QQAQ: Air Quality Prediction and Analysis Platform Using Deep Learning

Speaker: Jen-Wei Huang

https://www.youtube.com/watch?v=YDNhHdcCuw4
About Me

- Jen-Wei Huang 黃仁暐
  - Dept. of EE, NCKU
- Research Interest:
  - Data Mining / Machine Learning / AI
- Knowledge and Information Discovery lab
  - https://kid.ee.ncku.edu.tw/
Research Areas

● Data Mining
  □ Social network analysis
  □ Spatial temporal data mining

● Multimedia Information Retrieval
  □ Image search and recommendation
  □ Voice classification (incl. signal processing)
  □ Text mining

● Fintech & Bio-medical Data Analysis
Outline

- Introduction
- Related Works
- Preliminaries
- QQAQ System
- Experimental Results
- Conclusions
Introduction

https://www.epa.gov/sites/production/files/2016-09/pm2.5_scale_graphic-color_2.jpg

Knowledge and Information Discovery Lab, NCKU, Taiwan.
### PM2.5 Index Chart (Abolished in 2017)

<table>
<thead>
<tr>
<th>Index Level</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category</td>
<td>低</td>
<td>低</td>
<td>低</td>
<td>中</td>
<td>中</td>
<td>中</td>
<td>高</td>
<td>高</td>
<td>高</td>
<td>非常高</td>
</tr>
<tr>
<td>PM2.5 Conc. (μg/m³)</td>
<td>0-11</td>
<td>12-23</td>
<td>24-35</td>
<td>36-41</td>
<td>42-47</td>
<td>48-53</td>
<td>54-58</td>
<td>59-64</td>
<td>65-70</td>
<td>&gt;71</td>
</tr>
</tbody>
</table>

#### General Public Activity Recommendations
- Normal outdoor activities.
- Normal outdoor activities.  
- Any discomfort, such as eye pain, cough or throat pain, should consider reducing outdoor activities.  
- Anyone with discomfort, such as eye pain, cough or throat pain, should reduce physical activity, especially outdoor activities.

#### Sensitivity Group Activity Recommendations
- Normal outdoor activities.
- Normal outdoor activities.
- Heart, respiratory and cardiovascular diseases in adults and children should consider reducing physical activity, especially outdoor activities.
- Elderly people should reduce physical activity.
- Patients with asthma may need to increase the use of inhalation agents.

---

Knowledge and Information Discovery Lab, NCKU, Taiwan.
Introduction

Probability map of the daily PM2.5 over $35\mu g/m^3$
Goal

Illustration of Air Quality Forecasting
Goal

Illustration of the Air Quality Predictors
In This Work

- Objectives:
  - Find the most relevant spatial-temporal relations
  - Consider the impacts of elevations to the predictions
  - Provide an data-driven predictive model using spatial-temporal relations adaptively
  - Provide an Air Quality forecasting service
Outline

- Introduction
- Related Works
- Preliminaries
- QQAQ System
- Experimental Results
- Conclusions
Recent Models

- **Temporal Predictor**
  - Consider Local Data

- **Spatial Predictor**
  - Consider Neighbors' Data

- **Predictor Aggregator / Merge Layers**
  - Combine temporal predictor and spatial predictor
Temporal Predictor

\[ \text{PM2.5}_{t_c} - \text{PM2.5}_{t_c-1} - \text{PM2.5}_{t_c-2} + M_{t_c} \]

\[ \text{PM2.5}_{t_c} \rightarrow \text{PM2.5}_{t_c+h} \]
Spatial Predictor
Spatial-Temporal Predictor
Knowledge and Information Discovery Lab, NCKU, Taiwan.
Zheng’s Model

- Temporal Predictor:
  - Linear Regression Model

- Spatial Predictor:

- Predictor Aggregator:

[Forecasting Fine-Grained Air Quality Based on Big Data, SIGKDD2015]

Knowledge and Information Discovery Lab, NCKU, Taiwan.
Outline

- Introduction
- Related Works
- Preliminaries
  - Long-Short Term Memory (LSTM) & Convolutional Neural Network (CNN)
- QQAQ System
- Experimental Results
- Conclusions
Long-Short Term Memory

The repeating module in an LSTM contains four interacting layers.

