Oil Distillation Capacity (Thousand Barrels per Calendar Day)
 - 6I : Weekly U.S. Ending Stocks of Total Gasoline (Thousand Barrels)
 - 6J : Weekly U.S. Ending Stocks of Finished Motor Gasoline (Thousand Barrels)
 - [6M : Weekly U.S. Ending Stocks of Reformulated Motor Gasoline, Non-Oxygentated \(Thousand Barrels\)](#)
 - 7B : Weekly U.S. Days of Supply of Crude Oil excluding SPR (Number of Days)
- Label : (day t -> day t+1)
 - [6D : Weekly U.S. Ending Stocks of Crude Oil \(Thousand Barrels\)](#)

Result :

Training a GaussianNB using a training set size of 365. . .

Accuracy for training set: 0.6384.

Accuracy for test set: 0.6413.

Training a DecisionTreeClassifier using a training set size of 365. . .

Accuracy for training set: 0.7041.

Accuracy for test set: 0.6739.

Training a SVC using a training set size of 365. . .

Accuracy for training set: 0.6384.

Accuracy for test set: 0.6957.

Training a AdaBoostClassifier using a training set size of 365. . .

Accuracy for training set: 0.6575.

Accuracy for test set: 0.6848.

Training a RandomForestClassifier using a training set size of 365. . .

Accuracy for training set: 0.7041.

Accuracy for test set: 0.6739.

Experiment 5

- Datasource :
 - [U.S. Energy Information Administration](#)
- Dataset :
 - Weekly Supply Estimates
 - [Jun 4, 2010 ~ March 1, 2019.](#)
 - Total is 365 data points.
- Features : (day t-1 -> day t)
 - 2B : 2Weekly U.S. Refiner Net Input of Crude Oil (Thousand Barrels per Day)
 - 2D : Weekly U. S. Operable Crude Oil Distillation Capacity (Thousand Barrels per Calendar Day)
 - 6I : Weekly U.S. Ending Stocks of Total Gasoline (Thousand Barrels)
 - 6J : Weekly U.S. Ending Stocks of Finished Motor Gasoline (Thousand Barrels)
 - [6M : Weekly U.S. Ending Stocks of Reformulated Motor Gasoline, Non-Oxygentated \(Thousand Barrels\)](#)
 - 7B : Weekly U.S. Days of Supply of Crude Oil excluding SPR (Number of Days)
- Label : (day t -> day t+1)
 - [6C : Weekly U.S. Ending Stocks of Crude Oil \(Thousand Barrels\)](#)

Result :

Training a GaussianNB using a training set size of 365. . .

Accuracy for training set: 0.5781.

Accuracy for test set: 0.3913.

Training a DecisionTreeClassifier using a training set size of 365. . .

Accuracy for training set: 0.6493.

Accuracy for test set: 0.5217.

Training a SVC using a training set size of 365. . .

Accuracy for training set: 0.5808.

Accuracy for test set: 0.3913.

Training a AdaBoostClassifier using a training set size of 365. . .

Accuracy for training set: 0.5671.

Accuracy for test set: 0.3478.

Training a RandomForestClassifier using a training set size of 365. . .

Accuracy for training set: 0.6493.

Accuracy for test set: 0.4674.

Conclusion

Exp3 和 Exp4 的組合效果最佳：

- 取Jun 4, 2010 ~ March 1, 2019. 區間的資料 (共365筆)
- 以 2B, 2D, 6I, 6J, 6M, 7B(共6個欄位)之 $t-1$ 日與 t 日之差值(取正負)為特徵
- 以 6D (Weekly U.S. Ending Stocks of Crude Oil) 為標註集
- 以 Support-Vector-Machine(SVM)為演算法
- 得到 Accuracy for test set 為 **69.57%**