Behaviour Profiles for Evidence-based Policies Against Obesity

Ioannis Sarafis, Christos Diou and Anastasios Delopoulos

Abstract— **Obesity is a preventable disease that affects the health of a significant population percentage, reduces the life expectancy and encumbers the health care systems. The obesity epidemic is not caused by isolated factors, but it is the result of multiple behavioural patterns and complex interactions with the living environment. Therefore, in-depth understanding of the population behaviour is essential in order to create successful policies against obesity prevalence. To this end, the BigO system facilitates the collection, processing and modelling of behavioural data at population level to provide evidence for effective policy and interventions design. In this paper, we introduce the behaviour profiles mechanism of BigO that produces comprehensive models for the behavioural patterns of individuals, while maintaining high levels of privacy protection. We give examples for the proposed mechanism from real world data and we discuss usages for supporting various types of evidence-based policy design.**

I. INTRODUCTION

The obesity prevalence is soaring in children and adolescents and constitutes a major public health problem that burdens societies and affects the well-being of individuals [1]. Obesity is a multifactorial problem rooted in complex behaviour patterns; however, current policies aiming to halt the epidemic tend to be fragmented and, eventually, ineffective [2]. Furthermore, the policy making process is weakened by the lack of systems and methods that monitor and model population behaviour and the quantify effect of policies and interventions [3], [4].

To this end, the central aim of BigO system* is to facilitate counter-obesity policy making at population level based on measurements of the population's behavioural characteristics and the environmental factors. Figure 1 illustrates the behavioural data collection and processing pipeline of BigO, which includes the following stages:

1) Large scale behavioural data collection: Volunteer students and students participating in organised school efforts contribute data using the BigO smartphone application. Data from the everyday life of the participating individuals are collected using sensors (e.g., accelerometer, GPS, heart rate) of smartphones and smartwatches. In addition, the individuals contribute self-reported data (e.g., questionnaires, food photos, food advertisement photos) using the smartphone application.

*The work leading to these results has received funding from the European Community's Health, demographic change and well-being Programme under Grant Agreement No. 727688 (`http://bigoprogram.eu`), 01/12/2016 - 30/11/2020. The Institution's Ethical Review Board approved all experimental procedures involving human subjects.

All authors are with Multimedia Understanding Group, Information Processing Laboratory, Aristotle University of Thessaloniki, Greece.

2) Behaviour indicator extraction: The collected raw sensor measurements are processed to extract *behaviour indicators*, which quantify various behaviours linked with healthy or unhealthy lifestyles (e.g., physical activity levels during the day, visits to fast food outlets).

3) Behaviour models for individuals: The behaviour indicators are processed to produce meaningful summarisations that describe the overall behaviour for each individual. Essentially, the goal of this stage is to model the individual's habits, including: (a) the type of facilities visited; (b) the exhibited behaviour during the visits; and (c) the mobility patterns and transportation preferences between the facilities. At the same time, we must ensure high level of privacy protection. The more important output of this stage are the *behaviour profiles* generated for the participating individuals.

4) Analysis methods for population: The last stage is to use the behaviour models to apply data analysis for the monitored population. The goal of this stage is to support various forms of policies by providing evidence for: (a) population groups with similar behaviours; (b) associations between certain behavioural patterns and the living environment; and (c) the efficacy of possible interventions and their impact on population or targeted population groups.

The data collection (stage 1) and behaviour indicator extraction (stage 2) can be supported sufficiently by the current technology and state-of-the-art algorithms. However, existing behaviour modelling methods (stage 3) are not verbose enough to describe the behavioural aspects that play key role in obesity development and, subsequently, methods for supporting evidence-based policies at population level (stage 4) have not emerged.

The shortcomings of current behavioural modelling methods originate from the fact that they were developed mostly for network infrastructure management [5], [6]. For example, they distinguish the Points of Interest (POIS) into a small number of categories and ignore information about the visited facilities. In addition, they do not model important behavioural aspects during POI visits (e.g., exhibited physical activity levels, facility types) or the preferences regarding transportation between POIs (e.g., transportation method, distance, physical activity levels).

