Smart City Data in Urban Wellbeing Estimation

Tzirath Perez Oteiza
Hamilton Institute
Maynooth University
Maynooth, Ireland
https://orcid.org/0000-0003-1180-7689

Peter Mooney
Computer Science Department
Maynooth University
Maynooth, Ireland
https://orcid.org/0000-0002-2389-3783

Liadh Kelly
Computer Science Department
Maynooth University
Maynooth, Ireland
https://orcid.org/0000-0003-1131-5238

***Abstract*—Mental wellbeing is a global health concern. Within the urban context, it is well understood that factors associated with urban living can impact individuals' wellbeing. For example, the relationship between wellbeing and environmental factors, such as air quality, has been shown. We postulate that such factors can be automatically derived from data sources available in smart cities and used, as *contextual cues* to form automatic indicators of the wellbeing levels of individuals within a city. In this paper we provide a realistic vision for this new approach to mental wellbeing estimation for an urban population, focusing on contextual cues associated with environmental information and green infrastructure. Outcomes from initial investigations are discussed and a roadmap for the future is presented.**

***Index Terms*—Cities, smart cities, urban wellbeing, data management.**

I. Introduction

Most urban citizens are living in a smart city. Smart cities use information and communication technologies (ICT) to gather data and information continuously on how the city is operating and functioning. Smart cities usually share this information and generated knowledge with citizens. Using these rich, diverse, and dynamic data sources can provide a better quality of government service, citizen welfare, and living within the city. Smart cities strive to optimize how the city functions and promote economic prosperity. But these cities must also work to improve and enhance the quality of life for their citizens by using smart technologies, data analysis, and other computational and engineering solutions [10]. Value emerges from how smart cities use technology, analyze data and make decisions rather than simply how much technology or data is available and accessible. We see smart cities as urban areas where there is increased citizen engagement, appropriate infrastructure, social capital, and digital technologies which all combine to make our cities more livable, environmentally sustainable, resilient, and better able to respond to challenges and the needs of citizens [12]. However, not everything within smart cities can be easily monitored or computed regardless of the technologies and data available. There are challenges for smart cities in areas such as citizen mobility, environmental protection and sustainability, and balanced economic growth and development. The challenge we consider specifically in this paper is the assessment of the impact of urban life on the mental wellbeing of citizens.

In this paper, we describe ongoing work in the development of an innovative and novel approach to automatically assessing urban mental wellbeing in a given smart city context. Our proposed approach uses smart city data to derive *contextual cues* for mental wellbeing estimation. *Contextual cues* are patterns or indicators in smart city data which can be linked to established scientific research outcomes. For example, exposure to poor air quality [5] has been strongly linked to poor mental wellbeing while lack of access to infrastructures such as public transportation [18] or green spaces [7] have detrimental effects on mental wellbeing. We discuss the development of a customized data management framework to prepare and shape smart city data for the computation of contextual cues. This framework will guide the production of a harmonized dataset of contextual cue observations for a wide variety of urban domains (environment, transportation, green space, etc.). This dataset is then provided as input to an Artificial Intelligence (AI) or Machine Learning (ML) system used to generate predictions of urban mental wellbeing given the contextual cues generated from the smart city-data.

The remainder of the paper is structured as follows. In Section II we discuss contextual cues in support of our work. Section III presents our work to-date. Finally, in Section IV we conclude the paper and provide directions for future work.

II. Smart cities and contextual cues

A. Smart Cities Initiatives

The World Happiness report[1] states that cities may present many challenges that could affect the wellbeing of urban citizens with the report highlighting underdeveloped public transportation infrastructure potentially contributing to high air pollution in a city. They estimate that 90 percent of urban citizens breathe in dangerous air pollutants which results in 4.9 million yearly premature deaths. The same report highlights the importance of having accessible green space in a city since it has been shown to contribute to improving general wellbeing. Smart cities often have identified domains they deem as a priority depending on regional interests to improve

This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 18/CRT/6049. The opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Science Foundation Ireland.

