Towards Understanding Latent Relationships among Uncollectible Garbage and City Demographics

Koh Takeuchi, Yasue Kishino,
Yoshinari Shirai, Futoshi Naya, Naonori Ueda

NTT Communication Science Laboratories
2-4 Hikaridai, Kyoto 619-0237 Japan
Email: {takeuchi.koh, kishino.yasue, shirai.yoshinari,
naya.futoshi, ueda.naonori}@lab.ntt.co.jp

Takuro Yonezawa, Tomotaka Ito,
Jin Nakazawa

Graduate School of Media and Governance
Keio University
5322 Endo Fujisawa, Kanagawa 252-8520 Japan
Email: {takuro, tomotaka, jin}@ht.sfc.keio.ac.jp

Abstract—**We propose a preliminary work of a GIS-based participatory sensing system for collecting city information which works with a real-time collaboration of local government employees. With garbage management data collected by our application, we provide the running results of spatio-temporal data analysis that uncover the latent relationships among the spacial distributions of uncollectible garbage and city demographics. To discover such relationships, we conducted simple experiments that predicted the amount of uncollectible garbage from demographic and housing statistics by regularized linear regression methods and show that utilizing both population and housing statistics improved the predictive performance. We also report that the features of population and housing statistics, which are related to the lifestyles of citizens, greatly affect the amount of uncollectible garbage.**

Keywords–Participatory sensing, Data mining

I. Introduction

Understanding the relationships between citizen's activities and an urban information (e.g., types of housing) is one critical topic for both citizens and governments. These relationships could provide an understanding of cities from various viewpoints, so that we can utilize knowledge for urban management, urban planning and so on. For example, Budd [1] studied what type of house burglaries could be predicted from official British crime data. In a more recent study, Venerandi et al. [2] proposed a quantitative method to study the relationship between gang activity and a set of descriptors of urban forms extracted from open datasets for areas in London. Since providing security and safety are important for cities, these observations should encourage more practical urban planning.

In this paper, we focus on finding the relationship between garbage and urban information, including demographics and types of residences. Waste management and recycling are typical worldwide problems in various cities and countries for improving public health and reducing environmental footprints [3], [4]. In addition, since the cost of garbage management in cities is enormous (e.g., Fujisawa city in Japan annually pays more than seven billion yen for garbage-related city operations), garbage reducing initiatives are required for a city to be cost-effective [5]. Thus, the final goal of our research is to provide valuable knowledge for cities to reduce such management costs by analyzing the relationship between garbage and urban information.

Our paper tackles two challenges: collecting fine-grained garbage information in realtime and analyzing the latent relationship between the collected garbage information and the urban information. In most Japanese cities, the information about daily garbage amounts is measured and available only at garbage centers. There is no detailed information on collected garbage for each area of a city. Moreover, several types of specific garbage, such as illegally-dumped or uncollectible garbage, require a special cost on garbage management. However, such information has not been collected or stored at all. Thus, we propose a new system called MinaRepo that collects such information by participatory sensing of the daily tasks of city employee. Secondly, we investigate the basic statistics of the garbage data collected by MinaRepo and urban information. Then, we conduct a regression problem that predicts the amount of garbage in urban sites based on the population and the housing statistics of Fujisawa city and we identify the latent relationships among garbage and city demographics.

In summary, contributions of the paper are as follows:

- We introduce MinaRepo, a new way to collect fine-grained garbage-related information by piggybacking on the daily tasks of city employees.
- We reveal the latent relationships among garbage-related information and city demographics based on statistical machine learning methods.

II. Related works

Various kinds of attempts on urban waste monitoring, management, and its related technologies can be found in [6] and references therein. Recently, a remote sensing system has become a popular technology in those research topics. A combination of remote sensing and machine learning method was proposed by [7]. They conducted a linear regression problem to estimate the amount of solid waste produced by commercial activities on urban sites, but no population and household statistics were considered in their analysis.

Great efforts have been made to improve and deploy participatory sensing technology [8]–[10].These works show the possibility of participatory sensing involving the citizens. In contrast to these works, we especially focus on city officers as target users. In addition, our purpose is to provide efficiency to their daily works to gather a lot of data with correct labels.

In the context of sustainability of cities and environment, statistical analysis methods were utilized to detect relationships between recycling and consuming behaviors [11], [12]. However, those articles did not focus on revealing relationships among uncollectable garbage and city demographics and adopted participatory sensing technologies for gathering datasets.

