

Activity Recognition Enhancement Based on Ground-Truth: Introducing a New Method Including Accuracy and Granularity Metrics

Hamdi Aloulou^{1,2}(✉), Romain Endelin^{1,2}, Mounir Mokhtari^{1,2,3},
Bessam Abdulrazak⁴, Firas Kaddachi², and Joaquim Bellmunt^{1,3}

¹ Institut Mines Telecom, Paris, France

{hamdi.aloulou,romain.endelin,mounir.mokhtari}@mines-telecom.fr

² Laboratory of Informatics, Robotics and Microelectronics, Montpellier, France

firas.kaddachi@lirimm.fr

³ Image and Pervasive Access Laboratory, Singapore, Singapore

joaquim.bellmunt@ipal.cnrs.fr

⁴ University of Sherbrooke, Sherbrooke, Canada

bessam.abdulrazak@usherbrooke.ca

Abstract. The uncertainty associated with existing sensing technologies and reasoning methods affects the outcome of the activity recognition process (e.g., accuracy, precision, granularity). The activity recognition process is even challenging when switching from laboratory towards real deployments, where scenarios are not predefined and more complex. Therefore we propose a novel method to improve the activity recognition outcome, by finding a proper balance between accuracy and granularity. The method has been validated through the deployment of UbiSMART (an AAL framework) in 45 scenarios of ageing in place. We discuss in this paper our method and the validation results.

Keywords: Ambient assisted living · Activity recognition · Semantic reasoning · Quality insurance · Ground-truth acquisition

1 Introduction

The acceptance of Ambient Assisted Living (AAL) solutions by stakeholders is a critical factor for the success and large use of such systems. Different criteria influence the acceptance of AAL solutions such as the usability, usage, usefulness. According to EN ISO 9241-11 [8], usability is defined as the degree of suitability of use of a system, a prototype, or a service in a particular application environment to achieve specific goals in a satisfactory and efficient manner. In this perspective, context-awareness and activity recognition are critical components leading to the acceptability and usability of AAL solutions.

Context-awareness and activity recognition are necessary to trigger adequate services on appropriate media, given a certain situation/context. An example can

be to send an alert when a risk/dangerous situation is detected. Thus, fault inferences (i.e., lack of accuracy and granularity) in activity recognition are among the most critical issues we have identified in our experience with AAL. An inaccurate system generates a number of misleading reactions affecting the acceptability of the solution, whereas a coarse-grained system can possibly be accurate, but hardly useful.

Following, we present our approach to improve the activity recognition outcome based on collected ground-truth data. The approach adopts a hierarchical representation of activities and introduces two metrics “Accuracy” and “Granularity” in the activity recognition process. Section 2 illustrates the problems of fault inferences. Section 3 positions our contribution within the literature. Section 4 introduces our method to quantify the accuracy of a reasoning engine. Section 5 discusses our method to optimize the decision-making. We also introduce in this section a score of quality for activity recognition. Section 6 presents the validation of our method. Finally, Sect. 7 concludes the paper.

2 Granularity and Accuracy Issues in Activity Recognition

We propose to introduce the use of two metrics “accuracy” and “granularity” in the activity recognition process to improve its outcome and reduce fault inferences. For clarity purpose, it is important to distinguish the concepts of *granularity* and *accuracy*: The granularity represents how much precise an activity recognition engine was; for instance, a fine-grained inferred activity would be “on the phone with his daughter,” and a coarse-grained activity would be “in the living-room.” Therefore, we assign a granularity level to each specific activity. On the other hand, accuracy is defined in the literature [7] as the confidence that the inferred activity matches the reality (i.e., the confidence of having a correct inference). We use accuracy to determine the ability of an activity recognition engine to properly detect a specific activity. A system with 99% of accuracy would be considered highly accurate, whereas another system with 20% would be considered inaccurate.

From our experience, we can deduce that the more fine-grained an activity is, the less accurate it will be. To illustrate this, Fig. 1 represents two chains of hierarchical activities (one in circles and one in triangles). Activities on the upper left corner are the most fine-grained (for example eating pasta), however, they are difficult to infer and therefore they have low accuracy. On the other side, activities on the lower right corner are easy to infer (for example located in the kitchen) and therefore have the highest accuracy. However, they are too coarse-grained to provide appropriate services for the end-user. Therefore, our method presented in this paper intends to find the best balance between granularity and accuracy, in order to improve the outcome of the activity recognition process.

