Healthism@MediaEval 2019 - Insights for Wellbeing Task: Factors related to Subjective and Objective Health

Qi Huang[1, 2], Ailin Sheng[1, 2] Lei Pang[1], Xiaoyong Wei[1, 2], Ramesh Jain[1, 3]

[1]Peng Cheng Lab, Shenzhen, China
[2]Sichuan University, Chengdu, China
[3]University of California Irvine, Irvine, US

huangqi@stu.scu.edu.cn,2017223040013@stu.scu.edu.cn,pangl01@pcl.ac.cn,weixy@pcl.ac.cn,jain@ics.uci.edu

ABSTRACT

This paper presents an overview of our proposed methods for the two subtasks of MediaEval 2019 "Insights for Wellbeing". The goal is to investigate the factors related to personal environmental health conditions (PEH). We model this as a regression problem where environmental factors measured by both physical sensors and psychological perception ratings are used as indicators for predicting the target values of objective or subjective PEH measures. A variety of models (e.g., GBDT and LSTM) have been employed to conduct the regressions. The experimental results indicate that objective PEH is mainly determined by the factors indicated by physical sensors and the temperature and humidity contribute the most. Subjective PEH is dominated by the perceptual factors collected from questionnaires and indicated by urban natures mined from a GIS dataset.

1 INTRODUCTION

Plenty of studies have been carried out for investigating the relationship between the environmental conditions and the human wellbeing [5, 7]. While fruitful and usefully findings have been obtained, however, these demographic conclusions provide limited guidance when we are applying them to individual cases (e.g., to study the diseases that are tightly related to personal exposure) [2]. Therefore, there is an urgent need to measure the personal environmental health conditions (PEH). Thanks to the recent progress on wearable devices, we are able to measure PEH through a wide range of sensors such as those for general conditions (e.g., temperature, humidity) and those for specifical air pollution (e.g., PM2.5, NO_2, O_3). This makes datasets like SEPHLA [6] (the first PEH dataset) possible.

With SEPHLA, this paper presents our methods proposed for the two subtasks of MediaEval 2019 "Insights for Wellbeing". The goal is to investigate the main factors related to personal environmental health conditions (PEH). In this study, PEH has been measured objectively by the PM 2.5 and subjectively by P-AQI (personal air quality index which is determined by a human rating system). We model this as a regression problem in which PEH measures are the targets while factors are the indicators. We have investigated a wide of factors including those measured by physical sensors (e.g., temperature, humidity, heart rate), those collected from questionnaires for measuring people's psychological perceptions to the PEH,

and urban natures extracted from OpenStreetMap[1] (a third-party GIS dataset). A variety of models (e.g., GBDT and LSTM) have been employed to fulfill the regressions.

2 FACTORS OVERVIEW

2.1 Physical Sensors

Since PEH focuses on a road-level area, the factors from city-level sensors are discarded, including NO_2, O_3 and NO. The remaining factors are collected from wearable devices, such as PM2.5, temperature, humidity as well as heart rate. For PM2.5, a calibration strategy, which replace those values shifting from mean with more than three times of variation with mean of its surrounding 30 values, is adopted to filter outliers. In total, there are 43,684 and 24,055 samples in training and testing set respectively.

2.2 Questionnaires

In SEPHLA, the participants should answer 5 questions at each checkpoint according to the data collection instruction. The participants need to provide their subjective perception about the segment, including quietness, calmness, fun, easy of walking and crowdedness. To assign questionnaires to their corresponding checkpoints, we use k-means to group them into clusters based on time distribution and the number of clusters is equal to the number of checkpoints. Finally, we have 307 and 197 valid questionnaires for training and testing respectively and each questionnaire is represented by a 5-dimensional vector. The label (i.e. P-AQI) of questionnaires are directly derived from that of corresponding segments in a multiple instance learning strategy [1], since the P-AQI actually is generated based on the perceptions of all participants.

2.3 Urban Natures

Since questionnaire is inaccurate because of individual difference, we make use of OSM to have a relatively accurate description about the surrounding environment. In OpenStreetMap, each location is described by a JSON file, which contains the description of surrounding buildings, roads and scenery. We sample locations along the routine by sliding window with stride as 20 meters and a circle with radius as 25 meters is drawn around the location. Then we collect all the urban nature description inside this circle and all the name entities are extracted with Stanford CoreNLP [4]. Then, the name entities are manually split into road, building and scenery. Since human beings have different subjective perceptions under various urban nature, we further cluster the entities into different groups as shown in Table 1. Similar to questionnaires, All the locations

---

[1]www.openstreetmap.org

Table 1: The representative concepts in different classes of ubran nature are shown and the number in parentheses is the total number of concepts in corresponding class.

	A	B	C
road	subway, trunk (12)	lane (1)	pedestrian, footway (3)
building	shop, pitch (5)	museum, historic (3)	
scenery	no scenery (1)	forest, wetland, coastline (11)	

are assigned with the same label with that of the corresponding segment. Finally, each point is represented by a three-dimensional vector and there are 479 and 367 samples for training and testing respectively.

