An Interactive Visual Comparison System for Air Quality Data

Rongwen Zhao, rzhao17@ucsc.edu

Abstract—In recent years, due to the increasing development of industrialization and urbanization, air quality has become an important topic of discussion by the public in many countries. There are many studies have been conducted to reveal different aspects and discover the normal patterns of air quality. However, since the air quality data has many different types with geo-based data, existing methods cannot analyze such data comprehensively to the public. In this case, we introduce a visual analysis system to assist people to better analyze and explore air quality data.

Index Terms—air quality; geo-based data; data analysis; visual analysis;

1. Introduction

Nowadays, with the rapid development of economy, air pollution is becoming a hot issue in many developing countries. Frequent occurrences of hazy weather have attracted great attention on air quality among the public and the researchers. Therefore, various research studies in different areas have been conducted and tried to solve this problem through different ways. However, many methods are not quite intuitive and the public feel a little bit difficult to comprehend what these figures mean. Actually, many researchers tend to show their research findings using some special ways, either complicated mathematical equations or abstract figures, which are not appropriate for government managers to understand. In this way, they are not likely to adopt this approach, though it’s very useful to control the air quality and reduce the air pollution.

Figure 1. Picture showing air pollution

With the development of wireless sensors and remote sensing, air quality data can be collected by monitoring stations and remote sensing satellites. Usually, the collected data are spatial and temporal, which means each entity contains the time and location where it was recorded. Normally, air quality data consist of many air pollutant concentration values, such as SO, NO, SO2 and NO2. Each of them are recorded by monitoring stations hourly. Visual exploration of spatial-temporal air quality data can be more challenging and tough.

Figure 2. Air quality monitoring station

2. Related Work

In this section, we discuss the visual analysis for air quality data from two aspects. On the one hand, we need to summarize the existing techniques of data mining and machine learning to analyze air quality data. On the other hand, we should explore the visualization design for air quality data.

2.1 Algorithmic analysis for social media data

Normally, in order to do deal with analysis problems of air quality data, we have four different procedures, namely data discovery, data collection, data clean and data analysis[4,6,8]. According to distinct purposes, a lot of various applications of air quality data analysis methods have been classified into different types. Such as prediction the relationship of different areas through spatial links, patterns analysis of different areas.

There are a lot of famous algorithms applied into the field of air quality data[2,3,4,6]. Latent Dirichlet Allocation is one of the most popular methods. It is widely used to detect and model new topics in spatial-temporal data. Topic detection and modeling is a kind of statistical modeling task to disclose some latent or abstract topics which appears in a collection of temporal data. Usually, it builds a topic per document model and words per topic model, modeled as Dirichlet distributions. However, due to the huge volume of spatial-temporal data, LDA is not appropriate to analyze such data with different types and attributes.

2.2 Visualization for air quality data

Since the existence of machine learning and data mining techniques, visual analytics can be a beneficial helper for people to analyze social data. Visual analytics combines data visualization, data mining techniques and human-computer interaction skill to analyze spatial-temporal data thoroughly. Users can use some interactive interfaces to interact with system, and they can adjust the output of algorithm. Due to the large volume and complexity of air quality data, many researchers have developed a lot of various visualization systems to help people to analyze spatial-temporal data.

A lot of visualization systems or software have been developed to detect and explore anomalies and events for air quality data[4,5,7]. For example, Guo et al. proposes a visual analysis system, Vis-Stamp, for users to generate various visualization figures. It contains heat maps, parallel coordinate maps and map matrices. Moreover, it can be used to explain the divergence and evolution of spatial-temporal data with multi-variable patterns.

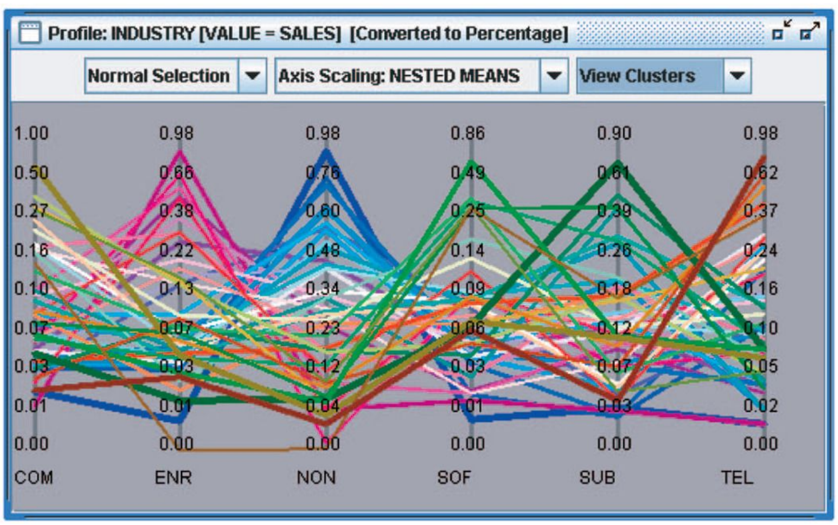

Figure 3. Overview of Vis-Stamp system

Lu et al. develops an interactive visual analysis tools for visualizing multi-granularity time-series data.