Following we discuss existing works on the evaluation and validation of reasoning engines, with techniques used to collect ground-truth and the employment of the “Accuracy” metric to evaluate the activity recognition engines outcome.

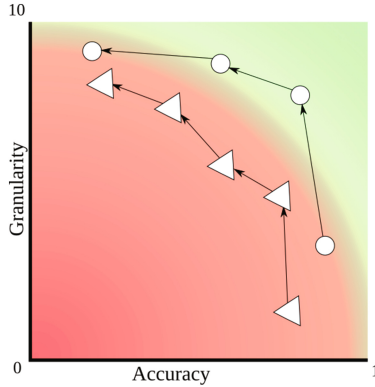


Fig. 1. Illustrating the relation between accuracy and granularity for activity recognition

3 Related Works

A large number of AAL solutions prototypes exist around the world [3, 14, 17]. Nevertheless only few of them involve long-term deployments with real data gathering [15]. Yet, we witness recently a shift in AAL research, from laboratory experiments towards real deployments. This shift is accompanied by the creation of several databases that enable researchers to share the real-world deployment recorded data. CASAS¹ is an example of these databases.

The limitation in the number of AAL solutions aiming for long-term deployment is a big burden towards the validation of the accuracy of such solutions. We believe that the availability of databases for real-world deployment records will promote the development and validation of approaches to improve accuracy of AAL solutions. In addition, a key step towards wide acceptance of AAL solutions is the evaluation and validation of reasoning engines [1]. The legacy approach for assessing elderly people Activities of Daily Livings (ADLs), whether by direct observations [10] or by questionnaires [6, 13], are very instructive, but they lack practical applications towards activity recognition in the AAL. These solutions are difficult to put in place in nursing homes and especially in personal residences for end-users living independently as the observations are limited to specific points of time when caregivers are available. They are based on a manual process which is tedious and time-consuming. Thus, some technological solutions should be put in place to easily collect elderly people ADLs and directly introduce them in the evaluation and improvement process of the AAL solution accuracy.

Several researchers have been interested in accuracy using the datasets gathered from the real-world deployment [4, 5, 9]. For example, Cook [5] has used CASAS datasets to perform machine-learning methods for activity recognition; the results of her team were around 75% on rich training datasets. Kleinberger et al. [11] performed a thorough validation of their system, measuring its

¹ <http://ailab.wsu.edu/casas/datasets.html>.

accuracy with the use of the well-established Goal-Question-Metric (GQM) approach [2]. They observe an accuracy of 92% of correct inferences on average for simple ADLs such as “Going to Toilet.” Kadouche et al. used Support Vector Machine (SVM) for activity recognition and they obtained an accuracy of 88% [9]. Chung et al. [4] applied activity recognition in an application targeting nursing home. Using Hierarchical Context Hidden Markov Model (hchmm), they obtain a recognition accuracy of 85%. Nevertheless, their activity recognition process relies on cameras, which is often associated with acceptability issues.

These approaches use “Accuracy” as a metric to evaluate the performance of their reasoning engine in activity recognition. On the other hand, the results are not used in a systematic process to improve the reasoning approach or methodology. The method we propose in this paper uses both metrics accuracy and granularity in an iterative process to gradually improve the reasoning engine outcome without changing the approach used or bringing new sensing technologies and complicate the deployment. In addition, researches to improve activity recognition have been supported by the development of novel algorithms based on Artificial Neural Network, Naive Bays, Support Vector Machine, etc. [16]. The method we introduce in this paper to improve the activity recognition outcome is reasoner-agnostic. It can be applied with any approach for activity recognition (e.g., machine-learning, ontological reasoning).

We believe that accuracy and granularity are key criteria in the validation of the activity recognition outcome, since accuracy is an indicator of the reliability of an AAL system and granularity is an indicator of its usefulness. We also believe that using these two indicators in an iterative process of activity recognition will improve the reasoning engine outcome. Our method improves the decision-making process on the output of the reasoning (i.e., the set of activities possibly being performed by the end-user), through accuracy, granularity, and score of the reasoning engine.

