Sensor2Vec: an Embedding Learning for Heterogeneous Sensors for Activity Classification

1st Junpei 'Joni' Zhong
School of Science and Technology
Nottingham Trent University
Nottingham, UK
joni.zhong@ntu.ac.uk

2nd Ahmad Lotfi
School of Science and Technology
Nottingham Trent University
Nottingham, UK
ahmad.lotfi@ntu.ac.uk

***Abstract*—Based on the idea of word2vec embedding method in NLP, this paper presents a novel idea called sensor2vec which captures the contextual information of the heterogeneous sensory information in the ambient assisted living setting. The contextual information is essential in order to classify and understand the human activity using multi-modal sensory data. In the activity classification, the sensor2vec embedding method is able to do the pre-processing which produce the embedding layer which represents the semantic value in the high-dimensional space. The preliminary experiment based on LSTM shows that the sensor2vec performs better classification result than the one-hot inputs.**

***Index Terms*—Activity Classification, Embedding Learning, LSTM, Ambient Assisted Living**

I. Introduction

The aging problem has been more and more serious in various countries. From the year 2010 and year 2050, it is predicted that the median age within the global population will increase from 29 years to 36 years old [32]. This problem is more obvious in some of the industrialised countries: as of December 2015, there are four countries whose elderly population exceeds 20% of the whole population: Germany, Italy and Japan. This figure is expected to rise to 13 countries by 2020 [2]. Nevertheless, among the aging population, elderly who live alone are increasing rapidly [29]. Therefore, the increased incidence of these conditions would put an increased burden on healthcare services, considering that the monitoring and healthcare of the elderly who live alone.

Because of this, a consistent surveillance technology is one of the strategies to improve the quality of life for those elderly who live alone. At present, there are plenty of video surveillance technologies used in the public space to recognize activities (e.g. [17]). However, the privacy in the ambient environment another issue that needs to be concerned which is different from the public space: they are sensitive areas, for instance bedrooms, toilets, which may involve the privacy of personal life. Therefore, although the classification of activities may develop quite well, it may not applicable in domestic environments [11].

978-1-7281-8741-9/20/$31.00 ©2020 IEEE

Alternatively, the detection of a day-to-day activity of elderly people should depends on another portable sensors [1], [22] or indoor sensors. These sensors usually hardly have concrete contents which may violate the privacy. Instead, their heterogeneous content may include the basic presence sensors which are designed to deliver Boolean values specifying the body movement within a room.

or can be connected and incorporated into these environments are versatile, ranging from basic presence sensors capable of providing Boolean values specifying the detection of movement within a room, to more invasive solutions capable of identifying specific inhabitants and their actions [7]. Examples of such sensing technologies include passive infrared (PIR) sensors or floor sensors detecting movement within an area, the pressure sensors which are capable of inferring object or door interactions, and pressure sensors embedded in chairs or beds capable of identifying occupancy. Besides of the Boolean values, some sensors could also provide digitized readings such as temperature, illumination, humidity or atmospheric pressure sensors, all of which are useful to detect the parameters in the environment. In all, when all of these technological devices interconnected in the ambient environment, a set of heterogeneous data could be collected which are capable of communicating and learning through the user's habits.

In particular, thanks to the rich content provided by the sensors, it is possible to build the ambient assisted living (AAL) system which is one way to give an non-invasive and technological assistance to the elderly. The AAL systems are integration of both emerging software and hardware technologies, for instance, various sensing techniques, machine learning methods, and even robots [25]. Within the field of AAL, the activity classification is one the essential processes for connecting the users' factor and the smart environment with certain computational intelligence [4], [14]. Such computational intelligence techniques may involve feature extraction, pattern recognition, reasoning, and decision-making procedures, etc.

II. Related Works

A. Human Activity Classification

Generally speaking, the human activity classification using in AAL is to detect and then identify certain human activities which are specialised in the ambient environment. To this aim, the techniques usually need to include both the spatial and temporal information from heterogeneous sensors [24]. The types of sensors used for the classification vary: it can be based on either wearable sensors [5] or sensors fixed at a certain place at home [16] or the combination of both [21].

