Air Quality and Meteorology Data

Description: The data set is comprised of one-year (2013-2-8 to 2014-2-8) air quality data from air quality monitoring stations in Beijing and Shanghai. There are two folders containing the data of Beijing and Shanghai separately. In each folder, there are two files: “Station.txt” and “CrawledData.txt”, specified as follows.

Station.txt: The detailed information of all air quality monitoring stations. Each row stands for a station, and the columns are defined as follows:

Columns	station_id	station_name	latitude	longitude
Unit	None	None	degree(°)	
Meaning	Each station has a unique id	Chinese name	A station’s position	

CrawledData.txt: Each row denotes the reading from one station at a timestamp. It has 10 columns defined as follows:

Columns	station_id	time	PM25_AQI_value	PM10_AQI_value	NO2_AQI_value	temperature	pressure	humidity	wind	weather
Unit	None	None	None	None	None	℃	hPa	%	km/h	None
Meaning	Corresponds to the station id in Station.txt	data updated time	Current AQI of PM2.5 [1]	Current AQI of PM10 [1]	Current AQI of NO2 [1]	[2]	[2]	[2]	Wind speed [2]	[2][3]

[1] The AQI is calculated based on GB3095-2012 [4], supposing the concentration of particles at this hour’s level lasts for 24 hours.

[2] The data of the district where the station locates.

[3] Each type of weather phenomena corresponds to a unique integer of the variable “weather”:

Snowy = 0, Cloudy = 1, Sunny = 2, Overcast = 3, Rainy = 4, Foggy = 5, Dusty = 6

[4] GB3095-2012: http://www.zzemc.cn/em_aw/Content/GB3095-2012.pdf

[5] Refer to the homepage of Urban Air: http://research.microsoft.com/en-us/projects/urbanair/ for more details.

References: Please cite the following two papers when using this dataset.

1. **Yu Zheng**, Furui Liu, Hsun-Ping Hsieh. U-Air: When Urban Air Quality Inference Meets Big Data. 19th SIGKDD conference on Knowledge Discovery and Data Mining (**KDD 2013**).

2. **Yu Zheng**, Xuxu Chen, Qiwei Jin, Yubiao Chen, Xiangyun Qu, Xin Liu, Eric Chang, Wei-Ying Ma, Yong Rui, Weiwei Sun. A Cloud-Based Knowledge Discovery System for Monitoring Fine-Grained Air Quality. MSR-TR-2014-40.

sContact: Yu Zheng, Email: yuzheng@microsoft.com

Homepage: http://research.microsoft.com/en-us/people/yuzheng/