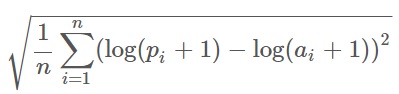
Topic : Kaggle Competition – Bike Sharing Demand

組員：李高丞、許成家、蔣承儒

Final Project Report

Proposal

Bike sharing systems主要用途是藉由網路來達成自行車自助租借、歸還自行車等流程，本次Kaggle競賽所提供的資料為華盛頓2011至2012期間自行車租借情況，資料內容區分包含詳細的年月日、小時、季節、天氣概況、攝氏溫度、體感溫度、濕度、風速、是否為工作日及國定假期、租借人是否為登記用戶或是非登記用戶。

本次競賽目標要求競賽組別必須根據提供的**日期**、**天氣相關參數資料**，預測華盛頓**每一日中每個小時的自行車租借量**(00:00~23:00)，預測準度計算方式如下：

n代表當下小時數字，例如AM 09:00即 n = 9

代表預測租借數量 代表實際租借數量

根據計算公式了解**預測數字與實際數字差越少**，所計算出來的分數會越小，即表示越準確！

Data Fields

提供資料：train.csv & test.csv

資料參數

**datetime** - hourly date + timestamp    
**season** -  1 = spring, 2 = summer, 3 = fall, 4 = winter   
**holiday** - whether the day is considered a holiday  
**workingday** - whether the day is neither a weekend nor holiday  
**weather** - 1: Clear, Few clouds, Partly cloudy, Partly cloudy   
2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist   
3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds   
4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog   
**temp** - temperature in Celsius  
**atemp** - "feels like" temperature in Celsius  
**humidity** - relative humidity  
**windspeed** - wind speed  
**casual** - number of non-registered user rentals initiated  
**registered** - number of registered user rentals initiated  
**count** - number of total rentals

Method

**Algorithm** we use in code：

1. Decision Tree

2. Extra Tree

3. Random Forest

4. Conditional Random Forest (r package)

**Feature Engineering**

**(以下列出所有參數)**

**Date time**

星期1~4 : 01:00~05:00 06:00 (07:00、09:00) 08:00 10:00~11:00 12:00~16:00 17:00~19:00 20:00~22:00 23:00~00:00(隔天)

星期5 : 01:00~05:00 06:00 (07:00、09:00) 08:00 10:00~11:00 12:00~16:00 17:00~19:00 20:00~21:00

22:00~23:00

星期6 : 00:00~02:00 03:00~06:00 07:00 08:00 09:00~11:00 12:00~17:00 18:00~00:00(星期日)

星期7 : 01:00~02:00 03:00~06:00 07:00~08:00 09:00 (10:00、18:00) 11:00~17:00 19:00~21:00 22:00~00:00(星期一)

**提升準確率 0.0262**

**溫度/體感(temp/atemp) Feature**

test$atemp < 30 & test$temp < 11

test$atemp < 30 & 11 <= test$temp < 13

test$atemp < 30 & 13 <= test$temp < 15

test$atemp < 30 & 15 <= test$temp < 20

test$atemp < 30 & 20 <= test$temp < 23

test$atemp < 30 & 23 <= test$temp < 24

test$atemp < 30 & 24 <= test$temp

33 > test$atemp >= 30 & test$temp >= 41

33 > test$atemp >= 30 & 31 <= test$temp < 41

33 > test$atemp >= 30 & 27 <= test$temp < 31

33 > test$atemp >= 30 & 25 <= test$temp < 27

33 > test$atemp >= 30 & test$temp < 25

33 <= test$atemp < 34 & test$temp < 29

33 <= test$atemp < 34 & 29 <= test$temp < 30

33 <= test$atemp < 34 & 30 <= test$temp < 31

33 <= test$atemp < 34 & 31 <= test$temp

34 <= test$atemp < 41 & test$temp < 29

34 <= test$atemp < 41 & 29 <= test$temp < 30

34 <= test$atemp < 41 & 30 <= test$temp < 31

41 <= test$atemp & test$temp < 31

41 <= test$atemp & test$temp >= 31

**提升準確率 0.01328**

**濕度(humidity) Feature**

humidity >= 92

84 <= humidity < 92

82 <= humidity < 84

80 <= humidity < 82

76 <= humidity < 80

74 <= humidity < 76

66 <= humidity < 74

62 <= humidity < 66

58 <= humidity < 62

56 <= humidity < 58

48 <= humidity < 56

**提升準確率 0.0166**

**季節/風速/天氣(season/windspeed/weather) Feature**

season < 1.5 & weather >= 2.5

season < 1.5 & weather < 2.5

season >= 1.5 & winspeed < 10 & weather >= 2.5

season >= 1.5 & winspeed < 6.6 & weather < 2.5

season >= 1.5 & 6.6 <= winspeed < 10 & weather < 2.5

season >= 1.5 & winspeed >= 10 & weather >= 2.5

season >= 1.5 & winspeed >= 10 & 1.5 <= weather < 2.5

season >= 3.5 & 10 < winspeed < 25 & weather < 1.5

season >= 3.5 & 25 <= winspeed & weather < 1.5

1.5 < season < 3.5 & 10 < winspeed < 16 & weather < 1.5

1.5 < season < 3.5 & 16 <= winspeed & weather < 1.5

**提升準確率 0.0121**

Data Postprocessing

做法：

用train.csv data拿掉一部分資料，並執行預測演算法發現：我們所預測的數量在數值低於55的情況下，預測值與實際答案會平均以正負差分佈，而數字越高(大於55)，預測值比實際答案低的分佈居多，因此我們推定如果在大於55後的數字分區間，設定公式去彌補差值會讓準確度提高。

Count分布圖 (藍色表示預測值、紅色代表實際值)：

**公式**:

55 ~ 90 : + 20

91 ~ 165 : + 35

166 ~ 220 : + 40

221 ~ 270 : + 45

271 ~ 330 : + 50

331 ~ 400 : + 60

401 ~ 500 : + 65

501 up : + 70

**提升準確率 0.0154**