Nevertheless, data captured by the heterogeneous sensors could be ambiguous, as well as noisy and sparse. It would be necessary to build a constructed and consistent to discover the knowledge and to describe the actual real event [23]. With the a large amount of data, various classification algorithms have been employed, such as naive Bayes (NB) filters [8], which use a probabilistic framework to infer the maximum likelihood based on observations. In NB, parameter estimation usually plays in essential role. Based on the static information, the random forest (RF) [33], k-nearest neighbour (k-NN) [9], and support vector machine (SVM) [13] could be also used. Alternatively, since the contextual information is one of the important factors to be considered while the systems are extracting a higher level of information, the Hidden Markov (HMM) [10], conditional random field (CRF) [30] models could take the consideration of the temporal relations and discover the contextual information dynamically. As a furthermore development about the contextual information, [20], [37] use hierarchical methods for modelling high-level features as topics. Such a hierarchical representation may shed a light to link the activity classification with the Natural Language Processing (NLP) problems.

With the recent development of deep learning, they have also been applied in activity recognition as well. Their advantages are that the raw data can be processed within the models without feature pre-preprocessing. Despite the fact that their training needs even more data, the Convolutional Neural Network (CNN) [12], Recurrent neural network (RNN) [15], [36] and Long Short-term Memory (LSTM) [31] have been adopted in human activity classification. Their multi-modal fusion ability is advantageous to deal with the heterogeneous raw signals. But it is still necessary to develop methods in order to alleviate the training costs of deep learning methods caused by their size of data-set.

B. Contextual Models for AAL

The contextual information has been used in activity recognition using single or multiple sensors. For example, using accelerometer data could be used for context awareness [35], but using a single sensor, especially from the accelerometer, it is difficult to extract enough features from the activities.

In order to understand the construction of structured data, a more formal method [26], [28] used to extract the contextual sensory inputs is proposed to extract the ontology. Based on the temporal logic, [19] a representation language for general events can be used to recognize the composite events. This method has been tested based on the video sensors.

Besides of the activity classification, the NLP techniques have also been adopted in different applications of AAL. They are useful in terms of its enhancement by not only the understanding of the contextual information but also its speech recognition and understanding ability in the AAL setting [6], [34]. Using the NLP, after becoming a correctly-labelled instances, the signals could be easily processed by the statistical machine learning methods supervised or unsupervised language models. Specifically, due to the recent development of NLP method, if we manage to apply the state-of-the-art NLP techniques to the AAL, it would be beneficial to adds great value to deal with the noisy and unstructured information by the abilities of inference and filtering of the NLP methods in the AAL setting.

Finally, to give the sensors used in AAL a meaning, such NLP models which consider the contextual information would be one of the possible methods. In short, both of them have the following common characteristics:

1) both of them own large amounts of information in textual form, and,
2) the necessity to integrate information from heterogeneous sources (sensors or parts of speech).

III. Word2vec and Embedding Learning

Since the extraction of the heterogeneous sensory data in the AAL is one of the issues we are facing, we borrow the concept of word2vec from the NLP and solve the issue. The word2vec uses data-driven method to extract reasonable and efficient features from the language data.

A. Word2Vec

The target of the word2vec [18] is to build distributed representations of words, phrases and sentences as a large dimension of real-valued vectors. After a proper training on the large corpus, the vectors can capture the semantic similarities between each other based on the contextual information in the corpus. Using a large amount of corpus (e.g. the wikipedia text), it can learns relationships between different words by using the context (i.e. the words before and after the input word). Since the sequences of the words always provide some logical meanings, the resulting trained network preserve such logical relations in their the averaging weights. Therefore, this makes the trained vector of words useful for the word analogy task, solving questions such as "man + king − woman = queen". The word2vec model has particularly been widely used for the pre-processing of the large data-sets.

Compared with other conventional representation such as Latent Semantic Analysis (LSA), it automatically preserves the semantic meaning of each paragraph thanks to the distributed word representation, making it useful in various occasions for the natural language processing (NLP) problem. Although there have been quite a few developments based on word embedding, such as sentiment analysis and text classification, we shed a light into another direction of word2vec in the AAL setting. Instead of deriving the embedding structure from the text contexts, a heterogeneous context which comes from different representation can be considered as part of the input experience of the word2vec model too. Although they have different forms of formats (e.g. discrete/continuous, multi-modal distributions, etc), the word2vec would extract based on the contextual information.

