Revisiting Activity Theory within the Internet of Things

Adriana Wilde and Ed Zaluska

Web and Internet Science (WAIS) research group,
Electronics and Computer Science (ECS),
University of Southampton, SO17 1BJ, United Kingdom.
{agw106,ejz}@ecs.soton.ac.uk

Abstract. With the emergence of the Internet of Things, interactions between humans and machines and indeed amongst machines themselves can be better understood using Leontiev's activity theory. This theory has been relevant to Human-Computer Interaction research for some time, but this paper revisits the underlying concepts with a particular emphasis on the Internet of Things. Newer approaches may be more appropriate to represent complex activities within their wider context, as opposed to the traditional (and limited) view of mediated activities at an individual level.

Key words: Activity Theory, Internet of Things, context of an activity.

1 Introduction

The Internet of Things (IoT) is the concept of networking physical objects into a "global infrastructure", "a loosely coupled, decentralised system of *smart objects*" [1], which can thus realise Weiser's vision of ubiquitous computing [2]. This vision was of a world in which people could be supported in their daily lives in a non-intrusive way. However, in order to achieve such unobtrusiveness, one of the following two premises should be satisfied. Either objects/systems become able to interact with each other without human intervention, or human activity must be accurately detected (without users having to define it explicitly).

On the one hand, activity recognition is a challenge worth undertaking to improve the general user experience, but it is even more relevant whenever users find interaction difficult or impossible, for example, but not exclusively, in the case of elderly people, infants or chronically ill patients. Accurate and reliable activity detection has the potential for improving the quality of care for these groups and, consequently, their quality of life. However, this is not to dismiss potential significant gains for other applications such as the managing of urban waste, industrial productivity, urban planning, home automation and energy savings, are all areas for which "smart" solutions can be applied (i.e. solutions involving automatic monitoring of activities using sensors).

On the other hand, smaller, less obtrusive devices are increasingly becoming available at lower cost. Technology capable of collecting activity data now ranges from wearable devices to complex intelligent environments. The choice of technology may depend on issues such as technical, economic and social constraints. The precise mix of data to be collected will vary with the application. Data collection is an important phase of activity detection, followed by a classification phase in order to discriminate one activity from another. In this paper we consider data collection for activity detection in the context of the IoT as taking place via sensors embedded in everyday objects. In Section 2 the notion of human activity is discussed, as well as its relevance to the IoT, and Section 3 offers an overview of common techniques for collecting sensor data for activity recognition, For the concluding remarks we revisit activity recognition as a whole and consider the need for a new approach in Section 4.

2 Activities

Most of the research concerning human activity recognition consider only physical activities. These are defined by Preece *et al.* as "any bodily movement produced by skeletal muscles that results in energy expenditure above resting level" [3]. This definition is somewhat simplistic, because although bodily movements are relatively easy to measure and identify in laboratory conditions, the practical applicability of this process in analysing real-life situations is limited. This is because actions always take place inside a context, and they are often impossible to understand without that context [4].

Inspired by Leontiev's activity theory[1], Kuutti offers a far more general definition of activity: "a form of doing directed to an object" [4]. Transforming an object is the purpose of the activity and thus defines it. Activities are then distinguished from each other according to the outcomes of the transformations to objects. An *object* here could be physical, but could also be less tangible (such as plans and ideals). However, they must be shared for manipulation and *transformation* by the those taking part in the activity, be it as an individual *subject* or as a *community*. Figure 1 shows three mutual relationships that are formed between these entities whenever an activity is taking place.

As mediators, there are *tools* used in the transformation process (including material and abstract tools), *rules* (norms and conventions within the community) and *division of labour* (organisation of the community around the transformation process) [4].

In addition, an activity can be conceived as a systemic whole in which all elements are interrelated (which can be very complex to represent). In fact, activities require contextual information, emanating perhaps from related activities (by time, space or resources), and themselves can represent contextual information to other activities. If the context is stripped out of activities so as to isolate them, and the complex observable phenomena are reduced to idealised simpler mathematical descriptions (using a "positivist" approach [6]), the result

[1] A.N. Leontiev was a Soviet psychologist from Vygotsky's school. For Leontiev, "activity is not a reaction or aggregate of reactions, but a system with its own structure, its own internal transformations, and its own development" [5]

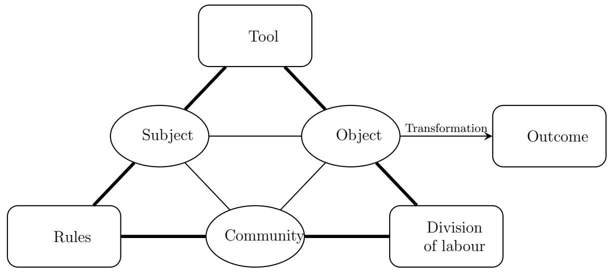

Fig. 1. Basic activity structure (adapted from [4]).

is a systemic model which applies at the individual level, such as the one shown in Figure 2.

