

The Role of Crowdsourcing in the Emerging Internet-Of-Things

Ramine Tinati
University of Southampton
Southampton, United Kingdom
r.tinati@soton.ac.uk

Aastha Madaan
University of Southampton
Southampton, United Kingdom
a.madaan@soton.ac.uk

Wendy Hall
University of Southampton
Southampton, United Kingdom
wh@ecs.soton.ac.uk

ABSTRACT

In this position paper we wish to propose and discuss several open research questions associated with the IoT. In particular, we wish to consider how crowdsourcing can be used as a scalable, reliable, and sustainable approach to support various computationally difficult and ambiguous tasks recognised in IoT research. We illustrate our work by examining a number of use cases related to healthcare and smart cities, and finally consider the future development of the IoT ecosystem with respect to the socio-technical philosophy and implementation of the Web Observatory.

Keywords

Internet of Things, IoT, crowdsourcing, Web Observatories

1. INTRODUCTION

The Internet-of-Things is quickly becoming one of the fastest growing areas of interest due to the vast spectrum of devices available to both the research community, and domestic market. In this position paper, we explore some of the challenges faced when designing and deploying IoT platforms, and explore the role of crowdsourcing as a socio-technical approach to address them.

One of the main barriers recognised across many IoT-supported infrastructures is the ability to accurately identify a device's footprint; that being, the type of device, the manufacturers, and ultimately, what data they are producing. As there currently exists no agreed standard for device metadata, much of this work becomes a manual process, and assumes an agreed level of trust between the provider and consumer, similarly, the process of data integration often requires hard coding, and automatic approaches which require human intervention. As a result, both device identification and data integration becomes difficult at scale, which in many scenarios, is where the true potential of the IoT eco-system resides.

In addition to the challenges around the identification and integration of devices and their data, management and le-

gitimate access to them is also of concern. Many of these devices will provide high-resolution data, which may be highly sensitive in the content of the data being collected and transmitted, and this level of sensitivity is dramatically increased with the analytical opportunities that data integration can provide. Consequently, the management, access, and control over require serious consideration.

In this position paper we wish to propose and discuss several open research questions associated with the IoT. In particular, we wish to consider how crowdsourcing can be used as a scalable, reliable, and sustainable approach to support various computationally difficult and ambiguous tasks recognised in IoT research. We illustrate our work by examining a number of use cases related to healthcare and smart cities, and finally consider the future development of the IoT ecosystem with respect to the socio-technical philosophy and implementation of the Web Observatory.

2. CROWDSOURCING AND IOT

There are several open research questions within the field of IoT, including topics such as device detection, data integration, schema alignment, and access control and data management [4]. Existing research has show the heterogeneity of IoT devices, as well as the data which they produce, this is further exacerbated by the lack of metadata (and standards for metadata schema) associated with the devices [7]. Therefore, using computational/automated methods for identify the type of devices, the nature of the data being produced, and its compatibility for data integration tasks, is often challenging. However, more recently, there has been a growing interest towards the role of of crowdsourcing as an approach to improve the current research challenges in IoT [15, 10].

Crowdsourcing, by definition, is the use of humans (at scale) to complete computationally difficult or time consuming tasks [6]. Traditionally, this involves human participants completing simple, short timeframe classification exercises in order to validate and verify device and data related questions.

Existing research outside IoT related-research has shown crowdsourcing, and in particular, citizen science approaches to crowdsourcing, can form reliable, scalable, and sustainable solutions to supporting the problem of annotating large complex datasets[14, 11]. Traditionally, crowdsourcing for citizen science has been used to help annotate scientific datasets, such as a collection of Hubble Telescope images[3]. Users are asked a series of simple questions (e.g. is there an object in

©2017 International World Wide Web Conference Committee (IW3C2), published under Creative Commons CC BY 4.0 License.
WWW'17 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
dx.doi.org/10.1145/3041021.3051693



the image), which then verification algorithms are applied to determine the most statistically valid answer.

One of the biggest difficulties with crowdsourcing is developing incentive mechanisms to recruit and sustain an active community of users[12]. It is important for the community to explore and work through different scenarios where users will engage based on a mutual value exchange (non-monetary). In order to achieve this, research will be conducted to investigate the motivations of participants, and the necessary extrinsic and intrinsic rewards suitable for sustained recruitment.

Adopting the crowd-based citizen science workflow in IoT platforms, it may be possible to exploit the use of citizen science techniques in order to help improve the identification of devices and data sources. In the simplest of use cases, participants will be asked a series of questions related to the device (e.g. “is the device a thermostat”) and about the data (e.g. “does the device have a timestamp field”). These answers will then be used to improve the current Machine Learning models for automated device detection and data integration. However, as of current, this is still a new research area, requiring extensive experiments and studies in order to demonstrate the capabilities of this approach [16].