In this work, we present an overview of BigO's processing and analysis pipeline for behavioural data. The presentation focuses on the proposed behaviour profiles mechanism and its fitness for daily habits and lifestyle modelling. It is complemented with a complete profile example from real world data and a discussion on possible use cases.

II. MEASUREMENTS OF BEHAVIOUR

In BigO, the behaviour indicators are measurable quantities that provide information for an individual's behaviour. The behaviour indicators fall into two categories: (a) self-reported, based on the direct user feedback through the smartphone application of BigO (e.g., questionnaires); or (b) automatically extracted, using the sensors of smartphones and smartwatches. For example:

- Using the 3-dimensional accelerometer signal, we automatically extract physical activity indicators, such as activity counts [7] and activity type (sitting, walking, running, completing household chores, etc) [8]. Typically, these indicators are calculated at minute intervals.
- Using the GPS location signal, we calculate indicators, such as: the individual's POIs [9] and the POI types (public parks, food outlets, recreation facilities, etc.) using external data sources (Google maps, Foursquare).
- Combining accelerometer and GPS signal, we calculate transportation mode indicators (e.g., walking, bicycle, bus or car) [10].

The self-reported indicators tend to be unreliable, whereas the automatically extracted ones provide objective, continuous measurements of behaviour. To this end, we rely on the latter for behaviour modelling (see also Section III).

A. *Limitations of the behavioural indicators*

The extraction of the behavioural indicators from recorded signals is the first step to measure behaviour. Nevertheless, there are various behavioural aspects that cannot be captured solely with the use of indicators. For example, consider the following questions:

- When the individual is at home in the afternoon, what facility types are more likely to be visited afterwards?
- What are the physical activity levels at these facilities?
- Does the individual exhibit sedentary behaviour when staying at home during afternoon hours?

The common element in these questions is the temporal aspect of human behaviour. Our ability to answer these questions is important since the temporal characteristics of the behaviour reflect the individual's daily habits and, subsequently, are linked with the risk of developing obesity.

To this end, we propose a methodology to systematically model the temporal characteristics of behaviour and the exhibited behaviour at different POI types, presented next.

III. BEHAVIOUR PROFILES

BigO uses two mechanisms to process the behaviour indicator time series in order to produce concise summarisations of the human behaviour, following a privacy-by-design approach. Section III-A presents the timelines mechanism, which provides the first level of summarisation; and Section III-B presents the behavioural profiles mechanism, which aggregates the timelines to uncover recurring behavioural patterns.

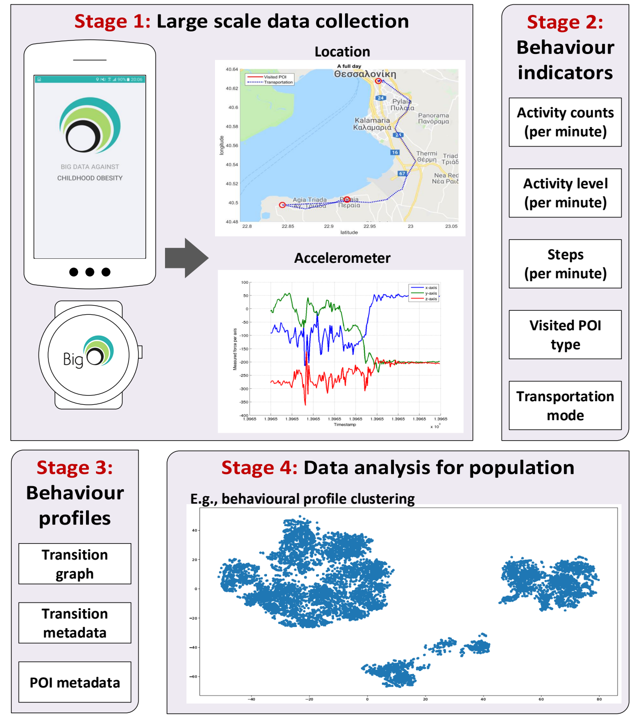

Fig. 1. The behavioural data collection and processing pipeline of BigO

A. *Timelines*

The timelines mechanism performs a summarisation of the temporal behaviour for an individual, for a specific day. As we will see next, the most important advantage is the high level of privacy protection since the actual visited locations are not stored in the timeline, only their types.