[1] https://worldhappiness.report/ed/2020/cities-and-happiness-a-global-ranking-and-analysis/

quality of life [8] by creating projects and setting up initiatives through policy change or by open access data hubs [1]. Smart cities like Amsterdam and Utrecht emphasize collecting data on mobility, bicycle usage and bicycle infrastructure. IBM specializes in smart mobility, where they collect and monitor Dublin, Ireland's vehicular traffic trends using sensor technology. iCity and SmartSantander are examples of initiatives that create a centralized platform where one can access data from sensors in Europe [15]. We take one step further and take advantage of these initiatives to develop a computational system to estimate and predict urban wellbeing.

B. Contextual Cues

We adapt and use the term 'contextual cue' in our work by relating it to the input datasets and data sources available. Contextual cues, in our work, are computed quantities, markers, or flags that relate to characteristics of the input dataset(s) [13]. Many aspects of urban life have been shown to lead to increased stress in citizens. The most widely studied environmental stressors are noise, crowding, poor housing quality, poor neighborhood quality, environmental quality, and traffic congestion [4] which can be considered as contextual cues. Some of these can be computed, while others require the integration of several datasets/sources. Traffic congestion in a city can be computed by analyzing traffic flow datasets while a cue such as poor neighborhood quality could require several sources of data to be integrated. Air quality as a contextual cue requires calculation of several different air quality indicators related to public health. Therefore, it is helpful to think of a contextual cue as a computable value where research has established a linkage between this cue and citizen mental wellbeing [13]. Considerable literature exists linking various urban systems and smart city data sources to mental health and wellbeing. This paper focuses on Green Infrastructure and Air Quality as these contextual cues are reasonably data-rich and are available for most urban areas.

Air quality: In [5] researchers state that exposure to high air pollutants may cause new symptoms of depression while other pollutants may worsen existing depressive symptoms. This study also explains that high pollution rates can thus increase the risk of emergency depression visits.

Green Infrastructure: Symptoms of anxiety and depression are lower when urban citizens live near green spaces [9]. Research in [18] indicates that participants visit green spaces to reduce their stress and achieve better mental health.

C. Application

As smart cities develop and become richer in data it is feasible to implement innovative approaches to derive estimations of urban wellbeing. The major advantage of analyzing urban citizens' mental wellbeing based on multiple contextual cues is the comparatively lower cost than large-scale clinical studies which are labor-intensive, expensive, and time-consuming. While our approach does not intend to replace these clinical assessments, it does facilitate faster computation of reliable wellbeing estimation results. Compared to clinical assessments, which must survey and collect data across a selective region, our approach has the advantage of using diverse data generated by cities enabling us to replicate the process in other smart cities. This is an opportunity to use the continuous availability of new and updated data to adapt to changes within the city over time. This approach could help governments develop initiatives based on more frequently available estimations of urban wellbeing and contextual cues.

III. Our Proposed Approach

The demand for adopting data management frameworks for smart city data is growing as data is diversifying, increasing in size and becoming difficult to manage [2]. Recent work has emphasized the creation of frameworks that include data collection, processing, analysis, management, and visualization features for robust smart city applications [3]. Previous work has focused on the technical application of data frameworks [15], while in our work, we use a framework for wellness prediction. Collected datasets for each wellness related contextual cue have different dimensions resulting in the need for data harmonization techniques in these frameworks.

Our initial exploration of this problem space has focused on environmental contextual cues and specifically air quality and green space data. In this section, we introduce the data management framework that is part of the overall AI or ML computational-based software system that uses the smart city data that will compute a prediction of the current mental wellbeing. The purpose of this framework is to prepare and harmonize data to be continuously inputted into the AI computational component. The output will follow the Likert-type scoring scale, which means the better the perceived urban mental wellbeing is, the higher the score. Low scores are associated with less positive indications of good urban mental wellbeing. Section Section III-A describes the initial steps to harmonize data through data collection, data quality filtering, and data processing, as seen in Figure 1. While in section Section III-B, we focus on choosing correct labels for air quality and green space data.

