Department of
Information Engineering
and Computer Science
DISI

UNIVERSITY
OF TRENTO - Italy

DISI - Via Sommarive 14 - 38123 Povo - Trento (Italy)
http://www.disi.unitn.it

Assessing Annotation Consistency in the Wild

Fausto Giunchiglia, Mattia Zeni, Enrico Bignotti and Wanyi Zhang

2018

Technical Report # DISI-18-015

Proceedings in: Pervasive Computing and Communications Workshops (PerCom Workshops), 2018 IEEE International Conference

Assessing Annotation Consistency in the Wild

Fausto Giunchiglia*, Mattia Zeni†, Enrico Bignotti‡ and Wanyi Zhang§
Department of Information Engineering and Computer Science - University of Trento
Via Sommarive, 9
I-38123 Povo (TN)
Email: *fausto.giunchiglia@unitn.it, †mattia.zeni.1@unitn.it, ‡enrico.bignotti@unitn.it, §wanyi.zhang@unitn.it

***Abstract*—The process of human annotation of sensor data is at the base of research areas such as participatory sensing and mobile crowdsensing. While much research has been devoted to assessing the quality of sensor data, the same cannot be said about annotations, which are fundamental to obtain a clear understanding of users experience. We present an evaluation of an interdisciplinary annotation methodology allowing users to continuously annotate their everyday life. The evaluation is done on a dataset from a project focused on the behaviour of students and how this impacts on their academic performance. We focus on those annotations concerning locations and movements of students, and we evaluate the annotations quality by checking their consistency. Results show that students are highly consistent with respect to the random baseline, and that these results can be improved by exploiting the semantics of annotations.**

***Index Terms*—User-generated content, Annotations, Smartphone sensing, Behavioral analysis, Crowdsensing**

I. Introduction

Understanding human behavior in real life scenarios is becoming an increasingly investigated issue. To do so, smart devices such as smartphones, thanks to their sensing capabilities and their pervasiveness in our everyday lives, are widely adopted to understand user behavior automatically. Smartphone pervasiveness also explains the rise in popularity of personal assistant applications such as Google Assistant or Apple's Siri. However, in order to move towards an improved personalization of applications and services, an even better understanding of human context is required. Acquiring this type of knowledge cannot be done by relying on sensor data alone — it requires involving humans and human annotations.

A way to add human knowledge to real-world sensor information is to have users annotate their everyday behavior. Users' annotations are at the base of research areas such as participatory sensing [16] and mobile crowdsensing [13], where mobile users actively participate in collecting data and, in some recent work, annotate their data as well [4].

A relevant issue in these areas is the quality of the data provided by users [20], [18]. However, the majority of the work focuses on the quality of sensor data, e.g., lack of sensor calibration, environmental noise, and redundancy of data [19], while there is still a lack of a systematic approach concerning user annotations of their data [14]. Ensuring quality in annotations is vital to obtain a clear understanding of users experience without resorting to time-consuming manual validation by experts, which only works within controlled environments [17], [15]. In fact, experts are unlikely to be able to exactly mirror the experience of users, except in very simple tasks, e.g., traveling [4].

In this work, we propose the evaluation of the annotation methodology developed in [10], where annotations are built semantically to account for users' understanding of their surroundings and can be deployed in real life scenarios. Thus, users can continuously annotate their experience during their everyday life, providing a detailed and personalized account of their behavior.

The annotation process was first performed during the SmartUnitn project, which aims at recognizing behavioral patterns of students to see how their everyday lifestyle affects their academic performance. During the study, students annotated their everyday life via a dedicated application on their smartphones that collected sensor data at the same time.

The evaluation of the annotation process is performed on the annotated dataset from SmartUnitn, where we focus on the annotations done by students when describing their locations and traveling habits. The evaluation consists in checking the consistency of the annotations, formalized as the pair <label, location coordinates> as collected from users, where the higher the number of different unique annotations referring to the same location coordinates, the less consistent the annotation task. Furthermore, we propose to exploit the ontological information used to build the labels to group semantically close labels to improve the results. Results show that our users are consistent, even more so if considering the semantics provided by ontologies.