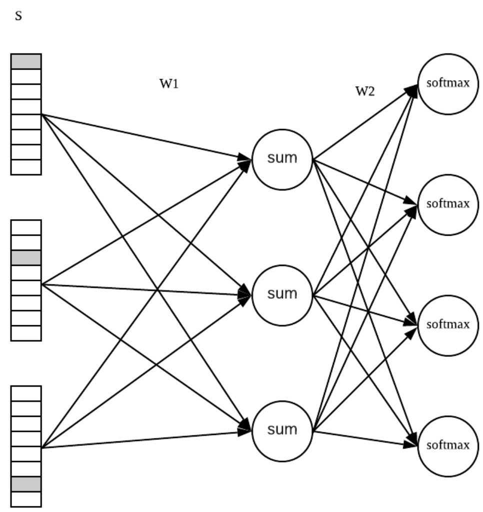

Fig. 1. The Sensor2vec model

B. Learning

Based on the concept of understanding the meaning of words based on the context, the word2vec model works as a shallow, two-layer feed-forward neural networks which is trained to reconstruct linguistic contexts of words. It takes the inputs from the word and its surrounding words. In the context of the AAL setting which employs various kinds of heterogeneous sensors, we propose a similar model called *sensor2vec*. As shown in Fig.1, a distributed semantic representation is used to encode the heterogeneous sensory input as a semantic vector space. Such a space usually has a high-dimension so that the sensory inputs can be learnt to encode in a sparse way. With such an encoding, similar as the usage of word2vec, it can be expected that the heterogeneous sensors which are trigger similarly can be encoded similarly in the embedding space.

Specifically, for the training of the sensor2vec model, assuming that the one-hot vectors are used to encode the sensor signals. The one-hot vector uses one single activation (i.e. 1) in the vector to represent a specific sensory event is triggered in the dataset. Since the rest of the units in the one-hot vector are all 0 except the activated one, it is a common way to represent categorical variables using discrete data but it is usually not the efficient one. Among the heterogeneous sensors, each signal from a single sensor occurs at a continuous time-block can be seen as a 'sensory block'. Note that with some types of sensors, one single sensor may produce different signals. For instance, one single switch sensor can produce either an 'on' or 'off' signal at different but consecutive time-steps. They can be encoded in two sensory blocks. On the other hand, for some of the sensors which produce the continuous readings, the discretized method is used to convert the continuous reading into the corresponding sensory block. As such, one sensory block is represented in one unit of a vector with the length of S, where S is called the total number of unique sensory blocks available in the entire data-set. S corresponds to the number of the words in the vocabulary in the NLP domain. In the heterogeneous sensors domain, the vocabulary here can be understood as a single semantic meaning of the event-related encoding. Besides, in the one-hot encoding, any unit is essentially a binary encoding with the value 1 being in a unique index for each word and the value 0 being in every other index of the vector.

From the one-hot vectors, we can further convert them into the Sensor2vec representation. In this experiment, the CBOW (continuous bag-of-word) method is used to do the training. The steps are as follows:

1) The input layer and the target, both are one- hot encoded of size $[1 \times S]$.
2) Only the linear activation is used between the layers, with the softmax function in the output layer.
3) The training of the sensor2vec model is based on the back-propagation.

Given the input data x, the output y is the softmax function from the hidden layer:

$$y = Softmax(W_2 \cdot W_1 x) \tag{1}$$

where W_1 are the weighting matrix between the input and the hidden layer and W_2 are the weighting matrix between the hidden and the output layer. The function $Softmax(\cdot)$ is used to calculate the normalised probability of selecting neuron i which is defined as

$$Softmax(z) = \frac{e^z}{\sum_{i=0}^{k} e^i} \tag{2}$$

The loss function of the sensor2vec comprises of 2 parts [27]. The first part is the negative of the sum for all the elements in the output layer (before softmax). The second part takes the number of the context words and multiplies the log of sum for all elements (after exponential) in the output layer.

$$L = -\sum_{c=1}^{C} u_{j_c^*} + C \cdot \sum_{j'=1}^{V} exp(u'_j) \tag{3}$$

where j is the index of the actual output in the output layer, u_j is the score for each sensory block, defined as:

$$u_j = {v'_{Wj}}^T \cdot v_{Wi}^T \tag{4}$$

The v is the i-th and v' is the j-th corresponding row and column instead of the whole matrices W_2 and W_1, since the input is an one-hot encoding.

