[DATA-02P] Bike sharing demand

Miguel-Angel Canela, IESE Business School

December 20, 2015

Introduction

Bike sharing systems are a means of renting bicycles where the process of obtaining membership, rental, and bike return is automated via a network of kiosk locations throughout a city. Using these systems, people are able rent a bike from a one location and return it to a different place on an as-needed basis. Currently, there are over 500 bike-sharing programs around the world, involving more than 500 thousands bicycles. Nowadays, there is a great interest in these systems due to their important role in traffic, environmental and health issues.

Apart from interesting real world applications of bike sharing systems, the characteristics of data being generated by these systems makes them attractive for researchers. Opposed to other transport services such as bus or subway, the duration of travel, departure location, arrival location, and time elapsed is explicitly recorded. This feature turns bike sharing system into a virtual sensor network, which can be used for studying mobility in a city. Hence, it is expected that most of important events in the city could be detected via monitoring these data.

Bike-sharing rental process is highly correlated to the environmental and seasonal settings. For instance, weather conditions, precipitation, day of week, season, hour of the day, etc. can affect the rental behaviors. In a Kaggle competition, participants were asked to combine historical usage patterns with weather data in order to forecast bike rental demand in the Capital Bikeshare program in Washington, D.C.

The data set

The data set is related to the two-year historical log corresponding to years 2011 and 2012, which is publicly available in `http://capitalbikeshare.com/system-data`. The weather information has been extracted from `http://www.freemeteo.com`.

The variables included are:

- Date time (`datetime`): hourly date and timestamp.

- Season (`season`), with spring coded as 1, summer as 2, fall as 3 and winter as 4.

- A dummy indicating whether the day is considered a holiday (`holiday`).

- A dummy indicating whether the day is neither a weekend nor holiday (`workingday`).

- Weather (`weather`), with "clear, few clouds or partly cloudy" coded as 1, "mist and cloudy, mist and broken clouds, mist and few clouds or mist" as 2, "light snow, light rain and thunderstorm and scattered clouds or light rain and scattered clouds" as 3, and "heavy rain and ice pallets and thunderstorm and mist or snow and fog" as 4.

- Temperature in Celsius (`temp`).

- "Feels like" temperature in Celsius (`atemp`).

- Relative humidity (`humidity`).

- Wind speed (`windspeed`).

- Number of non-registered user rentals initiated (`casual`).

- Number of registered user rentals initiated (`registered`).

- Number of total rentals (`count`).

Source: Kaggle competition.