The remainder of this paper is organized as follows. Section II provides an overview of our methodology for building annotations, while Section III presents SmartUnitn and the dataset. Then, Section IV explains the notion of consistency and shows our results concerning the quality of the students' spatial annotations. Section V describes works similar to ours and Section VI concludes the paper.

II. The Annotation Methodology

The building process for our annotations is based on the notion of context, i.e., "a theory of the world which encodes an individual's subjective perspective about it" [9]. In our approach, context is the mechanism that people use to make sense of their surroundings to decide what is relevant to their current state of affairs [3]. For instance, a student focus will be on specific elements of her environment, e.g., a lesson and the class where it is taking place, which means that others may not be considered relevant, e.g., the teacher gesturing or a classmates' smartphone lying in the next desk.

To account for the structure of context, we model it as a tuple:

$$Cxt = < me, WA, WE, WO, WI > \quad (1)$$

where:

- **me** is the person on which the context is centered, e.g., a student;
- WA is the Temporal component, i.e., the dimension that answers the question "**W**h**A**t are you doing?".
  It covers all the relevant activities for a person in the current context, e.g., attending a lesson;
- WE is the Spatial component, i.e., the dimension that answers the question "**W**h**E**re are you?".
  It covers all the relevant locations for a person in the current context, e.g., a classroom;
- WO is the Social component, i.e., the dimension that answers the question "**W**h**O** are you with?".
  It covers all the relevant people for a person in the current context, e.g., the teacher and classmates
- WI is the Object component, i.e., the dimension that answers the question "**W**hat are you w**I**th?".
  It covers all the relevant objects for a person in the current context, e.g., his or her smartphone

We model each dimension as an ontology based on the general ontology in [11] unifying human perception and knowledge representation. Figure 1 shows an example of an ontology specifying the WE component, i.e., locations; notice that this ontology can become arbitrarily complex and general since this example is not linked to any particular standard.

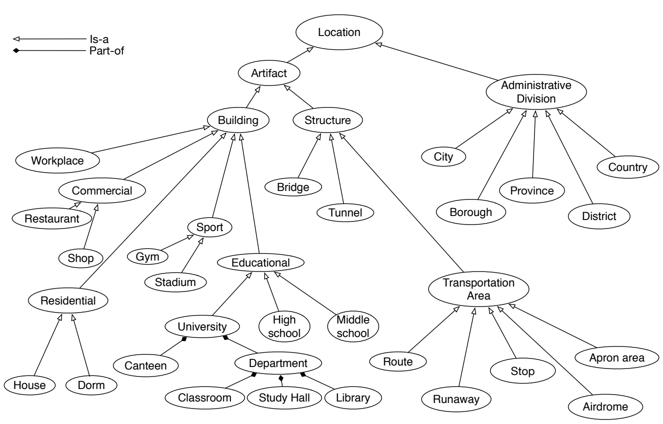

Fig. 1: An example of an ontology representing WE, i.e., locations.

Ontologies can act as a hierarchy of labels to be used as annotations. However, they must adapt to be deployable on smartphones and usable by people in real life. Thus, we present them to the users as time diaries, which are widely used in sociology to analyze human behavior and consist in logs where respondents report activities performed, locations visited and people encountered during their day [23].

This presentation of ontological information in the form of time diaries consists of two main steps. Firstly, the appropriate time use classification standard must be agreed upon, considering its adaptability and coverage of the area to be explored. Secondly, the context dimensions to be covered must be chosen, i.e., locations, activities, and people. Then, the resulting ontologies must be adapted, with the help of sociologists, based on the research scope and aim to become the coded entries of time diaries to be deployed on smartphones.