III. MINAREPO

A. Expert Crowd Sensing

To understand a citys situation or its citizens activities, it is necessary to obtain various city related information generated by citizens. One way to get such information is by monitoring the daily works of city administrative employees because they are the ones dealing with city information and citizens demands on a daily basis. For example, Fujisawa city has about $3,500$ local government employees, some of whom work outside of city hall, in areas such as garbage collection, firefighting, or maintaining city infrastructure like roads. By integrating their daily works with city sensing, large amounts of useful data can be collected every day. We call such piggyback sensing on the daily work of city employees *Expert Crowd Sensing*. In addition, city employees are expected to have more task specific knowledge of the city in which they are working than general residents. For example, garbage collectors, who daily travel through their assigned area, should be knowledgeable about the whole area and could detect subtle changes in a city. Consequently, expert crowd sensing enables us to acquire more accurate human sensing data with greater sensitivity. To achieve expert crowd sensing, a sensing tool must satisfy the following requirements:

- Easy integration with the daily responsibilities of city employees
  To enhance their daily work by such sensing, the tool must not disturb their current daily activities and has to efficiently support them.
- Easy usage
  Since most city employees are not information technology experts, our tool has to be easily understood and simply leveraged.
- Providing reliability, dependability, and security
  The tool must be used in daily city functions, which sometimes provide critical services for citizens. Reliability and dependability are also needed for it. Since city tasks often deal with the private data of citizens, data must be securely protected.

To satisfy the above requirements, we took an approach of user-engaged system development. We first interviewed city employees about the details of their work flow in their daily works and what kind of problems they presently face. We also specified the types of city data that can be collected by their works. After designing and implementing a prototype system, city employees reviewed it and gave feedback to us. Then, we reflected on such feedback and incorporated it into our system. Such user-engaged development and refinement were repeated several times.

B. System Design and Implementation

Through user-engaged development, we developed MinaRepo, an expert crowd sensing system that satisfies the above requirements. MinaRepo is composed of smartphone applications and server-side software. Fig. 1 represents its usage flow. If a city employee notices incidents that should be reported, he/she opens the MinaRepo application of his/her smartphone/tablet. First, he/she selects the type of a report based on such city incidents as illegally-dumped garbage, graffiti, uncollectible garbage, road damage, and so on. Then he/she takes a photo of the city incident and inputs a brief description for the report: ”these garbage items are plastic bottles and are not allowed to be dumped in this area.” At the same time, the GPS(Global Positioning System) location information is automatically added to the report. After inputting the needed information, he/she sends the report to the MinaRepo server, where it is stored in a secure database. The report can be accessed by a web interface, which visualizes all the reports with tables and a map interface, where city employees can check the details of all reports by clicking on the report item. In addition, the web interface provides a search functionality and filters the reports by type, reporter’s name, or date. If an additional action is needed (e.g., erasing graffiti or disposing of illegally-dumped garbage), city employees can contact the appropriate city officials.

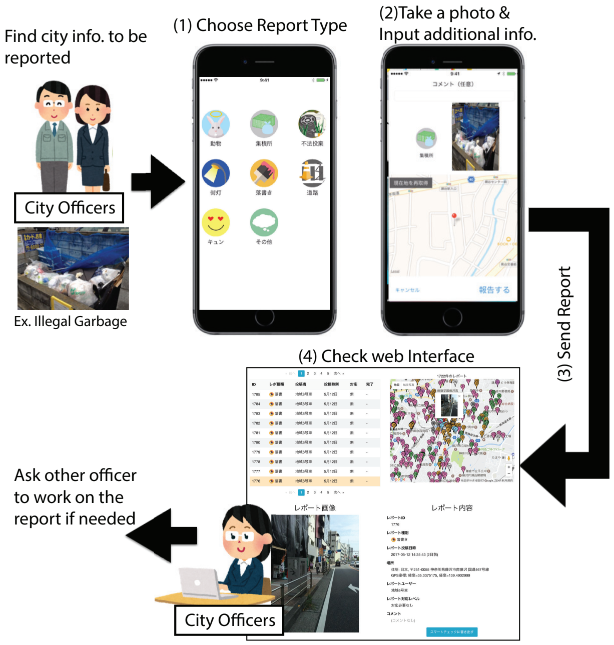
Find city info. to be reported
(1) Choose Report Type
(2)Take a photo & Input additional info.
City Officers
Ex. Illegal Garbage
(3) Send Report
(4) Check web Interface
Ask other officer to work on the report if needed
City Officers