4 Measuring Accuracy of a Rule Engine

We discuss in this section the measurement of the deployed solution accuracy. As an example of inaccurate inference, the system infers that the end-user “*Watches TV*” at 14:00, whereas we know from direct observations that he actually “*Takes a Nap*.” There is a need to know how accurate a reasoning engine is, in order to improve the quality of activity recognition. Thus, there is a need to measure the confidence towards the fact that an activity is actually being performed, given that it has been inferred. To obtain this confidence, we first need to enrich our datasets with real observations (*ground-truth*), in order to confront inferred activities with observed ones. The relation between real activities, sensor events, and inferred activities can be summarized in Fig. 2. Following we discuss the process of gathering ground-truth and measuring the accuracy of activities.

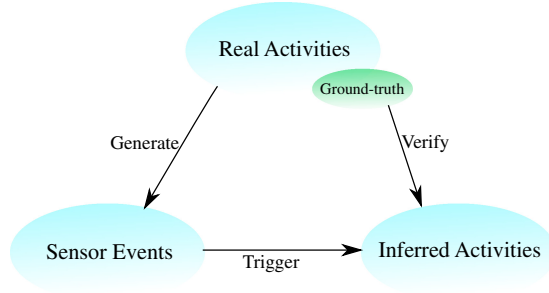


Fig. 2. Relation between Real Activities, Inferred Activities, and Sensor Events

4.1 Gathering Ground-Truth

The most straightforward method to gather ground-truth is to perform direct observation in a real-world environment. A human observer regularly observes and records what end-users do. The result is a list of punctual observations of activities over a period of time. Through this method, we can be certain that the observations represent a real situation. However a bias may be introduced by the sampling periods: the observers are more likely to perform the ground-truth acquisition only at specific hours of the day (e.g., in morning time for nurses), ending up with an heterogeneous density of data, that must be translated later in the measurement process. The ground-truth acquisition is a manual process, and it is not immune to human errors. This manual process is time-consuming, which restricts observers recruiting.

This method may also have logistic difficulties and acceptance issues, particularly in the case of collecting data linked to end-user living independently. We have experienced this situation ourselves in individual houses. A solution would be to ask caregivers to perform acquisition, but it would only bring little benefit, as caregivers would actually influence the environment they observe. In fact, in this situation, ground-truth acquisitions would take place in a multi-user situations (i.e., end-user and caregiver), in which the acquired data sensors cannot be considered linked only to the end-user.

We have provided a dedicated mobile application which sends a quick questionnaire when the system wants to verify the reasoning output or a risky situation is detected. This solution was used by our collaborating nursing home's caregivers and helped to semi-automatically collect ground-truth data without seriously affecting the caregivers daily routines. It also allowed integrating data directly into our process of activities accuracy measurement.

Another approach to overcome the limitations of the direct observation by human would be using cameras. This method is rich in data, however, we did not consider it due to acceptability and privacy concerns.

4.2 Measuring Accuracy of Activities

One goal of our research is to give a confidence value to an inferred activity based on our ground-truth observations of real activities. More precisely, we are interested in a metric based on the probability $P(A = a|I = a)$ that an activity a is being actually performed (A), given that it has been inferred (I). In other words: “is the person really doing a , when a reasoner says a ?” In order to measure the accuracy of activity recognition, we apply the Bayes equation of probability [12]. We define the metric as follows:

$$P(A = a|I = i) = \frac{P(A = a \cap I = i)}{P(I = i)} = \frac{|a \cap i|}{|A \cap I|P(I = i)} = \frac{|a \cap i| \sum_{X \in I} duration(X)}{|A \cap I| \sum_{x \in i} duration(x)} \quad (1)$$

where:

$|a \cap i|$ is the number of occurrences when i is inferred and a is observed

$|A \cap I|$ is the number of observations made while inferring an activity

$\sum_{X \in I} duration(X)$ is the total duration covered by our inferences

$\sum_{x \in i} duration(x)$ is the total duration when i is inferred

In the case when the activity has never been observed or never been inferred, we set accuracy value of this activity as 0.