The time diary used in this work was presented in [10], so we will not go into too much detail on the building process. To summarize, it relied on ATUS, given its potential for activity recognition [2], to obtain an ontology consisting in over 80 candidate labels for three dimensions, each being a question to be asked: activities, locations, and people. Then, the second step led sociologists to reduce the total number of labels to 43, including "Other", which is a standard option in time use survey to represent that any activity not previously listed is outside the research scope [5].

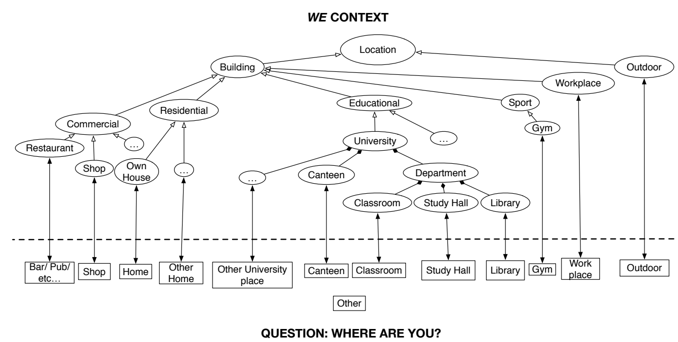

Fig. 2: The mapping from WE ontology to the locations labels.

As an example, since it will be relevant in Section IV, Figure 2 shows the mapping from the location dimension, i.e., the *WE* component, to the question about locations. Notice how the mapping is almost one to one with the lowest tiers, except for "Other University place" and "Other Home", since they group more specific types of buildings. Notice how the majority of the granularity goes towards academic buildings since the major focus of is the students' life.

TABLE I: The time diary used in [10].

What are you doing?	Where are you?	Who is with you?
Lesson	Class	Alone
Study	Study Hall	Classmate(s)
Eating	Library	Friend(s)
Personal Care	Other University place	Roomate(s)
En route (*)	Canteen	Partner(s)
Social life	Bar/ Pub/etc	Colleague(s)
Social media & internet	Relative(s)	Other
Cultural Activity	Home	
Sport	Other Home	**(*) How are you travelling?**
Shopping	Workplace	By Foot
Hobbies	Outdoors	By Bus
Other Free Time	Gym	By Train
Work	Shop	By Car
Housework	Other Place	By Bike
Volunteering		Other
Other		

This process of adaptation was then carried out also for the other dimensions investigated, i.e., activities and social roles, resulting in a complete time diary, as shown in Table I. Each dimension is mirrored by a list of possible closed answers, where each question refers to the corresponding context component. In fact "What are you doing?" accounts for activities (WA), "Where are you?" accounts for locations (WE), and "Who is with you?" accounts for social relations (WO); notice that no WI is present in this specific case. The link between the fourth question "How are you traveling?" and the "En route" activity, shown via an asterisk, represents that, although "En route" would qualify as an activity, it refers to traveling habits, which refer to locations. If a user selects this option, instead of the "Where are you?" options, a list of possible means of transportation is provided: "By Foot", "By Bus", "By Train", "By Car", "By Motorbike", and "By Bike".

III. SmartUnitn

We validate our proposed solution on the data from the SmartUnitn project, which belongs to a family of projects called ƧMAЯTRAMS[1] that leverages on smartphones to extract behavioral patterns from people and develop systems that assist users in their everyday life. The main aim of SmartUnitn is to fill the empirical gap concerning students' time allocation and academic performance by providing a detailed description of how their time management affects their academic achievement.

SmartUnitn relies on the dedicated mobile application called i-Log [24] for two functions.

1) **Data collection:** it collects data from up to 30 streams simultaneously, both from hardware sensors (e.g., GPS, accelerometer, gyroscope, among others) and from software sensors (e.g., in/out calls, running applications).
2) **Annotation:** it administers the time diary from Section II asking students about their activities, location and social relations with a fixed time interval between questions, as shown in Figure 3.