Figure 1. Overview of MinaRepo work flow

We first started to collaborate with the garbage-collecting section of Fujisawa city. Through engaged development, we confirmed that the work flow must comfortably fit their daily works. Usually, those reports are shared among city employees in traditional analog ways; when a city employee notices a city incident, he phones a manager and describes it. Then the manager records it in a map document and faxes it to another employee. This traditional procedure is time-consuming and wastes human resources. MinaRepo enables city employees to report city incidents easily and efficiently.

We defined seven types of reports for the sensing tasks of the garbage section and identified three types of actions needed in their work: urgent action is needed, action is needed but not urgent, and no action is needed. We also provide an interface with which to choose the type of action required for a report. When an ”urgent action is needed” report is input, an e-mail, which includes detailed information of the report, is automatically sent to city personnel. This functionality increases efficiency.

For providing system reliability and dependability, we also developed automatic system monitoring and backup functionality. When our systems is out-of-use, city employees can get status notification by e-mail so that they can work using their traditional procedures.

IV. Summary of the MinaRepo dataset

Our dataset, which consists of $1,173$ reports recorded from October 6, 2016, to April 25, 2017, includes seven types of labels. The first label, residue, reports garbage that was dumped or discarded in the wrong place or on a wrong day. The second label, forgotten-garbage, is legal solid waste that was overlooked by the garbage trucks. The third label, illegally-dumped garbage, is solid waste that cannot be legally picked up in Fujisawa city. The fourth label, garbage-station, indicates a report about garbage problems at the specially designated drop sites. The fifth label, graffiti, reports a place where graffiti has been written on walls or buildings. The sixth label, animal corpse, denotes where an animals body has been found. The remaining label, disaster, report places where a relatively major incident happened and the damage caused by it. Examples of reports with residue and graffiti labels are shown in Figs. 2 and 3.

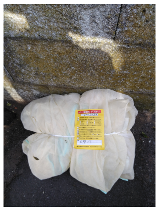

Figure 2. Photo from reports with residue label

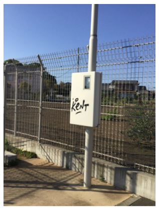

Figure 3. Photo from reports with illegal grafitti label

Reports were submitted by 55 users. The average number of daily reports was 8.65. We show the numbers of reports by label and user in Figs. 4 and 5. Residue labels were the most frequently reported type. Graffiti reports were the second largest type. Fig. 5 indicates that our system works with users of different activity levels, including highly active users whose number of reports are in the hundreds. In contrast, about 63% of the users submitted fewer than ten reports in this period because they had just started to use our application.

The time series plot of the numbers of reports is shown in Fig. 6. One particular day got more than 100 reports because on that day the local government was sponsoring a special campaign to detect graffiti. The Lag-N autocorrelations of this time series with $N = (1, 2, 3, 4, 5, 6, 7)$ were $-0.07, 0.29, 0.19, -0.01, -0.02, 0.02$, and 0.27, respectively. Thus no significant autocorrelation was found from this dataset. We also checked the spatial autocorrelation of each label. The spatial autocorrelations for the ratio of each label by population were calculated on 192 areas of administrative districts with K-nearest neighbor graphs ($K = 4, 5, 6, 7, 8, 9, 10$). Moran’s I value with the highest absolute value and its p-values for each label are shown in Table I. The residue label got a positive spatial autocorrelation, and its p-value was smaller than 0.01. Thus, reports with this label are suitable for detailed statistical analysis.