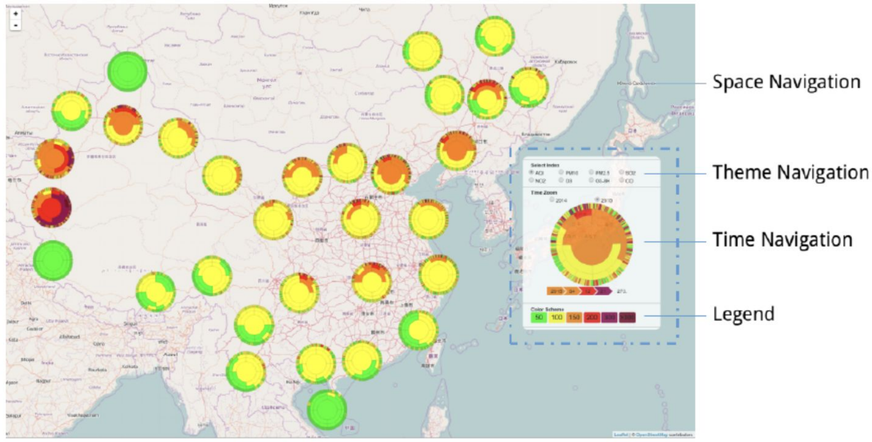

Figure 4. Interface of Lu's system

3. Data Description

This section aims to talk about how to collect data described blow from the website. As for our research project, we used the datasets from the website of National Environmental Protection Agency. These datasets were recorded hourly by many monitoring stations. And it contains about 1,500 monitoring stations. Each station record six common pollutants, including nitrogen dioxide (NO2), sulfur dioxide (SO2), ozone (O3), carbon monoxide (CO), and particulate matter (PM2.5, PM10). As shown in Figure 5.

Station list:

Station ID	City or County Name	Station Name	Local name (optional)	Latitude/Longitude	Timezone (optional)
ID_BEI_DC	Beijing	Dongcheng	东城东四	39.929/116.417	+0800
ID_BEI_WP	Beijing	West Park	西城官园	39.929/116.339	+0800
ID_BEI_OP	Beijing	Olympic Park	朝阳奥体中心	39.982/116.397	+0800
...	...	...	...	...	...

Real-time pollutant list:

Station ID	Pollutant	Unit	Update time	Value	Averaging
ID_BEI_DC	PM10	mg/m3	2019/03/19 00:00:00	27.8	1 hour
ID_BEI_DC	PM25	mg/m3	2019/03/19 00:00:00	10.8	1 hour
ID_BEI_DC	Ozone	mg/m3	2019/03/19 00:00:00	15.2	1 hour
ID_BEI_DC	Ozone	mg/m3	2019/03/19 00:00:00	18.2	8 hours
ID_BEI_DC	Temperature	Celcius	2019/03/19 00:00:00	22.3	1 hour
ID_BEI_WP	PM10	mg/m3	2019/03/19 00:00:00	27.8	1 hour
ID_BEI_WP	PM25	mg/m3	2019/03/19 00:00:00	10.8	1 hour
ID_BEI_WP	SO2	ppb	2019/03/19 00:00:00	15.2	1 hour
ID_BEI_WP	Humidiy	%	2019/03/19 00:00:00	88	1 hour
...	...	...	...	...	

Figure 5. Description of dataset collected

4. System Overview

In this section, we introduce how to build the whole system after collecting the dataset from website. Figure 6 shows the whole system overview.