SmartUnitn involved 72 students selected from the ones enrolled at our university during the Academic year 2015-2016, and in particular, only those fulfilling these criteria:

1) to have filled three university surveys to obtain their socio-demographic data, shown in Table II, and other characteristics, e.g., psychological and time use related;
2) to attend lessons during the period of our experiment, so that they could describe their daily behavior during the university experience;
3) to have an Android smartphone with an Android version 5.0.0 or higher.

TABLE II: Socio-demographics of students from SmartUnitn.

Gender		Departments		Scholarship	
Male	Female	Scientific	Humanities	True	False
61.1%	39.9%	56.9%	43.1%	37.5%	62.5%

The students were asked to attend a presentation where they were presented with the aims of SmartUnitn and how to use the

[1] See http://trams.disi.unitn.it for more information

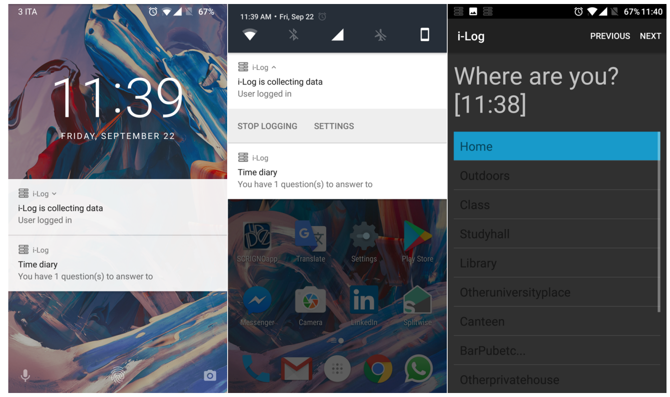

Fig. 3: i-Log is unobtrusive and does not alter the user experience. It only creates a notification to tell the user that the data collection is running and a second notification when a new question is generated.

application. If they wished to participate, after the presentation they signed a consent form and then installed i-Log on their smartphones. Students were informed about all aspects of their personal information treatment in terms of privacy, from data collection to storage to processing. Furthermore, before starting the data collection, we obtained the approval from our university ethical committee.

SmartUnitn lasted two weeks: during the first week, students were asked to answer a time diary on their smartphone, while the application was collecting sensor data in the background. The questions had a 30-minute interval and could be answered only in the 150 minutes immediately after they were generated; otherwise, they expired. During the second week, students were only required to have the application running for collecting sensor data. After the end of SmartUnitn, students received an economic compensation for their participation.

The resulting 110 Gb dataset is a behavioral annotated dataset that exploits sociological insights from the very beginning and, also with sensor data and answers, it is also merged both with socio-demographic characteristics of students provided by our university.

IV. Evaluating Annotation Consistency

Evaluating the quality of the annotations in the wild is not easy. In fact, labels cannot be evaluated without the corresponding sensory data, which makes developing a general approach a challenging task. To the best of our knowledge, there are still no systematic approaches for assessing the quality of in the wild annotations in the literature.

We propose to evaluate our annotations quality by focusing on two specific annotations types, i.e., locations and movements. The main reason for choosing these types of annotations is that both locations and movements are relatively easy to recognize from the point of view of sensing strategies since there is no need to rely on complex combinations of sensor data to identify them. An additional reason for locations is that they are unlikely to change function or position during the time of the project. This last element is especially relevant because we had no external ground truth to compare the annotations with, e.g., students' home addresses were not provided for privacy concerns.

A. Annotation Clustering

An annotation A is formalized as a tuple $< L, LOC >$, where L is the label representing the answers to the questions "Where are you?" or "How are you traveling?", and LOC is the physical location represented as a triple $<$*latitude*, *longitude*, *altitude*$>$ that was collected by the smartphone sensors at the time of the question generation. This location was collected either from the GPS sensor or calculated through the network Wi-Fi connection.