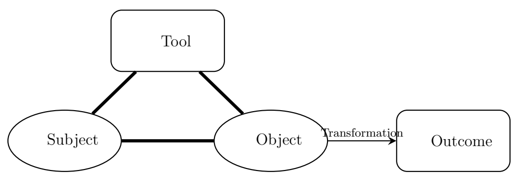

Fig. 2. Mediated relationship at the individual level (adapted from [4]).

This view individualises the notion of an activity, and this is the exact scope of concern within most of the literature. To illustrate the limitations of this view, consider a system that is able to detect that some users are running, others are standing, others walking, and so on, yet is unable to identify that the activity taking place is a football match. As a community, all users perform certain actions/operations that define an overarching activity: a common goal is being pursued, subject to how the labour was divided and other restrictions given by the rules of the game and other contextual cues that the system is not considering. Such a system focuses only on the individual level, as in Figure 2, failing to consider the role of the community into the transformation process as represented in Figure 1.

To emphasize this point, Kuutti notes that the real world experience always involve an intertwined network of activities [4]. Activities are hence seen over longer periods, as objects are transformed into outcomes through a process that can be deconstructed as a sequence of *actions*, which in turn consists of *operations*. This is a crucial distinction, commonly blurred in the literature. As an example, Loke defines an activity as typically referring to some action or operation, undertaken by a person [7]. An activity is then considered as contextual information which can be used to characterise someone's situation, so that the activity becomes "the minimal meaningful context for understanding individual actions".

Furthermore, stripping the context out is an undesired oversimplification of the system, since a given action might belong to different activities, in which case the contextual information would provide the only means to ascertain to which activity it belongs. For example, consider the action of "sitting", as part of the activities "driving" and "watching television". Detecting the posture is not sufficient for the differentiation between these activities. Only if the context is captured too, could these activities be discriminated accurately. The next section deals briefly with the technical aspects of such a process: the data collection (via sensors) and the recognition algorithms.

3 Sensing and recognising

Activity recognition is the process in which human behaviour is monitored and processed to infer the underlying activities [8]. *Sensors* are essential for data collection, as they detect stimuli by generating measurable signals [9]. Typically, sensors fit into the activity model presented in Section 2, as a mechanism to measure the transformation process. However, within the IoT framework, they could be considered instead as measuring the interaction between entities directly (e.g. an object being manipulated by a subject). In what follows, we look at the use of sensors for activity recognition purposes.

A great variety of different sensors can be embedded in a device (e.g. cameras, accelerometers, rate gyros, light sensors, microphones, location sensors, barometers, passive infra-red sensors, magnetic field sensors, metal ball switches, mercury switches and solar panels [10]). Despite this variety, the vast majority of activity classification systems have used inertial sensors, notably accelerometers, because they directly record motion data [3].

The wearable sensor approach is an effective and inexpensive way to recognise certain types of activities, e.g. physical movements [8]. However, recognition accuracy varies with the number of sensors and their placement on one or several locations on the body [11–13]. With regards to the placement as a variable, Bao and Intille [11] showed that placing biaxial accelerometers in as few as two locations affected accuracy only marginally (less than 5%) when compared to a five-accelerometers arrangement. The extra cost of providing extra sensors did not provide a significant improvement in the data. Furthermore, Ravi *et al.* [13] uses data from only one triaxial accelerometer (on the hip), and so did Wilde [14], achieving accurate recognition rates. In the latter study, accurate activity recognition was possible even for activities which intuition dictated would be difficult to discriminate with a single sensor on the hip, such as brushing teeth and writing on the blackboard. However, Wilde reported less accurate discrimination between hiking uphill and downhill [14].

Lester *et al.* [12] used a variety of sensing modalities to minimise the information loss from using a single device. In other words, when the sensors are in a single placement (as they would be in a smartphone), it is useful to utilise multimodal information to record more contextual cues in the environment. More recent works use a single accelerometer [14, 15] yet are able to achive high recognition rates despite of the simplicity of the system. Ravi *et al.* suggest that "short activities", such as arm gestures, might be recognised using accelerometer data with greater difficulty than general activities of cyclical or repetitive nature which span over longer periods [13]. The length of the activity is highly sensitive to the sampling rate, since with a low rate short activities can be missed altogether. Brezmes *et al.* perform activity recognition with a sampling rate of up to 30 samples per second [15], but accelerometers are increasingly capable of fast sampling rates, to frequencies in the region of 160Hz and higher, allowing for the creation of vast datasets for activity recognition. However, to put in perspective how much the data collected can grow when considering multimodal sensors, Lester *et al.* sampled at only 4Hz, yet the amount of data collected was over 18,000 samples every second, requiring preprocessing of the data to reduce the high dimensionality before it could be input to a recognition algorithm. Once activity data has been collected, it can then be processed, using techniques based on either statistical modelling, or logical reasoning [8]. Most of the literature to date uses the first approach, in which activity recognition is commonly achieved by identifying patterns in the raw data, leading to its classification.