A. Data Collection, Filtering, and Processing

Initially, Smart City data was collected from open-sourced websites based in Dublin, Ireland[2] that also includes Irish government-funded programs. These datasets contain data related to the environment, mobility, and infrastructure from which contextual cues can be derived. In particular, the datasets contain air pollution levels, geospatial green space data, public transportation usage, bike path length, etc. This approach is generalised and does not take into account the movement and divisions of a city due to data availability. Extensive geospatial green space data is available as open data in Manchester in contrast to limited availability in other cities. The UK for example has rich green space data publicly available. Hence, to gain greater insight into the potential of green space data as a wellness supporting contextual cue we

[2]https://smartdublin.ie/

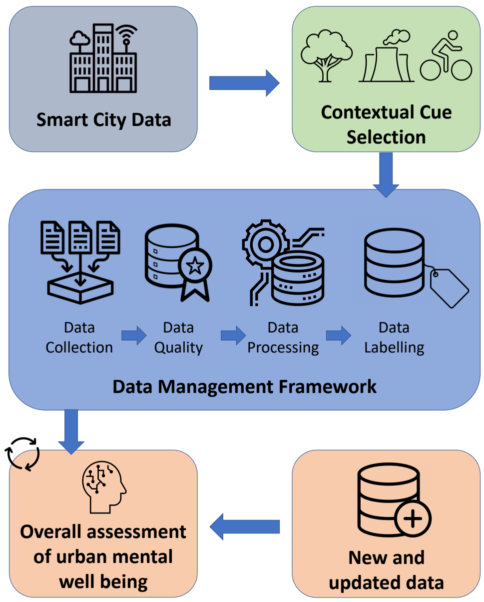

Fig. 1. Overall schematic of our proposed computational software systems

consider for our initial investigations we use green space data from the UK Ordnance Survey[3] using the OS Open Green Space dataset. For this initial stage of the integration of the data management framework, we decided to focus on using the best-case examples of data even if that means that the data acquired are from different sources. To assess the data quality for each of the available datasets they were evaluated under a criterion based on the dimensions [14]: availability, scale, coverage, aggregation, and completeness. Using these dimensions we can proceed to narrow down our initial key questions:

1) How much data is there?
2) How frequently is the data updated?
3) Is the data format usable?

These key questions will help determine what type of dataset is most suitable for this proposed method and are addressed below. Manchester air quality data and UK green space data are taken into consideration and analyzed.

Air quality: The Manchester air pollution dataset contained hourly averages of fine particle matter ($PM_{2.5}$), particulate matter (PM_{10}), and Nitrogen dioxide (NO_2) from the years 2015 to 20202. The data is provided by the Manchester city council from two main stations in the city. The completeness of the data varied by year with some years only having data from one particular site. However, when considering the three key questions mentioned before, there is enough Manchester air quality data for data analysis and training. The lack of consistency across the dimensions can be addressed with data processing techniques.

Green space: The Ordnance Survey (OS) UK Open green space dataset contains the geometric location and other associated attribute information for UK's green space locations. Manchester was chosen as the model smart city for green space. The OS Boundary-Line dataset was used to extract Manchester green spaces as well as include only green spaces that were publicly accessible, leaving out golf courses, for example. This dataset was satisfactory across all data dimensions and passed all three key questions. This data is updated every six months but it can be assumed that green space is highly static even with government efforts to build more green spaces in the next few years.

B. Data Labeling

A harmonized dataset is required for training any proposed AI or ML-based system. This harmonized data must integrate all of the chosen contextual cues with a uniform labeling classification. For example, a recorded air quality observation is labeled as having a positive, negative or neutral impact on urban mental wellbeing. In a similar way, we must also label green space accessibility or availability with an appropriate label. We need to take into consideration the various possibilities of green space interpretation: is green space based on access to public spaces or does it also include private gardens? It is important to note that there is not always broad agreement in the literature around what contextual cues relate directly to wellbeing. Furthermore, there is no agreement on what range of quantitative values contextual cues should have.