5 Improving the Decision-Making Process by Introducing a Score of Quality

We believe that the accuracy metric can be helpful to evaluate the quality of a reasoning engine. By coupling accuracy with granularity, and introducing hierarchical activities models, we propose a systematic method to improve the activity recognition outcome and help the reasoner to conclude with more effective inferences.

5.1 Introducing Granularity to the Reasoner

The reasoner may infer “Takes Shower” whereas the ground-truth shows a faulty inference, as the end-user “Goes to Toilet.” It could also infer “Is Busy”, which is not an adequate inference for a service delivery. These two cases are caused by an irrelevant granularity. In fact, in some cases, the reasoner tends to be too fine-grained, and leads to an inaccurate inference. The reasoner could infer a slightly less precise activity with more accuracy.

Generally, a human observer intuitively adjusts his conclusions to the expected level of granularity. He also measures the risks of being inaccurate

when he makes fine-grained conclusions. Our approach is to apply the same process for decision-making in activity recognition, so that the system will minimize the risk of inaccuracy and maximize granularity. Thus, we represent granularity in our model as a value ranging from “1” to “10,” determined by an expert. “1” would be an extremely coarse-grained activity and “10” would be an extremely fine-grained one. We argue that introducing granularity into the model as an arbitrary value is acceptable because an acceptable granularity is a non-functional requirement of a system, which is by nature arbitrary.

5.2 Introducing Hierarchical Activities

“Takes Shower” and “Goes to Toilet” can both be generalized as “Is in the Bathroom.” In other words, “Is in the Bathroom” is a parent activity of “Takes Shower” and “Goes to Toilet.” More generally, activities exist at different granularities, and can be modeled as hierarchies. With a hierarchical model, several activities are inferred at a given time, and a person can be both “Going To Toilet” and “In the Bathroom.” When an activity occurs, all of its generalizations occur, recursively. Similarly, we can affirm that when an activity occurs, some of its specializations may occur too. An example of a hierarchical model is presented in Fig. 3, and the generalization inference goes from right to left.

Hierarchical models are richer than linear models, and the ability to infer activities on several layers has powerful applications when combined with granularity and accuracy. In a hierarchical model, an activity always has a higher granularity than its parent. On the other hand, an activity always has a lower accuracy than its parent: if an activity is accurate, its parent will always be accurate, but if an activity is not accurate, its parent will sometimes be accurate. Following this logic, we created a formal process to perform accurate inferences while being as fine-grained as possible. The reasoning process is similar to a tree exploration, where the system starts by reasoning on the most coarse-grained activities (i.e., the activities with no parent) to its specializations (more fine-grained). In this path (parent \rightarrow kid), the reasoning checks whether the context can be valid or not (for of each specialization). This process can be executed recursively across specializations until the inference of a chain of activities, from coarse-grained to fine-grained (Fig. 1). We expect that with this process, a system converges towards the activity that has the best balance between granularity and accuracy.

5.3 Measuring Performance of the Reasoning Engine

We propose to formalize the process of inferring activities with the right level of granularity. Therefore, we introduced a score of quality of activity recognition that enables the reasoner to converge towards the activity having the best balance of accuracy vs. granularity. Our proposed process starts by giving each activity a score, that is based on its accuracy and granularity. Then, the decision-making engine selects the activity with the best score. The system uses a weighted geometric mean of accuracy and granularity (Eq. 2) to measure

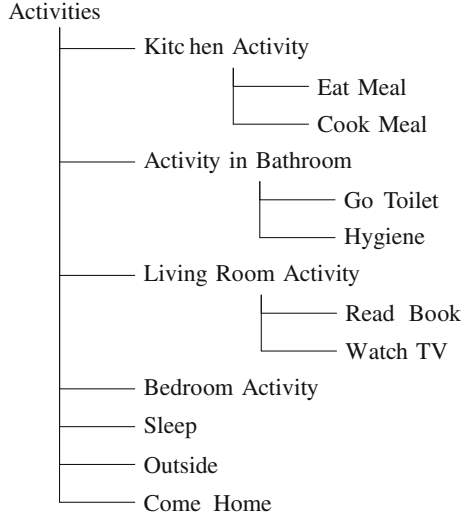


Fig. 3. A hierarchical set of activities

score. The advantage of a geometric mean is that a marginally low value of either granularity or accuracy has a dramatic impact on the resulting score, and an accuracy of 0% will result in a score of 0.

