

Leveraging Egocentric and Surrounding Environment Data to Adaptively Measure a Personal Air Quality Index

Dang-Hieu Nguyen¹, Minh-Tam Nguyen², Loc Tai Tan Nguyen³,

^{1,2,3}University of Information Technology, VietNam

hieund.12@grad.uit.edu.vn, tamnm.12@grad.uit.edu.vn, locntt.12@grad.uit.edu.vn

ABSTRACT

This paper introduces a new solution for measuring the personal air quality index that reflects the egocentric perspective of human beings with their surrounding environment. Two instances of the solution are introduced and evaluated by using the MediaEval 2019 Insights for wellbeing task dataset and evaluation metric. The first instance calculates the Air Quality Index (AQI) using sensors data, utilizes the user's tags and visual features to measure the personal AQI adaptively. The second instance leverages the average value of the user's tags and feature of the route to determine personal AQI. The performance of these two instances is also discussed.

1 INTRODUCTION

In [2], the author gives various evidence gathered from many reference sources and points out the impact of air pollution on individuals in many perspectives (health, psychology). The mentioned pollution factors include environmental factors (e.g. fine particulate matter $PM_{2.5}$, Nitrogen dioxide NO_2 , Ozone O_3 , Sulfur dioxide SO_2), weather variables (e.g. temperature, humidity), and urban nature, traffic. Unfortunately, most of investigations on this domain focusing on measuring the air quality index using sensors data regardless of understanding how people feel of air qualification around them.

MediaEval 2019 Insights for wellbeing task [1] introduces an interesting subtask of measuring personal air quality index (PAQI). The PAQI is defined as the personal feeling of AQI comparing to the real AQI calculated by using sensors data. The subtask requests to measure the PAQI using egocentric data (e.g., lifelog image, heartbeat, step counts, user's annotations) and surrounding environment data (e.g., air pollution, weather).

The definitions, dataset and evaluation metric of this subtask are described in [1].

2 METHODOLOGY

As mentioned above, environmental factors, weather variables, urban nature, and traffic impact on individuals. Observing the dataset provided by the subtask, we found that main streets with lots of traffic and fewer trees will have a low PAQI and vice versa. This observation gives a hint to propose the solution to measure PAQI using AQI, user's tags, and visual features. Two instances of the solution are introduced and evaluated by using the MediaEval 2019 Insights for wellbeing task dataset and evaluation metric. The first instance calculates the Air Quality Index (AQI) using sensors data, utilizes the user's tags and visual features to measure the personal AQI adaptively. The second instance leverages the average value of the user's tags and feature of the route to determine personal AQI.

2.1 Data Processing

First, data along each route are pre-processed to get rid of noises and outliers. Necessary interpolations are conducted to compensate for missing data. Then, two instances (runs) of the proposed solution are constructed as follows:

Run 1: From the dataset, we can identify a group of users walking along a specific route. Since the 2018 dataset is recorded by seconds, we convert a recording time to the minute to make sure the highest value of each factor within 1 minute is retained. Then we calculate AQI using these factors (e.g., $PM_{2.5}$, NO_2 , O_3). Next, visual features are extracted from images.

Run 2: We first collect all data in the same group, then we only keep data coordinating with user's tags. Next, we divide each segment of one route into four smaller segments. This task aims to have a segment as straight as possible so that the radius can sweep all the points tagged on the segment. Then, we scan the radius with a radius by the distance between the small segments, if any of the tag points are within this range, we collect them and calculate the average value of that user's tags (e.g., assume the distance between *line_start* and *line_end* is 100 m. We divide it into four road segment with 25 m each and get new 3 points in between).

2.2 Visual Features Extraction

We use the visual features provided by the task’s organizers. Besides, we develop a tool that crawls images from Google Street View using coordinates provided in the dataset. This tool aims to enrich the image dataset. Finally, we extracted traffic and tree density from these images.

2.3 PAQI Measurement

In this section, we use the data input obtains from section 2.1 corresponding to each run mentioned.

2.3.1 Run 1: PAQI measurement is first calculated by using the AQI calculation formula. Then we use the user’s tags and traffic density, tree density to adjust the AQI values. We build a function to adaptively adjust AQI into PAQI as follows:

$$f(x) = \frac{\sum_{i=1}^n (factor_i \cdot \alpha_i)}{n} \quad (1)$$

Where: $\sum_{i=1}^n (\alpha_i) = 1$; $factor_i$: input data such as user’s tags and visual feature.

The PAQI’s value are specified by:

$$PAQI = AQI \cdot f(x) \quad (2)$$

Finally, we adjust the value of α_i according to a route to get the final PAQI. The value of α_i is calculated based on the factors’ values. If the factors’ values are high, α_i increases and the PAQI is high. If factors’ values are low, α_i decreases and the PAQI is low.