We extract the timelines using modified versions of the DBSCAN algorithm specialised for geospatial trajectories (for example, the algorithm of Lue *et al.* [9]). The resulting timeline is a sequence of "stop" and "move" events. The "stop" events occur at the individual's POIs. Each detected event contains all the behaviour indicators time series recorded during its occurrence, except the location co-ordinates which are excluded for privacy protection.

For this reason, instead of the actual POI location, we perform geospatial queries to data sources (such as Foursquare and Google maps) using the recorded location co-ordinates. Then, we assign the event with the POI type according to a majority voting scheme for the retrieved facility types.

We look for facilities in categories: "School", "Fast food or take away", "Restaurant", "Café or café bar", "Supermarkets or grocery stores", "Wine and liquor stores", "Public parks", "Athletics and sports" and "Indoor recreation facilities". If no facilities in these categories of interest are retrieved, then the POI type is assigned as "Other".

Regarding the home of an individual, we apply heuristics for its detection based on that the home should be the first and the last visited POI almost each day. In addition, we split the home POI into two types based on the time of day: "Home before 12:00" and "Home after 12:00". The reason for this separation is that different transition patterns are exhibited when individuals leave home early in the morning and when they leave home during the rest of the day. Also, the behaviour at home may vary between morning and afternoon hours (for example, physical activity levels and activity types).

In the end, “stop” events consist of: start/end timestamps, the recorded behaviour indicators and the POI type. Whereas, “move” events consist of: start/end timestamps, POI type of origin, POI type of destination, the recorded behaviour indicators, the travel distance and the transportation mode.

Based on the individual’s timelines, we proceed to calculate an overall summarisation of the individuals behaviour using the behaviour profiles mechanism, described next.

B. Behavioural profiles

The behaviour profiles mechanism processes the individual’s timelines to build an overall behaviour summarisation, in a privacy-preserving manner. Because different behaviour is exhibited between school and non-school days (or working and non-working days for adults), we calculate two behavioural profiles per individual. An example behavioural profile is given in Section IV.

A behavioural profile consists of three parts:

1) Transition graph: The transition graph captures the general mobility patterns. We will follow the assumption that the individual’s timelines are the result of a first order Markov chain, similar to [6]. This is a logical assumption since each individual tends to have repeating transition patterns between POI types for the same type of day (school or non-school), and it is safe to assume that the underlying mobility patterns do not change rapidly [11].

Thus, the transition graph of a behaviour profile is a directed graph describing a Markov process, where the edge starting from POI type i and ending to POI type j has a transition probability:

$$P_{ij} = Pr\{\text{Transition from POI type } i \text{ to POI type } j | \\ \text{Individual is at POI type } i\}$$

2) Transition metadata: Apart from the transition frequencies, we are also interested in the modelling the individual’s behaviour and transportation preferences between POI types. To this end, each edge of the transition graph with non-zero P_{ij} is accompanied with transition metadata that aggregate its characteristics.

The transition metadata contain the transportation mode distribution, which is a probability mass function (pmf) with domain the types of transportation mode. In addition, for each transportation mode with non-zero probability, the transition metadata hold the average travel distance and the average travel duration.

3) POI metadata: The transition graph and transition metadata offer a thorough summarisation of the individual’s mobility habits. To complete the behavioural model, we need to summarise how the individual behaves during a visit at different POI types (i.e., during the “stop” timeline events). This information is provided by the POI metadata. For a given POI type, the metadata contain: a probability distribution for the duration of “stop” events; and for each behaviour indicator, a probability distribution using the measurements recorded during “stop” events.