Air quality data labeling: There are various approaches when measuring the impact of pollutants on mental wellbeing. Szyszkowicz et al. [17] conducted in Toronto, researchers decided to analyze the mental health associated emergency department (ED) visits of teenagers against the levels of ambient air pollution. Results indicate an association between exposure to urban air pollutants and the number of ED visits for various mental disorders in children and young individuals [17]. Research by Chen et al., [6] provides an appropriate labeling for collected air quality data because this can be applied to air quality datasets from different sources. Chen et al. find that "a one-standard-deviation (18.04 $\mu g/m^3$) increase in average $PM_{2.5}$ concentrations in the past month increases the probability of having a score that is associated with severe mental illness by 6.67 percentage points."

Green space data labeling: Research by Labib et al. [7], used spatial approaches to determine the impact green space can have on citizens' physical and mental wellbeing in an urban area. Research in areas such as geocomputation and urban studies have often considered determine whether green space has an impact on urban wellbeing and used the calculated the distance between locations in urban areas and the access points of green spaces as a basis for this work. Labib et al. [7] shows that most approaches used the Euclidean distance approaches in these calculations using a buffer of anything between 300 and 600 meters. Here we use the

[3]https://www.ordnancesurvey.co.uk/business-government/products/open-map-greenspace

Euclidean distance approach with an average of 450 meters for buffer size as a good intermediate distance. Using the appropriate green space datasets we can use this Euclidean approach to measure the distance from urban residences to green spaces.

Labeling considerations: When applying the green space labeling system to the Manchester spatial green space map, there was a noticeable difference in green space square meters per population in a given area. In areas like the city center of Manchester, there is less green space square meters per person compared to the suburbs of Manchester. This leads us to consider a further question:

How do we allocate labels for contextual cues based on the spatial aspects of the data?

In a given city, green space may be impacted by climatic factors. Regarding air quality, air pollution levels in a city may have different impacts on citizens depending on where the citizens live. If a citizen lives in a neighborhood with a high number of factories, the impact may be quite different than if that citizen lived close to green space.

IV. Conclusions and future work

This paper proposes an innovative computational system that combines multiple techniques which utilize smart city data to assess urban mental wellbeing based on chosen contextual cues, air quality, and green space availability. This approach highlights the importance of creating a customized data management framework where smart city data is collected, filtered, processed, and labeled to create a harmonized dataset. The aim of this computational system is to use AI to predict the overall urban mental wellbeing in a given city based on selected contextual cues. The demand for urban cities to sustain the quality of life for citizens has become a priority as cities are growing. Mounton et al. [11] argue that population health and equity have been noticeably absent from much of the smart cities research and policy agenda. They urge researchers to begin to integrate the "healthy cities" work with the current interest in "smart cities" research. The description of our research is a very positive step in this direction. But as commented by Pykett et al. [16] the linkages between mental health, city life, and urban landscapes are becoming increasingly scientifically researched and these will hopefully lead to new approaches to the active promotion of urban health and wellbeing. This is where we see the potential for positive impacts of this type of work on urban and regional policy.

This work forms the early stages of a PhD program for the lead author. The next steps include (i) implementing the AI phase and assessing the accuracy or plausibility of results. Performance evaluations will be published when the system is completely implemented; and (ii) extending this approach to other case-study cities or particular regions within a city where appropriate datasets and data sources are available. In the end, the availability of smart city data will dictate which contextual cues are most relevant for a given urban area. The output of the computational systems can be used for analyses and to track the changes in a city's behaviour.

References

[1] Eiman Al Nuaimi, Hind Al Neyadi, Nader Mohamed, and Jameela Al-Jaroodi. Applications of big data to smart cities. *Journal of Internet Services and Applications*, 6(1):25, Dec 2015.

[2] Iman Avazpour, John Grundy, and Liming Zhu. V for variety: Lessons learned from complex smart cities data harmonization and integration. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pages 1–6, 2016.