$$Score(a) = \left(accuracy(a)^A \times (0.1 \times granularity(a))^G \right)^{\frac{1}{A+G}} \tag{2}$$

Where:

A is the weight given to accuracy

G is the weight given to granularity

The 0.1 factor normalizes granularity in range [0, 1]

A and G are to be defined by experts. However, there is no objective criteria to set them. From our empirical experience, we choose to give more importance to accuracy than to granularity and we set $A = 3$ and $G = 1$. The goal is to have the most useful sets for the end-user experience. We also propose to run reasoners with various values of A and G , and ask end-users which reasoner generates the most useful conclusions.

6 Validation

Our proposed method was validated using our AAL framework Ubiquitous Service Management & Reasoning sysTem (UbiSMART). UbiSMART was deployed in a real environment where several scenarios of aging in place were performed. The first version did neither include the hierarchical activities approach, nor the metrics of Accuracy and Granularity. We have updated UbiSMART in order to

support accuracy and granularity. The new UbiSMART reasoner can be executed in two different ways:

1. **In calibration mode:** we execute the reasoner in order to measure the accuracy of all activities in hierarchical chain. We expect the reasoner to return all possible valid inferred activities, without making a conclusion. The result of the calibration mode is used as an input for the production mode. The calibration mode is run only at the first phases of deployment. Once the reasoner has converged towards stationary accuracy values, there is no need to run calibration mode anymore.
2. **In production mode:** we execute the reasoner in order to infer a single activity at a given time. In this case, accuracy measurement is predefined and calculated during the calibration mode. The production mode is the mode used by UbiSMART framework to deliver adaptable services.

We use the same collected datasets to run the UbiSMART’s reasoner, before and after applying the proposed method. We have experimented our system in the environment represented in Table 1 and with the activities summarized in Table 2.

Table 1. Topology of the experiment

Room	Bedroom	Bathroom	Living Room	Kitchen	Entrance
Sensors	Motion sensor	Motion sensor	Motion sensor	Motion sensor Door sensor (fridge)	Motion sensor Door sensor (main door)

6.1 Introducing Validation Metrics

We propose four metrics to validate our proposed method. The first metric is the Recall R (e.i. total measured accuracy vs. ground-truth) as defined in Eq. 3. R is useful to measure the exact accuracy of the reasoning, given that we have a ground-truth on the executed dataset. R is similar to the measured accuracy of an activity (Eq. 1), but it is measured on all activities at once (A), not on a specific activity. It is defined as the number of times the system inferred correctly (based on ground-truth observations), divided by the total number of activities that are both inferred and observed.

$$R(A) = \frac{|\{groundtruth(a) = inferred(a) | a \in A\}|}{|groundtruth(A) \cap inferred(A)|} \quad (3)$$

Three other metrics have been introduced: the average value of accuracy (\bar{A}), granularity (\bar{G}) and score (\bar{S}) for the inferred activities (Eq. 4). These three metrics provide an estimation of accuracy, granularity, and score of reasoning engine in a dataset, even in the absence of ground-truth. \bar{A} is not to be confused with the total measured accuracy R . \bar{A} is an indicator that can be obtained at any time, whereas R is exact, but requires ground-truth to be measured.

$$\bar{X}(act) = \frac{\sum_{a \in act} (X(a) \times d(a))}{\sum_{a \in act} d(a)} \quad (4)$$

where: X = accuracy — granularity — score; act = activities; d = duration

6.2 Results

We run UbiSMART reasoning engine in both modes (i.e., calibration and production). Table 2 presents accuracy, granularity and score values for each activity, after the calibration mode. Score is calculated using Eq. 2 from Sect. 5.3, with an accuracy weight $A = 3$ and a granularity weight $G = 1$. For comparison, a second score is calculated, using $A = 1$ and $G = 1$. We observe that activities that are coarse-grained but extremely accurate, such as “*Kitchen Activity*”, are more valued with a higher value of A . With $A = 3$, it has a score of 79.5%, whereas it only has a score of 63.2% with $A = 1$. On the opposite, a more fine-grained and less accurate activity, such as “*Cook Meal*” has a score of 64.5% with $A = 3$, and 69.3% with $A = 1$. When both “*Kitchen Activity*” and “*Cook Meal*” are inferred, the reasoner will conclude with “*Kitchen Activity*” if $A = 3$, and with “*Cook Meal*” if $A = 1$.