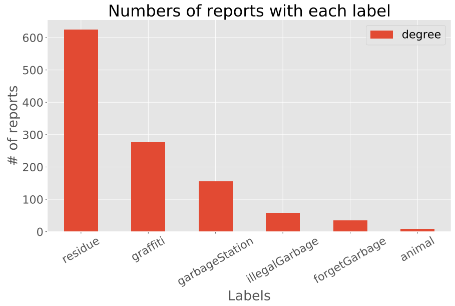

Figure 4. Number of reports with seven labels

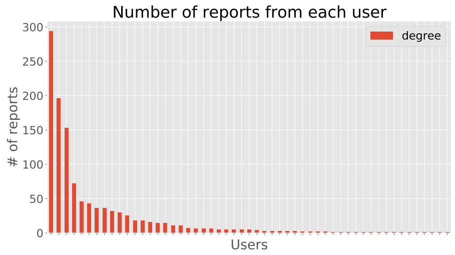

Figure 5. Number of reports from 61 users

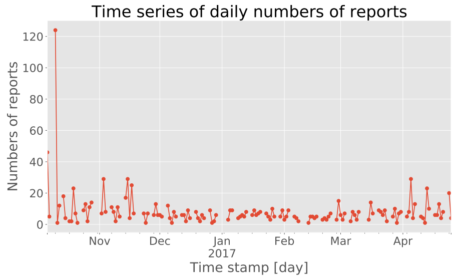

Figure 6. Daily reports submitted on weekdays

TABLE I. Spatial autocorrelation of each label

Label	*Moran’s I*	*K*	*p-value*
Residue	0.154	6	0.000022
Graffiti	0.075	5	0.030596
Garbage-station	0.116	4	0.005596
Illegal-dumped Garbage	0.025	3	0.290003
Forgotten-garbage	0.035	4	0.197829

We utilized the demographic and housing statistics and describe basic summaries to understand the spatial demographics in this city. Housing statistics consisted of information about rental properties, such as identifying whether they are condos, apartments, or houses. From this dataset, we obtained 87 features that include the number of renting rooms, the average housing prices, average occupied areas, and so on. Other characteristic features also seem to have correlations with the amount of uncollectible garbage, for example, information about eligible renters, since some rooms can only be inhabited by one person, and such convenient amenities, including automatically locking doors, lockers for deliveries, and access to the internet. We averaged these values for each site. To compare our dataset and these statistics, we illustrated the amount of residue per population, the spatial density of the population, and the number of apartments per area in Figs. 7, 8, and 9.

V. Experiments

We conducted experiments that predicted the amount of uncollectible garbage with residue label by population on each site. We employed linear regression methods: Elastic Net, Lasso, Ridge, and Ordinal Least Squares (OLS) [13]. We denote the target values, input values, and coefficient vectors as $y_n \in \mathbb{R}$, $\boldsymbol{x}_n \in \mathbb{R}^d$ for $n = (1, \cdots, N)$, and $\boldsymbol{\beta} \in \mathbb{R}^d$, respectively. Then, we defined the linear regression problem with both the $\ell 1$ and $\ell 2$ penalty terms as:

$$\min_{\boldsymbol{\beta}} \sum_{n=1}^{N} \|y_n - \boldsymbol{\beta}^\top \boldsymbol{x}_n\|_2^2 + \lambda_1 \sum_{i=1}^{d} |\beta_i|_1 + \lambda_2 \sum_{i=1}^{d} \|\beta_i\|_2^2,$$

where we denote the hyper parameters as $\lambda_1, \lambda_2 \in \mathbb{R}$. When $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$ this method corresponds to Elastic net and includes Lasso, Ridge and OLS as special cases with ($\lambda_1 \neq 0$ and $\lambda_2 = 0$), ($\lambda_1 = 0$ and $\lambda_2 \neq 0$), and ($\lambda_1 \neq 0$ and $\lambda_2 \neq 0$), respectively.

We used the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) to assess the predictive performance of these methods,

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \|y_n - \boldsymbol{\beta}^\top \boldsymbol{x}_n\|_2^2},$$

$$\text{MAE} = \frac{1}{N} \sum_{n=1}^{N} |y_n - \boldsymbol{\beta}^\top \boldsymbol{x}_n|.$$

We employed three types of input features in our experiments in which we used housing statistics ($d = 87$), and population statistics ($d = 12$), and both ($d = 99$) were used as inputs. We randomly picked 80% of the 192 sites as a training data set and used the rest as a test dataset. Hyperparameters λ_1 and λ_2 for the regularized linear regression methods were selected by one-leave-out cross validation. We ran 10 experiments and got the average and the standard deviations of the predictive errors.

We show the experimental results in Tables II and III. Ridge achieved the best predictive performances on RMSE and MAE. Utilizing both the housing and population statistics improved the performances over just using the housing or population statistics.