All the annotations were then clustered together by using the DBSCAN algorithm [6] regardless of the collection time. DBSCAN is a density-based clustering algorithm that, given a set of points in some space, groups together points that are closely packed together (points with many nearby neighbors), marking as outliers points that lie alone in low-density regions (whose nearest neighbors are too far away). The purpose of clustering by locations is to obtain a series of proxy locations or movements for every user. More formally, a cluster of annotations is defined as a set of labels $CL = [L_0, L_1, ..., L_N]$, where L_i is the *i*-th label, N is the total number of labels in the cluster, and *M* is the number of total unique labels, with $M \leq N$. An alternative way to see a cluster is to group together all the instances of the same M labels in the cluster and count their occurrence O so that $CL = [L_0 : O_0, L_1 : O_1, \ldots, L_M : O_M]$, which is a vector of pairs $L_i : O_i$. For instance, consider these two clusters belonging to the same user: $CL_1 = [Home : 132, Work : 1, Library : 1]$ and $CL_2 = [Home : 48, Work : 40, Library : 46]$. We can see in CL_1 that the user meant to annotate his or her own home, while "Work" and "Library" are clearly outliers, due to, e.g., uncertainty in the measurement accuracy or the wideness of the window for collecting the sensor information. On the other hand, it is unclear in CL_2 what was the actual location the user was referring to. Finally, a user U can be represented as the set of Q clusters of his or her annotations $U = [CL_0, CL_1, \ldots, CL_Q]$.

B. Defining Consistency

We base our intuition of consistency on the entropy $H(X)$ as defined in information theory [22]

$$H(X) = E[-\ln P(X)] = -\sum_{i=1}^{n} P(x_i) \log_b P(x_i) \quad (2)$$

where $H(X)$ is the entropy of a discrete random variable X with possible values $\{x_1, ..., x_n\}$ and probability mass function $P(X)$. The entropy is a measure of unpredictability of the information represented as a number between 0 and 1, where 0 means completely predictable and 1 means completely unpredictable. If all the labels in a cluster are the same, this means that the annotation task is consistent, and then the unpredictability is null. On the other hand, if all the labels in the cluster are different, this means that the cluster is highly unpredictable. Since the consistency should intuitively be better when high, while entropy behaves in the opposite way, we decided to define the C_{CL} consistency of a cluster CL as

$$C(CL) = 1 - H(CL) \quad (3)$$

where $CL \in [0, 1]$ is the random variable, and, recalling the definition $CL = [L_0 : O_0, L_1 : O_1, \ldots, L_M : O_M]$, the consistency formula becomes

$$C(CL) = 1 + \sum_{i=1}^{M} P(L_i) \log_M P(L_i) \quad (4)$$

where

$$P(L_i) = \frac{O_i}{\sum_{j=0}^{M} O_j} \quad (5)$$

is the probability of the occurrence of the *i*-th label L_i in the cluster CL. After these considerations, it can be represented as $CL = C_{CL} : NA$ composed by the consistency value C_{CL} and the number of annotations in the cluster $NA = \sum_{i=0}^{M} O_i$.

In order to compute the user consistency value $\overline{C}_U$ recalling that $U = [CL_0, CL_1, \ldots, CL_Q]$, we compute the weighted average of the consistency value of each cluster accounting for the number of annotations in each of them

$$\overline{C}_U = \frac{1}{\sum_{i=0}^{Q} NA_i} \sum_{i=0}^{Q} C_i(CL) NA_i \quad (6)$$

C. Assessing Location and Movement Consistency

We computed the consistency values $\overline{C}_U$ using Equation 6 for all the 72 students in our dataset. In order to better evaluate the results of the consistency analysis, we decided to use four different annotation sets for each user: *i) Labels*, *ii) Semantic Labels*, *iii) Random Baseline*, and *iv) Random Stratified Baseline*, as shown in Figure 4.