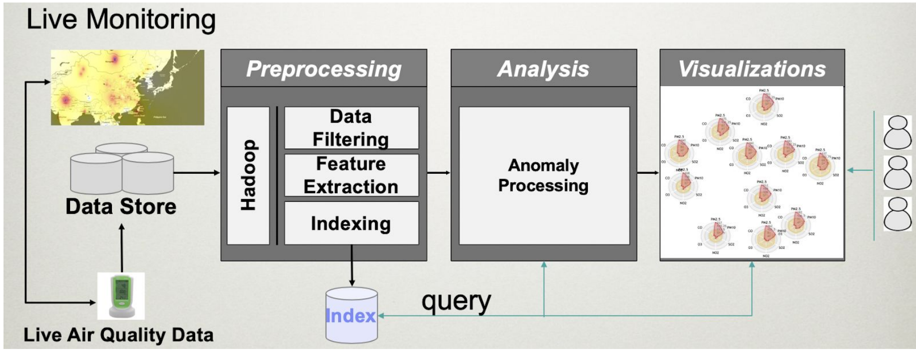

Figure 6. System overview

First, we get the real-time data from the target website and store it in database. After that, we need to conduct some preprocessing works. For example, some stations may lose some feature values at some points, in this case, we need to delete these stations. The preprocessing has three steps: (1) Data filtering: filter some special stations. (2) Feature extraction: extract the features you want to explore in the original datasets. (3) Indexing.

5. System Design

In this section, we introduce our interactive visual comparison framework in details. The whole system contains two parts: map view and radar view. Figure displays the design of the system.

5.1 Design of map view

We divided the whole map into many small region based on the same spatial granularity, as shown in Figure 7. Each region encodes by a different color. We used identical initialization color mapping by computing from the features distribution. In additions, about 1,500 monitoring station are divided into these various regions. Each region contains several monitoring stations.

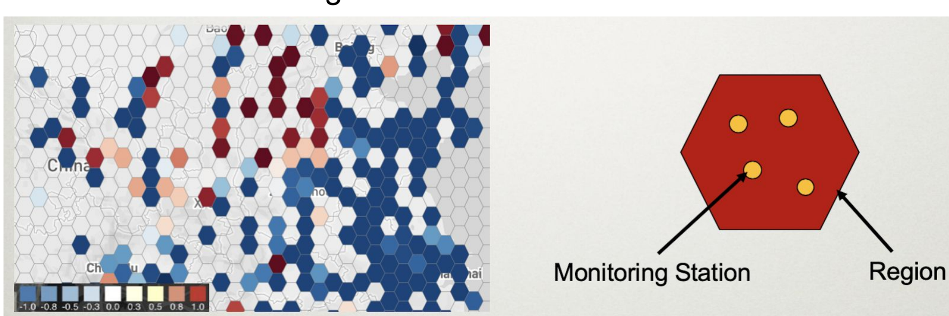

Figure 7. Design of map view

5.2 Design of radar view

In the radar view, we visualized each monitoring station as a six-dimensional radar. Each dimension represent the value of one type of pollutants. And we plot the feature values of each station along the feature axes surrounding the baseline. As shown in the Figure 8.

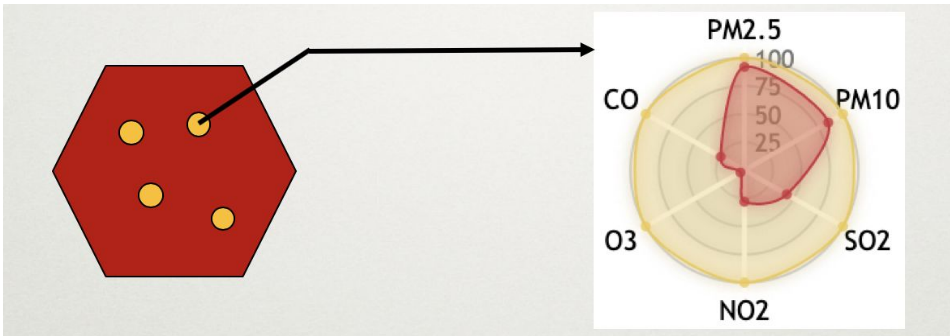

Figure 8. Design of radar view

6. Visualization Results

In this section, we show some initial results generated by our system.