To check the effects of the input features, we show the coefficient values learned by Ridge in Fig. 10. We also show the top 15 highest or lowest coefficient values and their correlation with the target values in Tables IV and V. We confirmed that the features of the housing statistics obtained the highest and lowest coefficient values among the features whose correlation was also high or low. We found various kinds of input features, which might indicate the relationship

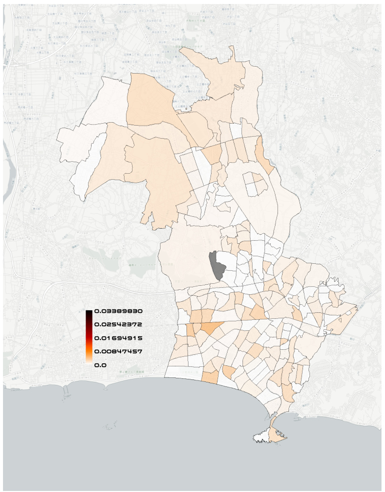

Figure 7. Amount of residue per population

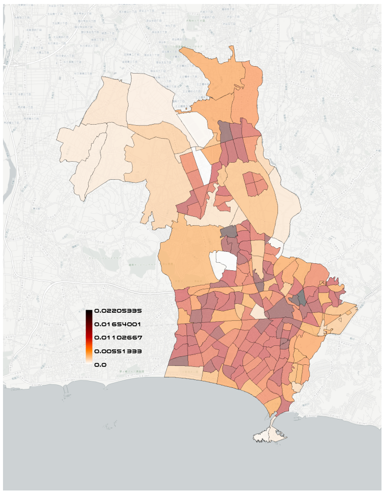

Figure 8. Spatial density of population

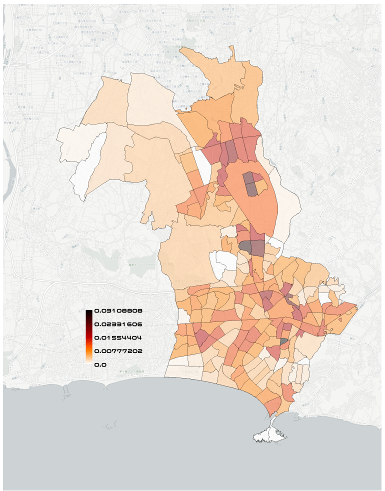

Figure 9. Number of apartments per area

ISBN: 978-1-61208-569-2

between garbage and city demographics, in Tables IV and V. For example, housing characteristics, such as home security companies, single-rentals only, self-locking doors, and free internet, all of which are favored by young people living alone, got the highest coefficient values. The ratio of 30's and 50's inhabitants, which were features of population statistics, obtained high values. On the other hand, different features such as apartments that allow room-sharing also had high values. This feature seems to be favored by young people with roommates; the popularity of such room-sharing is increasing in Japan. In contrast, with Table V, we found completely contrary features in Table IV. Such housing features as condominiums, terraces, lightings, and floor heating, which seem to be favored by families or senior citizens, had lower values. Population statistics including 70's and 90's inhabitants also obtained lower values.

TABLE II. AVERAGE AND STANDARD DEVIATION OF RMSE FOR PREDICTING SOLID WASTE AMOUNTS (RMSE$*10^3$))

Method	*Housing*	*Population*	*Housing + Population*
Elastic Net	1.44(0.28)	1.42(0.28)	1.43(0.27)
Lasso	1.52(0.21)	1.51(0.20)	1.52(0.20)
Ridge	**1.39(0.19)**	**1.36(0.22)**	**1.35(0.21)**
OLS	2.44(0.69)	1.50(0.23)	2.73(0.71)

TABLE III. AVERAGE AND STANDARD DEVIATION OF MAE FOR PREDICTING SOLID WASTE AMOUNTS (MAE$*10^3$)

Method	*Housing*	*Population*	*Housing + Population*
Elastic Net	1.10(0.16)	1.08(0.17)	1.10(0.16)
Lasso	1.13(0.09)	1.14(0.09)	1.14(0.09)
Ridge	**1.07(0.09)**	**1.05(0.11)**	**1.05(0.11)**
OLS	1.73(0.26)	1.12(0.13)	1.94(0.32)

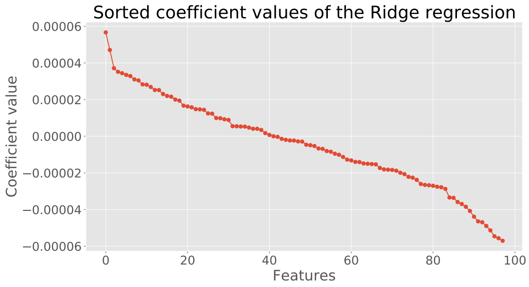

Figure 10. Learned coefficient values.