Another research challenge in drawing on crowdsourcing as a means for improving accuracy is investigating the spectrum of socio-technical platforms used to perform citizen science activities. Traditionally, citizen science platforms (e.g. Zooniverse [11], ESPGame [14]) have been Web-based, requiring users to navigate their way to a given platform and perform tasks in solo modes of operation. However, more recently, ‘reverse citizen science’ has been attempted; citizens themselves produce data using their own devices (e.g. taking pictures of the night sky using a mobile application). Data produced using this approach becomes stored centrally, and if designed appropriately, other participants of the system may be able to validate the collected data [8]. Thus not only is the collection of data crowdsourced, but so does the process of validation and verification. In light of these new approaches, we argue that there needs to be significant effort in how to engineer similar environments for engaging participants in IoT Citizen Science.

3. USE CASES

In this section we wish to consider several use cases where crowdsourcing could plan an important role in the overall architecture and work-flows of the Internet-of-Things infrastructure.

3.1 Healthcare Data Integration

Hundreds of thousands of medical devices such as patient monitors, infusion pumps, ventilators, and imaging modalities – many of which are life-sustaining or life-supporting – currently reside on hospital networks across the United States. Even more medical devices are accessible via wireless technologies, for example, insulin pumps and pacemakers.

Diabetes a lifestyle disease which is increasingly becoming common in the UK with almost 2.9 million people were diagnosed with diabetes in 2013. With 1 in 20 people estimated to have diagnosed or undiagnosed diabetes self-management is critical including lifestyle changes, complexities and possible side-effects of therapy, and patient education [2]. Diabetes digital coach is an IoT enabled test-bed to support healthcare commissioners, hospitals and community providers

to work with self-management products and evaluate latest developments in connecting monitoring devices. In addition to timely interventions from peers, healthcare professionals, carers and social networks, the testbed aims to enable the individuals to can gain comprehensive, real-time view of their own data to formulate self-management strategies based on hidden patterns, trends and relationships that are not considered through conventional treatment options. These individuals can now share this information and knowledge with relevant healthcare professionals for support, advice and care plan. Further data, information and knowledge from a variety of sources can be aggregated to gain a real-time and population-wide view of the health status of people and promote behaviours to improve health.

3.2 Collaborative Smart City Initiative

For building and sustaining smart cities in a democratic (bottom-up) manner, citizens need to be active participants in policy making, problem solving and not just data providers (e.g. crowdsensing [5]). Crowdsourcing can support integration of data from different services, sensors deployed by city management and citizens. It can also play a significant role towards rapid problem solving where citizens can easily report problems they observe or face and city management has appropriate and simple instruments to ask for help from citizens to solve the problem. For example, the CityVerve demonstrator aims to convert ‘flag and pole’ bus stops into safe places with location-based services, sensors and beacons, mobile apps and intelligent digital signage. People will then be able check-in to their bus stop and let bus operators know they are waiting for their service [1]. A similar application is envisaged for improving local healthcare services through ‘biometric sensor network’. Real-time sensors will be able to report on the current state of well-being of individuals who are using specific sensors. One could envision a ‘smart city’ which may contain various geographically-located ‘hotspots’; citizens can gather in order to share access or upload their data, which is particularly useful in rural areas where Internet and network access is limited.

4. RESEARCH CHALLENGES FOR CROWD-SOURCING

In this position paper, we argued for the role of crowdsourcing in the emerging infrastructures of Internet-Of-Things. By examining the role of existing crowdsourcing approaches and their suitability for supporting computationally-challenging and time consuming tasks within IoT systems, we presented a number of research challenges which could be addressed.

1. **Semantic Interoperability:** IoT is an industry driven technology where every IoT vendor produces its own IoT platform. Moreover, most of the IoT solutions are case-centric and result in creation of “IoT silos” which require “inter-silo” interoperability for sharing data. Any protocol or standard needs to consider devices, their context-of-use and data emitted by these devices. The challenge in IoT domain is that a variety of ontologies dealing with various aspects of sensors and sensing (different scope, granularity and generality) have been proposed. This makes integration of a formal ontology with an implicit one reflected in a database schema or in a communication protocol specification or in a design document complex [9]. Appli-

cation of crowdsourcing and incentive engineering can support enrichment of metadata for data interoperability and data sharing purposes in different IoT enabled domains such as, air pollution monitoring, health status monitoring among others.

2. **Data Sharing and Access Control:** Another major research challenge in IoT systems is user privacy and data protection especially with respect to privacy associated with data collection, sharing, and management. Identification and management of billions of devices associated with each other, maintenance of trust between device interactions, and the human identification of devices raise a critical concern of **authorization**. This can determine the credibility and reputation of a person or object, which ultimately leads to access (or future granted access) to a resource. However, it is both a policy and technical challenge to assess the risk associated with sharing information and trust on a requester. Crowdsourcing methods can support understanding of stakeholders privacy concerns and their mental models for information sharing through micro-tasks which can be analysed as inputs for access and data sharing policies in IoT systems.
3. **Democratic Policy Making:** Specialised scientific domains such as healthcare and governance context such as smart-cities are on one-hand becoming largest consumers for IoT devices and on the other hand are increasingly becoming democratic in nature where individuals act not just as data providers but also participate actively in solving problems, sharing solutions and formulating policies. In such scenarios, modelling these stakeholders as part of the platforms deployed in these domains is critical and faces a number of a big challenge of “how”. Incentivisation of tasks within self-management of chronic illnesses and for issues arising in a city and the network of smart devices is imperative.