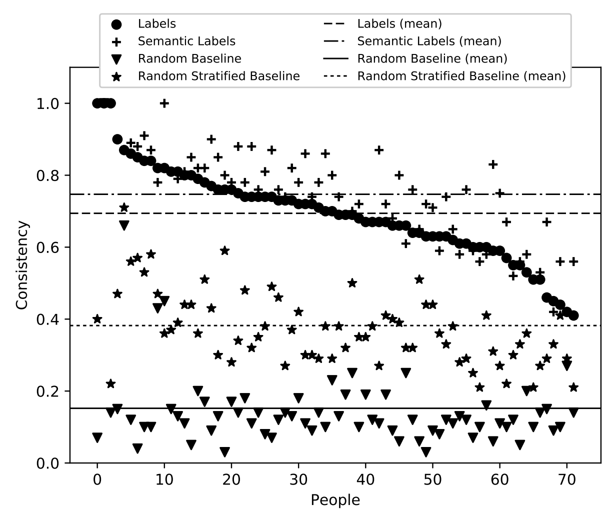

Fig. 4: The four consistency values, on the Y axis, of the students, on the X axis.

Labels are the user annotations as they were provided during the project. The result is a mean consistency value of 0.69 (Sd 0.12). As for *Semantic labels*, they are the semantically related labels that are grouped after a pre-processing step leveraging on the semantics of the ontology concepts as in Section II. This step followed two different levels of abstractions:

1) **Being in buildings vs. Travelling:** We divided the annotations according to whether they referred to actual buildings or whether they referred to any traveling. Dividing them meant grouping under the latter all the following annotations: “Outdoors”, “By Foot”, “By Bus”, “By Train”, “By Car”, “By Motorbike”, and “By Bike” and leaving the rest as they were.
2) **Home vs. University:** We further divided the annotations according to whether they referred to educational, residential or other buildings. We focused on the first two since they are the main contexts in which students spend their everyday lives. For educational buildings, we grouped “Class”, “Study hall”, “Library”, and “Other university place”, while for the residential buildings we grouped “Home” and “Other private house”.

These two categories allowed us to smooth the distribution of closely related spatial elements, e.g., study hall and libraries that belong to the same building and semantically refer to the educational context, and distinguishing when a user was moving and how. The final result obtained by adding this semantic step is 0.74 (Sd 0.13), which further improves the initial result by students.

To properly assess the consistency value as defined in this work, it must be compared with a baseline value. For this reason, we present two different baselines, as shown in Figure 4. The first baseline is the *Random Baseline*, which as the name suggests, simulated an entirely random behavior of each user. More specifically, we considered the same number of labels of each user annotation from the original dataset, and we replaced them by randomly choosing one of the 19 possible labels from the time diary. As expected, the consistency value is lower for all users, showing a final value of 0.15 (Sd 0.13). The second baseline, called *Random Stratified Baseline*, relied on stratified random sampling [21]. It randomly provides labels accounting for their original distribution for every user. The mean consistency value for the *Random Stratified Baseline* is 0.38 (Sd 0.12).

The results of the evaluation show that users’ annotated their own data almost five times more consistently than the entirely random baseline and 50% more consistently than the stratified random baseline. It also shows that exploiting semantics in the whole process from the building of annotations to the analysis can improve the results by an additional 7.2%. It is important to notice that, concerning standard approaches which rely on constrained environments and human expert intervention, the labels were collected at a significantly low rate (1 every 30 minutes). Given such a low value, because the labels were collected in a real-world scenario, the results are even more significant.