TABLE IV. FEATURES WITH HIGHEST COEFFICIENT VALUES

Feature	*Coefficient$*10^5$*	*Correlation*
Home security company	5.529	0.161
Single-rentals only	4.745	0.132
Toilet room	3.715	0.168
Bath room	3.678	0.158
30's	3.672	0.149
Room-sharing is allowed	3.408	0.087
Amounts of apartments looking for residents	3.224	0.126
Underfloor storage	3.079	0.092
Self-reheating bath	2.894	0.095
Self-locking doors	2.854	0.087
Free internet	2.545	0.075
Bathroom vanity	2.539	0.088
Pets are allowed	2.506	0.045
Tiled floors	2.329	0.052
50's	2.131	0.043

TABLE V. FEATURES WITH LOWEST COEFFICIENT VALUES

Feature	*Coefficient$*10^5$*	*Correlation*
Condominium	-5.705	-0.157
Terrace	-5.564	-0.124
Internet available at charge	-5.458	-0.177
70's	-5.133	-0.203
No room-sharing	-4.894	-0.205
Gus stove	-4.691	-0.121
Lighting	-4.645	-0.099
Free rent	-4.388	-0.101
Floor heating	-4.077	-0.097
Storage loft	-3.849	-0.179
Roommates are allowed	-3.697	-0.087
Access to parking	-3.59	-0.088
90's	-3.362	-0.143
Connected to sewage	-3.343	-0.045

VI. CONCLUSION

In this paper, we proposed a novel application for gathering information on uncollectible garbage in a city. We also showed the basic summaries of our dataset and visualized information such as housing and population statistics. To understand the relationships between garbage and demographics, we conducted simple regression problems and discovered a set of features that increases or decreases the amount of uncollectible garbage.

REFERENCES

[1] T. Budd, "Burglary of domestic dwellings: Findings from the british crime survey," 1999.

[2] A. Venerandi, G. Quattrone, and L. Capra, "Guns of brixton: Which london neighborhoods host gang activity?" in Proceedings of the Second International Conference on IoT in Urban Space, 2016.

[3] D. Hoornweg and L. Thomas, What a waste: solid waste management in Asia. The World Bank, 1999.

[4] L. A. Guerrero, G. Maas, and W. Hogland, "Solid waste management challenges for cities in developing countries," Waste management, vol. 33, no. 1, 2013, pp. 220–232.

[5] K. Palmer, H. Sigman, and M. Walls, "The cost of reducing municipal solid waste," Journal of Environmental Economics and Management, vol. 33, no. 2, 1997, pp. 128–150.

[6] M. Hannan, M. A. Al Mamun, A. Hussain, H. Basri, and R. A. Begum, "A review on technologies and their usage in solid waste monitoring and management systems: Issues and challenges," Waste Management, vol. 43, 2015, pp. 509–523.

[7] N. V. Karadimas and V. G. Loumos, "Gis-based modelling for the estimation of municipal solid waste generation and collection," Waste Management & Research, vol. 26, no. 4, 2008, pp. 337–346.

[8] R. K. Rana, C. T. Chou, S. S. Kanhere, N. Bulusu, and W. Hu, "Ear-phone: An end-to-end participatory urban noise mapping system," in Proceedings of IPSN, 2010.

[9] S. Kim, J. Mankoff, and E. Paulos, "Sensr: Evaluating a flexible framework for authoring mobile data-collection tools for citizen science," in Proceedings of CSCW, 2013.

[10] M.-R. Ra, B. Liu, T. F. La Porta, and R. Govindan, "Medusa: A programming framework for crowd-sensing applications," in Proceedings of MobiSys, 2012.

[11] I. E. Berger, "The demographics of recycling and the structure of environmental behavior," Environment and behavior, vol. 29, no. 4, 1997, pp. 515–531.

[12] A. Diamantopoulos, B. B. Schlegelmilch, R. R. Sinkovics, and G. M. Bohlen, "Can socio-demographics still play a role in profiling green consumers? a review of the evidence and an empirical investigation," Journal of Business research, vol. 56, no. 6, 2003, pp. 465–480.

[13] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 67, no. 2, 2005, pp. 301–320.