Knowledge and Information Discovery Lab, NCKU, Taiwan.

http://colah.github.io/posts/2015-08-Understanding-LSTMs/
Convolutional Neural Network

$F(I)$
Outline

- Introduction
- Related Works
- Preliminaries
- QQAQ System
  - Architecture & System Flow
  - Mining Spatial-Temporal Relations
  - Spatial-Temporal Deep Neural Network (ST-DNN)
- Experimental Results
- Conclusions
Architecture

Data sources

Data Crawler

Data & Results

Monitoring

Web Service

Predictive Data

Data Supplement

Model Training

Predictive Model

(online)

(offline)
Offline Process System Flow

Data Collecting and Preprocessing

Air Pollution & Station’s Data → Data Preprocessing → Air Pollution & Station’s Database

Model Evaluation and Comparison

Proposed Model

Select training data interval

kNN-ED

Calculate kNN-DTWD

Top k

k Stations’ data

k Stations’ data

Training data selecting flow
Mining Spatial-Temporal Relations

Adjacent in spatial
Similar in temporal

Knowledge and Information Discovery Lab, NCKU, Taiwan.
Euclidean Distance Similarity

- k-Nearest Neighbor by Euclidean Distance (kNN-ED)
kNN-ED Limitations
kNN-ED Limitations

PM2.5 time series

Knowledge and Information Discovery Lab, NCKU, Taiwan.
Temporal Similarity

- k-Nearest Neighbor by DTW Distance
- Based on Dynamic Time Warping (DTW)
Dynamic Time Warping

- **Problems:**
  - Missing values

- **Solutions:**
  - Shortest time interval threshold
  - Average unit distance
kNN-DTWD Flows

Set Minimum Time Interval $l_{\text{min}}$ → Determine missing values and common interval $>l_{\text{min}}$

Calculate DTW → Calculate average unit distance

Knowledge and Information Discovery Lab, NCKU, Taiwan.
Shortest Interval Threshold $l_{\text{min}}$

- The ignored interval may induce some loss of information, but can effectively remove most of the errors caused by the noise.
The average unit distance can combine multiple fragment sequences in order to facilitate the identification of the final similarity.

Eg. \[ d_\Delta = \frac{d_1 + d_2}{l_1 + l_2} \]
Spatial-Temporal Deep Neural Network Model Structure

- PM2.5 and other meteorological data (2k+1 stations)
- Terrain
- CNN
- LSTM
- Merge
- ANN
- Neural Network Model
- Target

Knowledge and Information Discovery Lab, NCKU, Taiwan.
Data Flow

Neighbor Data
- kNN-ED
  - Recent Meteorology
  - Recent PM2.5

Local Data
- kNN-DTWD
  - Recent Meteorology
  - Recent PM2.5
- Terrain
- ΔPM2.5

Spatial-Temporal Relations
- Terrain Extractor
- Temporal Relations
- Spatial-Temporal Predictor
- Predictions
Dealing with Terrain

- 11x11 points with spacing = 500m
- Inverse Distance Weighting (IDW)
  
  \[ IDW \text{ local mean}(n = 2) = \frac{\sum_{i=1}^{k} \frac{value_i}{(dist_i)^2}}{\sum_{i=1}^{k} \frac{1}{(dist_i)^2}} \]

- Hadamard product with relative elevations
Relative elevation function

\[ H_S = \frac{\text{elev} - \text{elev}_{st}}{\text{elev}_{st}}, \quad \text{elev}_{rel} = \frac{1}{e^{H_S}} \]

[Vertical profiles of aerosol absorption coefficient from micro-Aethalometer data and Mie calculation over Milan, *Science of the Total Environment 2011*]
Outline

- Introduction
- Related Works
- Preliminaries
- QQAQ System
- Experimental Results
- Conclusions
Experiments

- Taiwan Dataset
  - Provided by Taiwan Environment Protection Administration

- Beijing Dataset
  - Provided by “Forecasting Fine-Grained Air Quality Based on Big Data”, SIGKDD2015
Metrics