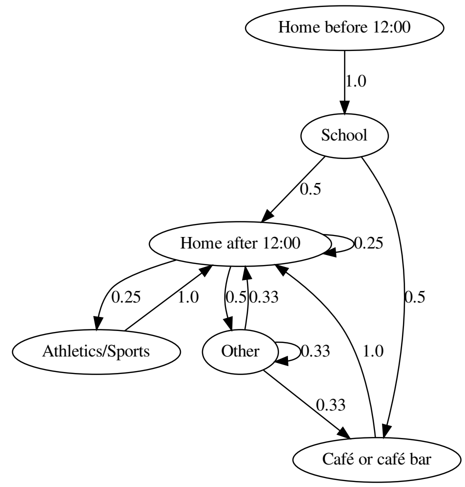

Fig. 2. Example graph visualisation for a student that participated in a BigO pilot. The transition probabilities were calculated on the timelines of 10 school days.

IV. AN EXAMPLE FROM REAL WORLD DATA

To comprehend the nature of the behavioural profiles, this section presents an example generated from real world data collected from BigO pilots.

Figure 2 shows the transition graph calculated using the timelines of 10 school days for a participating student. Table I shows the corresponding transition metadata for edges with $P_{ij} \neq 0$. In the current implementation, the transportation mode is distinguished between active type (e.g., walking, bicycle) or vehicle type (e.g., bus, car or other vehicle). Table II shows the metadata for the visited POI types for the same student, using: (a) normal distributions to model the activity counts indicator; and (b) a pmf to model stay duration, defined on the minute intervals: $[1, 10)$, $[10, 30)$, $[30, 60)$, $[60, 120)$, $[120, 240)$, $[240, 480)$, $[480, 1440]$.

Overall, we can see that the student: (a) transports using a vehicle (bus, car) for the more frequent transitions; (b) commutes actively only to POIs that are in close distance from home; (c) is sedentary at all POI types except “Athletics/Sports”; however, (d) visits to the latter are fairly rare.

Viewing this example behaviour profile from a policy makers perspective, gives us directions for possible interventions, for example: to increase physically active commute between POIs; or to make visits to “Athletics/Sports” more frequent?

Although these questions reveal possible uses of behaviour profiles, their more prominent use is for population groups. Next section discusses possible use cases for behaviour profiles at larger scale.

V. USE CASES

The concise representations given by behaviour profiles are most useful when applied at large scale; that is, when there are data for a significant number of individuals. Then, behaviour profiles can support multiple types of evidence-based policy making. For example:

TABLE I
THE TRANSITION METADATA OF THE EXAMPLE BEHAVIOUR PROFILE

Transition			walking or bicycle			car, bus or other vehicle		
			%	distance (km)	duration (min)	%	distance (km)	duration (min)
Home before 12:00	→	school	0	–	–	100	5.0	18.5
school	→	café or café bar	0	–	–	100	4.6	24.0
Home after 12:00	→	Athletics/sports	0	–	–	100	2.8	11.0
Athletics/sports	→	Home after 12:00	0	–	–	100	2.8	22.0
School	→	Home after 12:00	0	–	–	100	5.0	10.5
Café or café bar	→	Home after 12:00	100	0.4	5.0	0	–	–
Other	→	Home after 12:00	100	0.3	6.0	0	–	–
Home after 12:00	→	Other	100	0.8	7.0	0	–	–
Other	→	Café or café bar	100	0.4	4.0	0	–	–
Home after 12:00	→	Home after 12:00	100	0.7	15.0	0	–	–
Other	→	Other	75	0.1	1.0	25	2.9	11.5

1) Detecting associations between obesogenic factors of the environment and behaviours: Using statistical inference methods, we can quantify associations between behaviours and multiple environmental factors in the area of residence, such as: density of facilities types (fast food outlets, public parks, athletics etc.), availability of public transportation, and socioeconomic factors (e.g., average income or unemployment rate in neighbourhood).

2) Finding interesting population clusters: We can use unsupervised data analysis methods (for example, graph similarity methods [12]) to identify population clusters that exhibit similar behaviours. Then, we can provide insights regarding commonly occurring behaviours within the cluster using the POI metadata and/or the transition metadata. Furthermore, if measurements of the environment are available, we may recognise the possible presence of common, latent environmental factors affecting a cluster's behaviour.

3) Predicting the effect of interventions: Using supervised learning methods, we can build models that predict specific aspects of behaviour (i.e., behaviour indicators) as a function of the environment measurements. These models may estimate the expected behavioural change as a result of possible interventions in the environment, hence they can be used to prioritise possible interventions according to the expected impact on the population's behaviour. In addition, such models could estimate behavioural indicators for locations where no measurements from individual children are available.