We run the reasoner four times in production mode: with ($A = 3, G = 1$) and with ($A = 1, G = 1$), in both case *before* and *after* running the calibration (Table 3). Without calibration, each activity has a default accuracy of 10%, and scores are calculated accordingly. With ($A = 3, G = 1$), we measure $T = 93.8\%$ after the calibration, whereas R was only 63.4% before the calibration (+30.4%). \bar{G} has decreased from 6.75 to 4.15 (−2.60), and \bar{A} has increased from 55.5% to 89.4% (+33.9%). This illustrates the trade-off between accuracy and granularity. With ($A = 1, G = 1$), the calibration impact is less significant. R happens to increase by 3.3%. \bar{A} decreases by 11.8%, and \bar{G} increases by 0.26. This is explained by the fact that with $A = 1$, the reasoner is not allowed to decrease much granularity in favor of accuracy. Thus, it will tend to be as fine-grained as possible, which is similar to its default behavior, without the method introduced in this paper. Finally, we notice that \bar{S} increases with both values of (A, G): +16.5% with ($A = 3, G = 1$) and +1.7% with ($A = 1, G = 1$). This is inherent to the method we propose, which always selects the activity with the maximal score among all the inferred activity.

Table 2. Scores of the original UbiSMART reasoner

Activity	Kitchen Activity	Cook Meal	Bathroom Activity	Go Toilet	Living Room Activity	Read book	Bedroom Activity	Sleep
Accuracy	100%	60%	100%	50%	100%	50%	100%	100%
Granularity	4	8	4	8	4	8	6	7
Score $A = 3, G = 1$	79.5%	64.5%	79.5%	56.2%	79.5%	56.2%	88%	91.5%
Score $A = 1, G = 1$	63.2%	69.3%	63.2%	63.2%	63.2%	63.2%	77.5%	83.7%

Table 3. Results with ($A = 3, G = 1$) and ($A = 1, G = 1$), *before* and after the calibration

Sample	A	G	Total Accuracy	Average Accuracy	Average Granularity	Average Score
<i>Before calibration</i>	3	1	63.4%	55.5%	6.75	56.9%
<i>After calibration</i>			93.8%	89.4%	4.15	73.4%
<i>Before calibration</i>	1	1	63.4%	67.3%	5.56	59%
<i>After calibration</i>			66.7%	55.5%	5.82	60.7%

7 Conclusion

We have introduced in this paper our research on improving the activity recognition decision-making process in developing AAL solutions. This research has been motivated by the feedback we had from our real deployment in a nursing home and three individual houses. The observations from this real deployment brought our attention to the faulty results of our reasoning engine and stressed the need for a systematic method to evaluate reasoning engines in order to improve the activity recognition process.

We argue that an efficient systematic method to evaluate reasoning engines has to include ground-truth from deployment environment (e.g., elderly people house). Therefore, we proposed a method that includes observing ground-truth as an input for measuring the accuracy of a reasoning engine. We also introduce for the first time granularity and score of the quality of a reasoning engine. The score is derived from accuracy and granularity. Our method effectively leads to conclude on the most reasonable activity (i.e., the activity with the best balance between granularity and accuracy). We found that by giving more importance to accuracy over granularity, our reasoner infers more coarse-grained activities, in order to be more accurate.

Acknowledgments. This research project has been supported by the Quality Of Life Chair supported by Foundation Telecom of the Institut Mines-Telecom in France, La Mutuelle Generale and REUNICA which figure among the major health-care insurance companies in France. The work is also supported by the grand emprunt VHP inter@ctive project. We also wish to acknowledge the support of the Saint-Vincent-de-Paul nursing home and its director Brigitte Choquet, who kindly let us deploy our system within their environment.