5. TOWARDS AN IOT OBSERVATORY

Finally, we wish to consider the role that Web Observatories will play in the future of IoT development, with a particular focus on how an active community can contribute to, and benefit from, the Web Observatories distributed architecture for data access, sharing, and querying and crowdsourcing.

To actively engage communities of various stakeholders from different IoT application domains for sharing their resources and participate in various stages of the IoT data processing pipeline we envision the “**IoT Observatory**”. As IoT is considered as an extension of the Web, an IoT observatory can be considered as an extension of the Web Observatory proposed in [13]. The IoT Observatory will comprise of a distributed network of observatory nodes through which a number of devices, stakeholders participate for sharing data analysis, integration and correlation from different data streams coming from a heterogeneous set of devices. In addition to these, different communities can engage with relevant crowdsourcing tasks through the observatory interface. The IoT observatory will support the challenges mentioned in the previous section, including data interoperability through meta-data integration. It will also support real-time and historical data analysis and enable task organizers

and participants to engage with the provenance of similar tasks or activities supporting a life-long learning system.

Acknowledgements

This work is supported under SOCIAM: The Theory and Practice of Social Machines, funded by the UK EPSRC under grant EP/J017728/2, and also the PETRAS IoT Research Hub - Cybersecurity of the Internet of Things grant EP/N02334X/1.

6. REFERENCES

- [1] Cityverve. Open-Access Web Report, Last accessed: 22 January 2017.
- [2] Nhs england testbeds. Open-Access Web Report, Last accessed: 22 January 2017.
- [3] C. Christian, C. Lintott, A. Smith, L. Fortson, and S. Bamford. Citizen science: Contributions to astronomy research. *arXiv preprint arXiv:1202.2577*, 2012.
- [4] L. Da Xu, W. He, and S. Li. Internet of things in industries: A survey. *IEEE Transactions on industrial informatics*, 10(4):2233–2243, 2014.
- [5] R. K. Ganti, F. Ye, and H. Lei. Mobile crowdsensing: current state and future challenges. *IEEE Communications Magazine*, 49(11), 2011.
- [6] J. Howe. The rise of crowdsourcing. *Wired magazine*, 14(6):1–4, 2006.
- [7] S. Li, L. Da Xu, and S. Zhao. The internet of things: a survey. *Information Systems Frontiers*, 17(2):243–259, 2015.
- [8] M. Luczak-Rösch, R. Tinati, S. Aljaloud, W. Hall, and N. Shadbolt. A universal socio-technical computing machine. In *Web Engineering - 16th International Conference, ICWE 2016, Lugano, Switzerland, June 6-9, 2016. Proceedings*, pages 559–562, 2016.
- [9] D. Raggett. Building the web of things: Standards for semantic interoperability and end to end security. September 2016.
- [10] A. Rauniyar, P. Engelstad, B. Feng, et al. Crowdsourcing-based disaster management using fog computing in internet of things paradigm. In *Collaboration and Internet Computing (CIC), 2016 IEEE 2nd International Conference on*, pages 490–494. IEEE, 2016.
- [11] R. Simpson, K. R. Page, and D. De Roure. Zooniverse: observing the world’s largest citizen science platform. In *Proceedings of the 23rd international conference on world wide web*, pages 1049–1054. ACM, 2014.
- [12] R. Tinati, M. Van Kleek, E. Simperl, M. Luczak-Rösch, R. Simpson, and N. Shadbolt. Designing for citizen data analysis: a cross-sectional case study of a multi-domain citizen science platform. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 4069–4078. ACM, 2015.
- [13] T. Tiropanis, W. Hall, J. Hendler, and C. de Larrinaga. The web observatory: a middle layer for broad data. *Big Data*, 2(3):129–133, 2014.
- [14] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*,

CHI '04, pages 319–326, New York, NY, USA, 2004. ACM.

- [15] S. Ziegler, S. Nikoletsea, S. Krco, J. Rolim, and J. Fernandes. Internet of things and crowd sourcing-a paradigm change for the research on the internet of things. In *Internet of Things (WF-IoT), 2015 IEEE 2nd World Forum on*, pages 395–399. IEEE, 2015.
- [16] S. Ziegler, J. Rolim, and S. Nikoletsea. Internet of things, crowdsourcing and systemic risk management for smart cities and nations: Initial insight from iot lab european research project. In *Advanced Information Networking and Applications Workshops (WAINA), 2016 30th International Conference on*, pages 611–616. IEEE, 2016.