● Mean Absolute Error (MAE):

$$e = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n}$$

- $\hat{y}_i$: predictions
- $y_i$: ground truth
- $n$: number of instances measured for a time interval
Comparison Models

● ANN
  - kNN-DTWD
  - kNN-ED
  - Adaptive
  - All

● Linear Regression (LR)
  - kNN-DTWD
  - kNN-ED
  - Adaptive
  - All

● Zheng [KDD’15]
Proposed Model

- A (ANN only)
- L (LSTM only)
- C (CNN only)
- A+L
- A+C
- L+C
- A+L+C

Knowledge and Information Discovery Lab, NCKU, Taiwan.
Taiwan Dataset

- Training data: Jan 2014 ~ Sept 2016
- Testing data: Oct 2016 ~ Sept 2017
- Discussions group by:
  - Air Quality Regions
  - Terrains
Comparisons between Components

Short Period (1-6hrs) Prediction MAE

MAE ($\mu g/m^3$)

A L C A+L A+C L+C A+L+C

Time (hr)

1 2 3 4 5 6
Comparisons between Models

Short Period (1-6hrs) Prediction MAE

MAE ($\mu g/m^3$)

<table>
<thead>
<tr>
<th>Time (hr)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zeng</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptive_ANN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kNN-ED_ANN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kNN-ED_LR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALL_ANN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALL_LR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Knowledge and Information Discovery Lab, NCKU, Taiwan.
Summary

• In the 1st hour prediction, A+L+C almost outperforms all models

• For the 2nd-6th hour predictions, C almost outperforms all models
Beijing Dataset

- **Training data:**
  
  5-6, 8-9, 11-12/2014;
  
  2-3 /2015

- **Testing data:**

  7, 10/2014; 1, 4/2015

Knowledge and Information Discovery Lab, NCKU, Taiwan.
Features

- PM2.5
- PM10
- Pressure
- Weather
- Wind_speed
- Wind_direction
- Temperature
- Humidity
- No elevations
Prediction Results

Short Period (1-6hrs) Prediction MAE

MAE ($\mu g/m^3$)

Time (hr)

1 2 3 4 5 6

Zheng A+L Adaptive_LSTM kNN-DTWD_ANN Adaptive_ANN kNN-ED_ANN Adaptive_LR kNN-DTWD_LR Adaptive_LR kNN-ED_LR

Knowledge and Information Discovery Lab, NCKU, Taiwan.
Prediction Results

4 Time Intervals Prediction MAE

MAE ($\mu g/m^3$)

<table>
<thead>
<tr>
<th>Time (hr)</th>
<th>1~6</th>
<th>7~12</th>
<th>13~24</th>
<th>25~48</th>
<th>Time (hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zheng</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A+L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kNN-DTWD_ANN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptive_ANN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kNN-DTWD_LR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptive_LR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kNN-ED_ANN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kNN-ED_LR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Knowledge and Information Discovery Lab, NCKU, Taiwan.
DEMO

- QQ Air Quality - QQAQ
  - [https://qqaq.ee.ncku.edu.tw/](https://qqaq.ee.ncku.edu.tw/)
Web Pages

- https://qqaq.ee.ncku.edu.tw/dashboard/
- http://140.116.164.187/test/Visualize/TaiwanGrid.html
Conclusions

- This study proposed a spatial-temporal air quality predictive model based on deep neural networks to monitor and predict air quality.

- The proposed relation extractor kNN-DTWD mostly outperforms kNN-ED.

- Spatial distance in the air pollution is a certain impact, but the temporal relations can help to model changes.
Conclusions

- PM2.5 is indeed cyclical
- Southern Taiwan is more serious than northern Taiwan. Eastern Taiwan has the best air quality all around the year.
Acknowledgements

- Special thanks to:
  - Ping-Wei Soh
  - Meng-Xun Zhong
  - Jia-Wei Chang
  - Hsiao-Ting Huang
Thank you for your attention.

https://kid.ee.ncku.edu.tw