V. Related Work

Our work is within the research area of participatory sensing [16] and mobile crowdsensing [13]. The main idea is to have users collect, share and annotate sensed data from their surroundings using their smartphones. Recently, there has been increasing interest in both research areas in understanding the best approaches to elicit and assess the quality not only of sensor data, but their annotations as well. [4] analyze three approaches for the data collection, i.e., *Participatory* (PART), *Context-Triggered In Situ* (SITU), and *Context-Triggered Post Hoc* (POST), in an experiment on travelling habits. PART means that users actively collect data for the whole experiment. SITU and POST refer to obtaining annotations from user data either when a specific condition is triggered or to prompt users after the condition to obtain retrospective annotations, respectively. The results suggest the PART approach is the most effective one since it produces a larger amount of activity data and with less noise, although SITU and POST leads to more activity recordings. Our main difference with respect to these works is that we provide a simpler and standardized methodology for evaluating users’ annotations, thus removing the need of expert validators.

In terms of dedicated technologies for in the wild annotations, one early example is MyExperience [7]. It is a open source mobile data collection tool developed for Windows Mobile devices (including PDAs and mobile phones). It combines sensing and self-reports to collect both quantitative and qualitative data on human behaviors and activities in the field. MyExperience is based on a three-tier architecture of sensors, triggers and actions; triggers use sensor event data to conditionally launch actions. More recently, other applications able to annotate collected sensor data have been proposed. [15] proposed the idea of “mission”, i.e., a sequence of selecting activity class and device position as well as performing the activity. Over 35,000 activity data were gathered from more than 200 users over 13 months. However, only one type of sensor data was collected, i.e., the accelerometer of the participants’ smartphones. Instead, [14] presented an annotation system using multi-sensory stream for daily activity. It segments each day only in a small set of meaningful events which the user has to annotate with multiple tags categorized by activity, place and people, e.g., eating in a restaurant with friends. However, the system has only been evaluated by one volunteer. The major difference with respect to our work is our interdisciplinary methodology for building our annotations.

Our definition of context is quite different from that usually found in the pervasive computing community (see [1] for a survey). To us, context is an intermediate representation layer which bridges the human and machine representations of the world, rather than an aggregation mechanism exploited by machines to reason about sensor data [1]. As such, our notion of context does not need to represent uncertainty. We implement reasoning by representing context as an ontology and by exploiting efficient Modal/Description Logic inference engines [12], [8].

VI. Conclusions

In this work, we proposed an evaluation of an annotation process that relies on an interdisciplinary approach for obtaining annotations in the wild. The evaluation was done on a dataset from SmartUnitn that involved students in understanding how their everyday behavior impacts on their academic performance. We focused on those annotations concerning the movements and locations of students during the project, and we evaluated the annotations quality by checking their consistency. We showed that students were fairly consistent in annotating the spatial dimension, and we also showed that this result could be improved by accounting for the semantics of labels. Future work will consist of using the notion of consistency with other user dimensions, e.g., activities, and perform new iterations of SmartUnitn to increase both in size of participants and duration.

Acknowledgment

The project leading to this application has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 732194. Also special thanks to our colleague David Leoni for his invaluable input.

References

[1] C. Bettini, O. Brdiczka, K. Henricksen, J. Indulska, D. Nicklas, A. Ranganathan, and D. Riboni, "A survey of context modelling and reasoning techniques," *Pervasive and Mobile Computing*, vol. 6, no. 2, pp. 161–180, 2010.

[2] M. Borazio and K. Van Laerhoven, "Improving activity recognition without sensor data: a comparison study of time use surveys," in *Proceedings of the 4th Augmented Human International Conference*. ACM, 2013, pp. 108–115.

[3] P. Bouquet and F. Giunchiglia, "Reasoning about theory adequacy. a new solution to the qualification problem," *Fundamenta Informaticae*, vol. 23, no. 2, 3, 4, pp. 247–262, 1995.

[4] Y.-J. Chang, G. Paruthi, H.-Y. Wu, H.-Y. Lin, and M. W. Newman, "An investigation of using mobile and situated crowdsourcing to collect annotated travel activity data in real-word settings," *International Journal of Human-Computer Studies*, vol. 102, pp. 81–102, 2017.

[5] B. J. Claessens, W. Van Eerde, C. G. Rutte, and R. A. Roe, "A review of the time management literature," *Personnel review*, vol. 36, no. 2, pp. 255–276, 2007.