We set parameters as follows: $factor_1 \leftarrow$ user’s tags, $factor_2 \leftarrow$ traffic density, $factor_3 \leftarrow$ tree density, $\alpha_1 + \alpha_2 + \alpha_3 = 1$. First, we define $\alpha_1 = \alpha_2 = \alpha_3 = \frac{1}{3}$. Then we use ad-hoc-based approach to calculate factors’ values and adjust the values of α corresponding with each factor. With $factor_1$, if its value is larger than the predefined threshold (2.5 in our case), α_1 increases, otherwise α_1 decreases. With $factor_2$, if its value is high, α_2 decreases, otherwise, α_2 increases. With $factor_3$, if its value is high, α_3 increases, otherwise, α_3 decreases. This optimal loop is carried on until the convergence happens. With the maximum value of α is 1 and the minimum value of α is 0.

2.3.2 Run 2: First, we based on *line_start* and *line_end* points to determine the features of routes 1, 2, and 3 that are featured in Table 1. Second, we calculate the average value of the user’s tags. Third, we calculate the weight of routes:

$$w_r = \frac{PAQI_{input}}{avg(user's\ tags)} \quad (3)$$

Where: w_r : is weight of route.; $PAQI_{input}$: based on Development Dataset of routes 1, 2, and 3.; $avg(user's\ tags)$: the average user’s tags on routes 1, 2, and 3.

Because we can not find the “Mountain trail” feature in Development Dataset, we assume the weight of “Mountain

trail” feature same value with “Bayside path” feature. Table 1 shows the weight of routes 1, 2, and 3 when running on the development dataset.

Table 1. The weight of features

Feature	Weight	Feature	Weight
Main street	0.33	Shopping street	0.5
Path	0.67	Underground arcade	0.25
Sightseeing	1.5	Garden	1.5
Street	0.5	Bayside path	2
Park	1.5	Mountain trail	2

We use the weight of routes in Table 1 to infer PAQI’s value of routes 4 and 5 by the formula below:

$$PAQI_{output} = w_r \cdot avg(user's\ tags) \quad (4)$$

Where: $PAQI_{output}$: is value predicted of routes 4 and 5.; w_r : is weight from Table 1.; $avg(user's\ tags)$: the average user’s tags on routes 4 and 5.

3 RESULTS AND ANALYSIS

The experimental results running on the training dataset are denoted in Table 2. The results show we can measure PAQI with acceptable accuracy. Table 3 shows the results when running on the testing dataset. In Table 2 does not include the result of run 2 because we only obtain the weight of routes from Development data and use it to infer PAQI for Testing data.

Table 2. Results running on the training dataset

ROUTE/ List of course groudtruth	List of course run 1
Course 1/ (1, 2, 3, 4, 1)	1,2,2,3,2
Course 2/(1,2,3,4,1)	1,2,2,4,2
Course 3/(2,1,2,0,3,3,2)	2,1,2,2,4,2,2

Table 3. Results running on the testing dataset

Group_id	Subtask_id	Run_id	Score
SHT_UIT	2	1	0.8
SHT_UIT	2	2	1

In Table 3, we can see that run 1 has better performance than run 2. It shows that the approach of measuring PAQI by adaptively adjusting by user’s tags, density traffic, and density tree is more efficient than the approach that uses the average user’s tags and feature of the route. In run 2, since we use the rounding of numbers when calculating the average user’s tags, it could affect the whole performance.

4 CONCLUSION

In this paper, we report our solution for the challenge raised by MediaEval 2019 Insight for wellbeing task - subtask2. We introduce an ad-hoc approach that adaptively adjust user’s tags, traffic and tree density observed along a route to re-adjust the AQI value towards measuring an acceptable personal AQI value.

REFERENCES

- [1] Minh-Son Dao, Peijiang Zhao, Tomohiro Sato, Koji Zettsu, Duc-Tien Dang-Nguyen, Cathal Gurrin, and Ngoc-Thanh Nguyen. 2019. Overview of MediaEval 2019: Insights for Wellbeing Task: Multimodal Personal Health Lifelog Data Analysis. In *MediaEval2019 Working Notes (CEUR Workshop Proceedings)*. CEUR-WS.org <<http://ceur-ws.org>>, Sophia Antipolis, France.
- [2] Siqi Zheng, Jianghao Wang, Cong Sun, Xiaonan Zhang, and Matthew E Kahn. 2019. Air pollution lowers Chinese urbanites' expressed happiness on social media. *Nature Human Behaviour* 3, 3 (2019), 237.