4 Conclusions

As we have seen, activity recognition using body-mounted sensors has been extensively used. Despite this approach being broadly recognised as effective, it presents a number of problems. The majority of the contributions surveyed are limited to the study of physical activities, and furthermore, those defined using activity theory are mere actions or operations. We prefer to consider human activities (as well as those performed by Internet-enabled objects) to be part of a richer context. These activities happen concurrently and overlapping, with actions contributing to possibly several activities at the same time, and perhaps being performed by a community (or indeed more than one actor, as opposed to the current approaches). A model of activities suitable for the IoT is not sustainable if it only models mediated relationships at the individual level, as shown in Section 2. The community element, as well as the rules of play and how the tasks are subdivided amongst that "community" (be it of people or devices) has to be taken into account to be faithful to the emergent complexity.

More importantly, many real-world activities involve complex motions and complex interactions with the environment and its context, as pointed out in Section 2. Sensor readings alone may not be able to discriminate activities involving simple physical actions (e.g. making tea and making coffee), and thus another approach, object-based activity recognition, becomes necessary in such cases. The IoT promise is thus that the technical difficulties of this approach will be overcome with continued development, since the required complexity can only be captured through the use of a multitude of simpler systems functioning simultaneously and distributed throughout the environment of interest. Building e-Infrastructures allowing for such a development is a challenge, which can be undertaken with the lessons being learned now in Web and Internet sciences.

References

1. G. Kortuem, F. Kawsar, D. Fitton, and V. Sundramoorthy, “Smart objects as building blocks for the internet of things,” *Internet Computing, IEEE*, vol. 14, no. 1, pp. 44–51, 2010.
2. M. Weiser, “The Computer for the 21st Century,” *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 3, no. 3, pp. 3–11, 1999.
3. S. Preece, J. Goulermas, L. Kenney, D. Howard, K. Meijer, and R. Crompton, “Activity identification using body-mounted sensors - a review of classification techniques,” *Physiological Measurement*, vol. 30, pp. R1–R33, 2009.
4. K. Kuutti, “Activity Theory as a potential framework for human-computer interaction research framework,” in *Context and Consciousness: Activity Theory and Human Computer Interaction* (B. A. Nardi, ed.), vol. 1/1, ch. 2, pp. 17–44, 1995.
5. A. N. Leontiev, “The problem of activity in psychology,” *Journal of Russian and East European Psychology*, vol. 13, no. 2, pp. 4–33, 1974.
6. P. Dourish, “What we talk about when we talk about context,” *Personal and Ubiquitous Computing*, vol. 8, no. 1, pp. 19–30, 2004.
7. S. W. Loke, “Representing and reasoning with situations for context-aware pervasive computing: a logic programming perspective,” *The Knowledge Engineering Review*, vol. 19, no. 03, pp. 213–233, 2005.
8. L. Chen and C. Nugent, “Ontology-based activity recognition in intelligent pervasive environments,” *Int. J. Web Information Systems*, vol. 5, no. 4, pp. 410–430, 2009.
9. G. Francis, C. E. Rash, and M. B. Russo, “The human-machine interface challenge,” in *Helmet-Mounted Displays: Sensory, Perceptual and Cognitive Issues* (M. R. T. L. C.E. Rash and E. Schmeisser, eds.), ch. 2, pp. 29–44, 2009.
10. A. Schmidt and K. van Laerhoven, “How to build smart appliances?,” *IEEE Personal Communications*, vol. 8, no. 4, pp. 66–71, 2001.
11. L. Bao and S. S. Intille, “Activity recognition from user-annotated acceleration data,” *LNCS*, vol. 3001, pp. 1–17, April 2004.
12. J. Lester, T. Choudhury, and G. Borriello, “A practical approach to recognizing physical activities,” *LNCS*, vol. 3968, pp. 1–16, 2006.
13. N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, “Activity recognition from accelerometer data,” in *Proc. of the 17th Conference on Innovative Applications of Artificial Intelligence(IAAI)*, vol. 20, pp. 1541–1546, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press, AAAI Press, 2005.
14. A. Wilde, “Activity recognition for motion-aware systems,” master thesis, University of Fribourg (Switzerland), 2011.
15. T. Brezmes, J. Gorricho, and J. Cotrina, “Activity Recognition from Accelerometer Data on a Mobile Phone,” in *Proc. of the 10th Int. Work-Conference on Artificial Neural Networks: Part II: Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living*, p. 799, Springer, 2009.