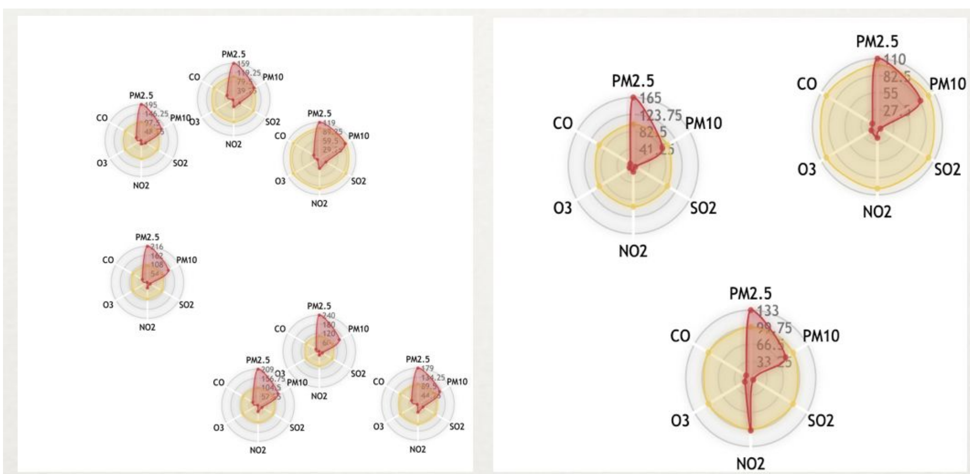

Figure 9. Visualization of regions with less stations

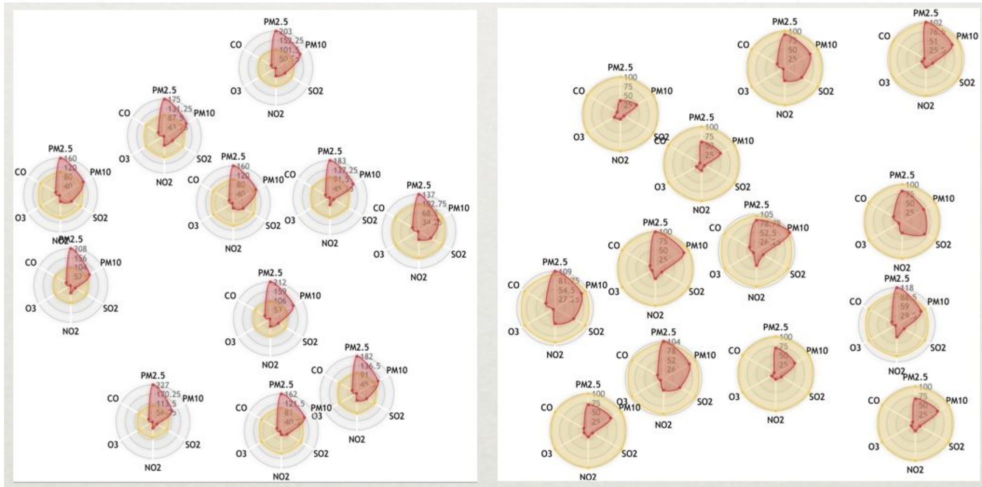

Figure 10. Visualization of regions with more stations

7. Conclusions and Future Works

In this project, we present an interactive visual comparison system for air quality data. In a brief, it proposes an online visual comparison framework for multi-variable data. Second, we adopt a position-aware graph layout algorithm. What’s more, it applies a novel method for interactively and progressively refining. For the future works, I will try to design intuitive visualization. In addition, we can choose better layout methods.

References

[1]] Li, Huan, Hong Fan, and Feiyue Mao. "A Visualization Approach to Air Pollution Data Exploration—A Case Study of Air Quality Index (PM2. 5) in Beijing, China." *Atmosphere* 7.3 (2016): 35.
[2] Guo, Diansheng, et al. "A visualization system for space-time and multivariate patterns (vis-stamp)." *IEEE transactions on visualization and computer graphics* 12.6 (2006): 1461-1474.
[3] Pebesma, Edzer J., Kor de Jong, and David Briggs. "Interactive visualization of uncertain spatial and spatio-temporal data under different scenarios: an air quality example." *International Journal of Geographical Information Science* 21.5 (2007): 515-527.
[4] Elbir, Tolga. "A GIS based decision support system for estimation, visualization and analysis of air pollution for large Turkish cities." *Atmospheric Environment* 38.27 (2004): 4509-4517.
[5] Wood, Jason, Ken Brodlie, and Helen Wright. "Visualization over the world wide web and its application to environmental data." *Proceedings of Seventh Annual IEEE Visualization'96*. IEEE, 1996.
[6] Junninen, Heikki, et al. "Smart-SMEAR: on-line data exploration and visualization tool for SMEAR stations." (2009).
[7] Novak, JOAN H., and ROBIN L. Dennis. *Regional air-quality and acid-deposition modeling and the role for visualization*. No. PB--92-124247/XAB. National Oceanic and Atmospheric Administration, 1991.
[8] Rebolj, Danijel, and Peter J. Sturm. "A GIS based component-oriented integrated system for estimation, visualization and analysis of road traffic air pollution." *Environmental Modelling & Software* 14.6 (1999): 531-539.

[9] Denzer, Ralf, and Gerald Schimak. "Visualization of an Air Quality Measurement Network." *Informatik für den Umweltschutz/Computer Science for Environmental Protection*. Springer, Berlin, Heidelberg, 1991. 78-86.
[10] Denzer, Ralf, and Gerald Schimak. "Visualization of an Air Quality Measurement Network." *Informatik für den Umweltschutz/Computer Science for Environmental Protection*. Springer, Berlin, Heidelberg, 1991. 78-86.