[6] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *Kdd*, vol. 96, no. 34, 1996, pp. 226–231.

[7] J. Froehlich, M. Y. Chen, S. Consolvo, B. Harrison, and J. A. Landay, "Myexperience: a system for in situ tracing and capturing of user feedback on mobile phones," in *Proceedings of the 5th international conference on Mobile systems, applications and services*. ACM, 2007, pp. 57–70.

[8] E. Giunchiglia, F. Giunchiglia, R. Sebastiani, and A. Tacchella, "More evaluation of decision procedures for modal logics," *KR*, vol. 98, pp. 626–635, 1998.

[9] F. Giunchiglia, "Contextual reasoning," *Epistemologia, special issue on I Linguaggi e le Macchine*, vol. 16, pp. 345–364, 1993.

[10] F. Giunchiglia, E. Bignotti, and M. Zeni, "Personal context modelling and annotation," in *Pervasive Computing and Communications Workshops (PerCom Workshops), 2017 IEEE International Conference on*. IEEE, 2017, pp. 117–122.

[11] F. Giunchiglia and M. Fumagalli, "Concepts as (recognition) abilities," in *Formal Ontology in Information Systems: Proceedings of the 9th International Conference (FOIS 2016)*, vol. 283. IOS Press, 2016, p. 153.

[12] F. Giunchiglia and R. Sebastiani, "Building decision procedures for modal logics from propositional decision procedures: The case study of modal K (m)," *Information and Computation*, vol. 162, no. 1-2, pp. 158–178, 2000.

[13] B. Guo, Z. Yu, X. Zhou, and D. Zhang, "From participatory sensing to mobile crowd sensing," in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference on*. IEEE, 2014, pp. 593–598.

[14] J. Hamm, B. Stone, M. Belkin, and S. Dennis, "Automatic annotation of daily activity from smartphone-based multisensory streams," in *International Conference on Mobile Computing, Applications, and Services*. Springer, 2012, pp. 328–342.

[15] Y. Hattori, S. Inoue, and G. Hirakawa, "A large scale gathering system for activity data with mobile sensors," in *Wearable Computers (ISWC), 2011 15th Annual International Symposium on*. IEEE, 2011, pp. 97–100.

[16] S. S. Kanhere, "Participatory sensing: Crowdsourcing data from mobile smartphones in urban spaces," in *Mobile Data Management (MDM), 2011 12th IEEE International Conference on*, vol. 2. IEEE, 2011, pp. 3–6.

[17] N. Kern, B. Schiele, and A. Schmidt, "Recognizing context for annotating a live life recording," *Personal and Ubiquitous Computing*, vol. 11, no. 4, pp. 251–263, 2007.

[18] W. Z. Khan, Y. Xiang, M. Y. Aalsalem, and Q. Arshad, "Mobile phone sensing systems: A survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 402–427, 2013.

[19] A. Melis, S. Mirri, C. Prandi, M. Prandini, P. Salomoni, and F. Callegati, "Crowdsensing for smart mobility through a service-oriented architecture," in *Smart Cities Conference (ISC2), 2016 IEEE International*. IEEE, 2016, pp. 1–2.

[20] D. Peng, F. Wu, and G. Chen, "Pay as how well you do: A quality based incentive mechanism for crowdsensing," in *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*. ACM, 2015, pp. 177–186.

[21] M. Shahrokh Esfahani and E. R. Dougherty, "Effect of separate sampling on classification accuracy," *Bioinformatics*, vol. 30, no. 2, pp. 242–250, 2013.

[22] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.

[23] P. A. Sorokin and C. Q. Berger, *Time-budgets of human behavior*. Harvard University Press, 1939, vol. 2.

[24] M. Zeni, I. Zaihrayeu, and F. Giunchiglia, "Multi-device activity logging," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 2014, pp. 299–302.