[3] Muhammad Bilal, Raja Sher Afgun Usmani, Muhammad Tayyab, Abdullahi Akibu Mahmoud, Reem Mohamed Abdalla, Mohsen Marjani, Thulasyammal Ramiah Pillai, and Ibrahim Abaker Targio Hashem. Smart cities data: Framework, applications, and challenges. *Handbook of smart cities*, pages 1–29, 2020.

[4] Elena Bilotta, Uchita Vaid, and Gary W. Evans. *Environmental Stress*, chapter 4, pages 36–44. John Wiley Sons, Ltd, 2018.

[5] Massimiliano Buoli, Silvia Grassi, Alice Caldiroli, Greta Silvia Carnevali, Francesco Mucci, Simona Iodice, Laura Cantone, Laura Pergoli, and Valentina Bollati. Is there a link between air pollution and mental disorders? *Environment International*, 118:154–168, 2018.

[6] Shuai Chen, Paulina Oliva, and Peng Zhang. Air pollution and mental health: Evidence from china. Working Paper 24686, National Bureau of Economic Research, June 2018.

[7] S.M. Labib, Sarah Lindley, and Jonny J. Huck. Spatial dimensions of the influence of urban green-blue spaces on human health: A systematic review. *Environmental Research*, 180:108869, 2020.

[8] Chun Sing Lai, Youwei Jia, Zhekang Dong, Dongxiao Wang, Yingshan Tao, Qi Hong Lai, Richard T. K. Wong, Ahmed F. Zobaa, Ruiheng Wu, and Loi Lei Lai. A review of technical standards for smart cities. *Clean Technologies*, 2(3):290–310, 2020.

[9] J Maas, R A Verheij, S de Vries, P Spreeuwenberg, F G Schellevis, and P P Groenewegen. Morbidity is related to a green living environment. 2009.

[10] Filipe Moura and João de Abreu e Silva. *Smart Cities: Definitions, Evolution of the Concept and Examples of Initiatives*, pages 1–9. Springer International Publishing, Cham, 2019.

[11] Morgan Mouton, Ariel Ducey, Judith Green, Lorian Hardcastle, Steven Hoffman, Myles Leslie, and Melanie Rock. Towards 'smart cities' as 'healthy cities': health equity in a digital age. *Canadian Journal of Public Health*, 110(3):331–334, 2019.

[12] Taewoo Nam and Theresa A Pardo. Conceptualizing smart city with dimensions of technology, people, and institutions. In *Proceedings of the 12th annual international digital government research conference: digital government innovation in challenging times*, 2011.

[13] Tzirath Perez Oteiza, Liadh Kelly, and Peter Mooney. Using contextual cues in understanding urban mental well-being. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 46:85–89, 2021.

[14] Leo L. Pipino, Yang W. Lee, and Richard Y. Wang. Data quality assessment. 45(4):211–218, apr 2002.

[15] Dan Puiu, Payam Barnaghi, Ralf Tönjes, Daniel Kümper, Muhammad Intizar Ali, Alessandra Mileo, Josiane Xavier Parreira, Marten Fischer, Sefki Kolozali, Nazli Farajidavar, Feng Gao, Thorben Iggena, Thu-Le Pham, Cosmin-Septimiu Nechifor, Daniel Puschmann, and Joao Fernandes. Citypulse: Large scale data analytics framework for smart cities. *IEEE Access*, 4:1086–1108, 2016.

[16] Jessica Pykett, Tess Osborne, and Bernd Resch. From urban stress to neurourbanism: How should we research city well-being? *Annals of the American Association of Geographers*, 110(6):1936–1951, 2020.

[17] Mieczysław Szyszkowicz, Roger Zemek, Ian Colman, William Gardner, Termeh Kousha, and Marc Smith-Doiron. Air pollution and emergency department visits for mental disorders among youth. *International Journal of Environmental Research and Public Health*, 17(12), 2020.

[18] Konstantinos Tzoulas, Kalevi Korpela, Stephen Venn, Vesa Yli-Pelkonen, Aleksandra Kaźmierczak, Jari Niemela, and Philip James. Promoting ecosystem and human health in urban areas using green infrastructure: A literature review. *Landscape and Urban Planning*, 2007.