3 RUN DESIGN AND RESULT ANALYSIS

We have submitted five runs for each task to explore the impact of different factors on PEH.

3.1 Segment Replacement

The five runs are listed as following:

- MEAN: The mean value of PM2.5 from the other participants in the routine is calculated and the hidden values in the replaced segment are filled with it. This is the baseline run based on the assumption that the PM2.5 should not largely fluctuate within a small region.
- LOCA: The hidden values is replaced with the PM2.5 value of the nearest point from other participants. Note that the map distance between two points is adopted rather than l2 distance. This run also bases on the same assumption of MEAN but further narrow down the region area.
- GBDT: LightGBM [3] is adopted to model the relationship between temperature, humidity and PM2.5. By observing the data, we find that the distribution of the development and testing data are quite different. Hence, we directly train the model on the data collected from other participants in the same routine with the hidden segment.
- FLSTM: Since the distribution in development and testing data are different, we propose to prediction the fluctuations by subtracting the mean value of routines. LSTM is adopted to incorporate the contextual information for precise prediction and the input features are also temperature and humidity. The hyperparameters are set as following: batch size as 1,000, learning rate as 0.0006, hidden units as 40. Adam optimizer is adopted for 500 epochs.
- LSTM: As run GBDT, the LSTM is directly trained on the temperature and humidity collected from other participants in the same routine with the hidden segment. The settings of hyperparameters are the same as that of FLSTM.

As shown in Table 2, GBDT achieves the best performance and LSTM performs even worse than LOCA. We attribute the low performance of LSTM to the random state in PM2.5, which means that LSTM is overfitting to model the contextual information. As for

GBDT, it well models the relationship between low-dimensional factors and PM2.5 with multiple weak classifiers and avoids the overfitting problem because of limited model capacity. In addition, we also try to incorporate heart rate into the input feature and get worse performance in development data, which means that individuals' physical status may be insensitive to the subtle change in PM2.5.

Table 2: Segment Replacement

Runs	normalized L2
MEAN	0.0303
LOCA	0.0012
GBDT	0.0004
FLSTM	0.0201
LSTM	0.0219

Table 3: Personal Air Quality

Runs	L1 distance
QUEST	0.9
WEAR	1.5
OSM	0.7
QOMER	0.5
WOMER	0.9

3.2 Personal Air Quality

All the runs are trained with LightGBM [3] but on different features as listed. In the first three runs, the label of one segment is inferred by max-voting on all the samples inside it.

- QUEST: As described in Section 2.2, the questionnaires are represented by a five dimensional vector with the same label as corresponding segment.
- WEAR: The physical sensor factors, including temperature, humidity and PM2.5 are assigned with the same label as corresponding segment.
- OSM: As described in Section 2.3, the locations inside a segment with urban nature are used as training samples.
- WOMER: The results of WEAR and OSM are further fused by average voting.
- QOMER: The results of QUEST and OSM are further fused by average voting.

The performance is listed in Table 3. WEAR is the worst run, which means that physical sensors actually basically provide no clues for P-AQI prediction. Meanwhile, QUEST and OSM achieve better performance. Questionnaire is easy to be affected by individual difference and OSM provides a better description of the environment. Hence, OSM achieves the best performance as single feature. In addition, by fusing different features, we further find that questionnaire provides complementary information to OSM and achieve the best performance among all the five runs.

4 FUTURE WORK

While the findings in this study are interesting, however, they are not conclusive, in the way that they are made with a biased set of factors with a limited number of sensors and a limited coverage of participants. In the future, We will extend the study by including more sensors such as ECG, EEG, and blood sugar, and a larger coverage of participants across age, gender, occupation, and education.

REFERENCES

[1] Marc-Andr Carbonneau, Veronika Cheplygina, Eric Granger, and Ghyslain Gagnon. 2018. Multiple Instance Learning. *Pattern Recogn.* 77, C (May 2018), 329–353.

[2] Minh-Son Dao, Peijiang Zhao, Tomohiro Sato, and Koji Zettsu. 2019. Overview of MediaEval 2019: Insights for Wellbeing Task Multimodal Personal Health Lifelog Data Analysis. In *MediaEval 2019 workshop.*

[3] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Advances in Neural Information Processing Systems 30.*

[4] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations.* 55–60.

[5] Darshan Santani, Salvador Ruiz-Correa, and Daniel Gatica-Perez. 2018. Looking South: Learning Urban Perception in Developing Cities. *Trans. Soc. Comput.* 1, 3 (Dec. 2018).

[6] Tomohiro Sato, Minh-Son Dao, Kota Kuribayashi, and Koji Zettsu. 2019. SEPHLA: Challenges and Opportunities Within Environment - Personal Health Archives. In *MMM.*

[7] Hyeonjin Song, Kevin James Lane, Honghyok Kim, Hyomi Kim, Garam Byun, Minh Le, Yongsoo Choi, Chan Ryul Park, and Jong-Tae Lee. 2019. Association between Urban Greenness and Depressive Symptoms: Evaluation of Greenness Using Various Indicators. *Int J Environ Res Public Health* (Jan. 2019).