TABLE II
THE POI METADATA OF THE EXAMPLE BEHAVIOUR PROFILE

POI type	Activity counts/min	Stay duration (min)
Home before 12:00	$\mathcal{N}(\mu = 528, \sigma = 105)$	$[10, 30) : 29\%$, $[30, 60) : 14\%$, $[480, 1440) : 57\%$
Home after 12:00	$\mathcal{N}(\mu = 528, \sigma = 139)$	$[1, 10) : 11\%$, $[10, 30) : 11\%$, $[30, 60) : 11\%$, $[60, 120) : 11\%$, $[240, 480) : 45\%$, $[480, 1440) : 11\%$
School	$\mathcal{N}(\mu = 531, \sigma = 171)$	$[240, 480) : 100\%$
Café or café bar	$\mathcal{N}(\mu = 523, \sigma = 124)$	$[30, 60) : 100\%$
Athletics/Sports	$\mathcal{N}(\mu = 4339, \sigma = 1581)$	$[60, 120) : 100\%$
Other	$\mathcal{N}(\mu = 472, \sigma = 205)$	$[1, 10) : 50\%$, $[10, 30) : 25\%$, $[30, 60) : 25\%$,

VI. CONCLUSIONS

We have outlined a methodology for modelling the behaviour of individuals that can be applied at population scale. The presented behaviour profile mechanism offers concise summarisations for the individual's habits, which include the temporal aspects of behaviour, the mobility patterns and the behaviour at visited POI types. With the collection of large scale data still on-going by BigO, the next step is to evaluate the use of behaviour profiles in analysis methods that aim to understand the complex characteristics of human behaviour and its relation with obesogenic environmental factors.

REFERENCES

[1] World Health Organisation, "Obesity and overweight, fact sheet," 2016, https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight, Last accessed on 2019-02-12.

[2] C. A. Roberto *et al.*, "Patchy progress on obesity prevention: emerging examples, entrenched barriers, and new thinking," *The Lancet*, vol. 385, no. 9985, pp. 2400 – 2409, 2015.

[3] S. Kleinert and R. Horton, "Rethinking and reframing obesity," *The Lancet*, vol. 385, no. 9985, pp. 2326 – 2328, 2015.

[4] B. Giles-Corti *et al.*, "City planning and population health: a global challenge," *The Lancet*, vol. 388, no. 10062, pp. 2912 – 2924, 2016.

[5] M. Papandrea, K. K. Jahromi, M. Zignani, S. Gaito, S. Giordano, and G. P. Rossi, "On the properties of human mobility," *Computer Communications*, vol. 87, pp. 19 – 36, 2016.

[6] K. K. Jahromi *et al.*, "Simulating human mobility patterns in urban areas," *Simulation Modelling Practice and Theory*, vol. 62, pp. 137 – 156, 2016.

[7] W. W. Tryon and R. Williams, "Fully proportional actigraphy: A new instrument," *Behavior Research Methods, Instruments, & Computers*, vol. 28, no. 3, pp. 392–403, Sep 1996.

[8] A. Reiss and D. Stricker, "Creating and benchmarking a new dataset for physical activity monitoring," in *Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments*. New York, NY, USA: ACM, 2012, pp. 40:1–40:8.

[9] T. Luo *et al.*, "An improved dbscan algorithm to detect stops in individual trajectories," *ISPRS International Journal of Geo-Information*, vol. 6, no. 3, 2017.

[10] M. A. Shafique and E. Hato, "Travel mode detection with varying smartphone data collection frequencies," *Sensors*, vol. 16, no. 5, 2016.

[11] M. Papandrea *et al.*, "How many places do you visit a day?" in *2013 IEEE International Conference on Pervasive Computing and Communications Workshops*, March 2013, pp. 218–223.

[12] L. A. Zager and G. C. Verghese, "Graph similarity scoring and matching," *Applied Mathematics Letters*, vol. 21, no. 1, pp. 86 – 94, 2008.