References

1. Abdulrazak, B., Malik, Y.: Review of challenges, requirements, and approaches of pervasive computing system evaluation. IETE Tech. Rev. **29**(6), 506–522 (2012)
2. Basili, V.R.: Software modeling and measurement: the goal/question/metric paradigm (1992)

3. Biswas, J., et al.: Activity recognition in assisted living facilities with incremental, approximate ground truth. In: Geissbühler, A., Demongeot, J., Mokhtari, M., Abdulrazak, B., Aloulou, H. (eds.) ICOST 2015. LNCS, vol. 9102, pp. 103–115. Springer, Cham (2015). doi:[10.1007/978-3-319-19312-0_9](https://doi.org/10.1007/978-3-319-19312-0_9)
4. Chung, P.C., Liu, C.D.: A daily behavior enabled hidden markov model for human behavior understanding. *Pattern Recogn.* **41**(5), 1572–1580 (2008)
5. Cook, D.J.: Learning setting-generalized activity models for smart spaces. *IEEE Intell. Syst.* **2010**(99), 1 (2010)
6. Fillenbaum, G.G.: *Multidimensional Functional Assessment of Older Adults: The Duke Older Americans Resources and Services procedures*. Psychology Press, London (2013)
7. Jahn, A., David, K.: Improved activity recognition by using grouped activities. In: 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops), pp. 1–5. IEEE (2016)
8. Jokela, T., Iivari, N., Matero, J., Karukka, M.: The standard of user-centered design and the standard definition of usability: analyzing ISO 13407 against ISO 9241–11. In: *Proceedings of the Latin American Conference on Human-Computer Interaction*, pp. 53–60. ACM (2003)
9. Kadouche, R., Abdulrazak, B., Mokhtari, M., Giroux, S., Pigot, H.: Semantic matching framework for handicap situation detection in smart environments. *J. Ambient Intell. Smart Environ.* **1**(3), 223–234 (2009)
10. Katz, S., Ford, A.B., Moskowitz, R.W., Jackson, B.A., Jaffe, M.W.: Studies of illness in the aged: the index of adl: a standardized measure of biological and psychosocial function. *JAMA* **185**(12), 914–919 (1963)
11. Kleinberger, T., Jedlitschka, A., Storf, H., Steinbach-Nordmann, S., Prueckner, S.: An approach to and evaluations of assisted living systems using ambient intelligence for emergency monitoring and prevention. In: Stephanidis, C. (ed.) *UAHCI 2009*. LNCS, vol. 5615, pp. 199–208. Springer, Heidelberg (2009). doi:[10.1007/978-3-642-02710-9_23](https://doi.org/10.1007/978-3-642-02710-9_23)
12. Koch, K.-R.: Bayes' Theorem. *Bayesian Inference with Geodetic Applications*. LNCS, vol. 31, pp. 4–8. Springer, Heidelberg (1990). doi:[10.1007/BFb0048702](https://doi.org/10.1007/BFb0048702)
13. Lawton, M.P.: Scales to measure competence in everyday activities. *Psychopharmacol. Bull.* **24**(4), 609–614 (1987)
14. Memon, M., Wagner, S.R., Pedersen, C.F., Beevi, F.H.A., Hansen, F.O.: Ambient assisted living healthcare frameworks, platforms, standards, and quality attributes. *Sensors* **14**(3), 4312–4341 (2014)
15. Mokhtari, M., Aloulou, H., Tiberghien, T., Biswas, J., Racoceanu, D., Yap, P.: New trends to support independence in persons with mild dementia—a mini-review. *Gerontology* **58**(6), 554–563 (2012)
16. Slim, S.O., Atia, A., Mostafa, M.-S.M.: An experimental comparison between seven classification algorithms for activity recognition. In: Gaber, T., Hassanien, A.E., El-Bendary, N., Dey, N. (eds.) *AISI 2015*. AISC, vol. 407, pp. 37–46. Springer, Cham (2016). doi:[10.1007/978-3-319-26690-9_4](https://doi.org/10.1007/978-3-319-26690-9_4)
17. Yared, R., Abdulrazak, B., Tessier, T., Mabilieu, P.: Cooking risk analysis to enhance safety of elderly people in smart kitchen. In: *Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, p. 12. ACM (2015)