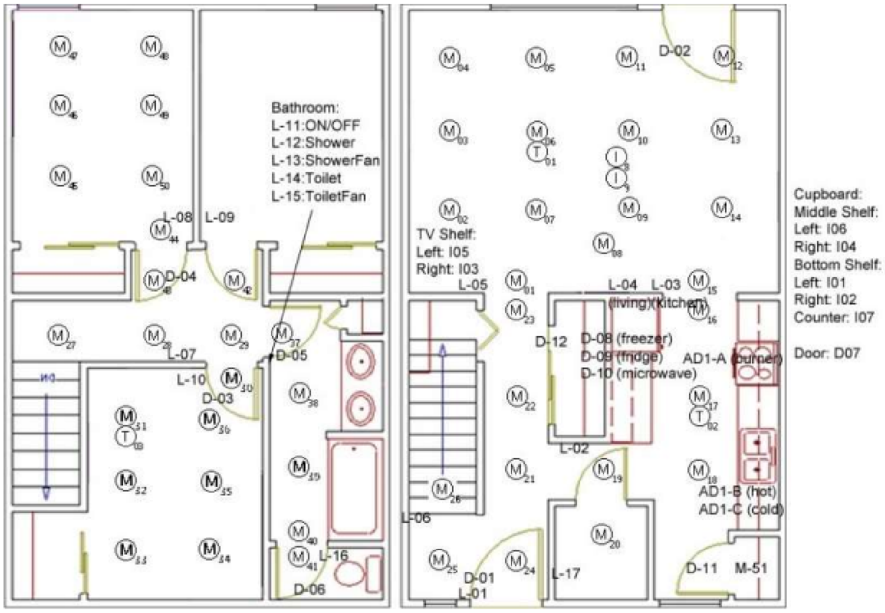

Fig. 2. The Sensors in Kyoto Test-bed (Redrawn from [3])

IV. Experimental Results

A. CASAS Data-sets

In this paper, the CASAS data-set[1] is used to examine our sensor2vec model. The CASAS provides 24 public data-sets describing the daily activities of the participants. Each CASAS data-set is tested in a smart-home setting. In the smart-home, it includes at least one bedroom, includes at least one bedroom, a kitchen, a dining area, and at least one bathroom. While the sizes and layouts of the apartments vary, each home is equipped with combination of heterogeneous sensors: motion/light sensors on the ceilings, combination door/temperature sensors on cabinets, and external doors sensors. Part of the data-set also includes the annotated activities including sleeping, bed-toilet transition, eating, taking medication, cleaning and relaxing.

B. Sensor2Vec Training

The Sensor2vec training is firstly conducted using the Kyoto7 and Kyoto8 data-sets in the CASAS. The Kyoto 7 and 8 datasets recorded the daily lives of two residents R1 and R2 in the test-bed apartment called Kyoto. Its sensor deployment as shown in Fig 2 (motion sensor M, sensor I for the kitchen, door sensor D, temperature sensor T, burner sensor AD1-A, hot water sensor AD1-B, cold water sensor AD1-C and electricity sensors P001).

As we introduced, we firstly train the Kyoto7 and Kyoto8 data-sets separately using the sensor2vec model. The Fig. 3 show the learning curves of the Kyoto8 dataset, from which we can observe that the training converges quite well after the first few epochs.

[1]http://casas.wsu.edu/datasets/

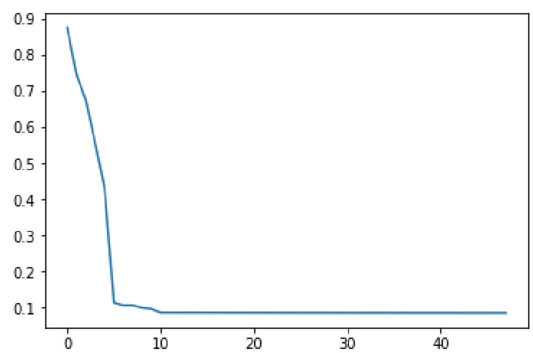

Fig. 3. Training Curve of Sensor2vec with Kyoto8 (RMS Error vs. Iteration)

After that, we tried to visualise the sensory blocks with the tSNE. A few out-liners which exceed a certain threshold of the average values are removed in the visualisation, in order to show the figures in a more clear way. As shown in Fig 4 and Fig. 5, the number of the sensory blocks in the visualisation figures is too large to distinguish. To better compare the results of the sensor2vec, we try to print out the closest sensory block with the cosine distances of certain sensory blocks, as shown in Tab. I and Tab. II. From the tables, we can see that the sensor blocks which are placed together may have a closer distance between each other. The Kyoto8 data-set is more accurate than the Kyoto7 probably because it has a larger amount of data. Nevertheless, a more detailed investigation should be conducted to find the rules of the sensor2vec model in the future work.

In order to test the performance of the sensor2vec learning, we further put the learnt embedded layer with the LSTM network to do the classification. The LSTM network includes

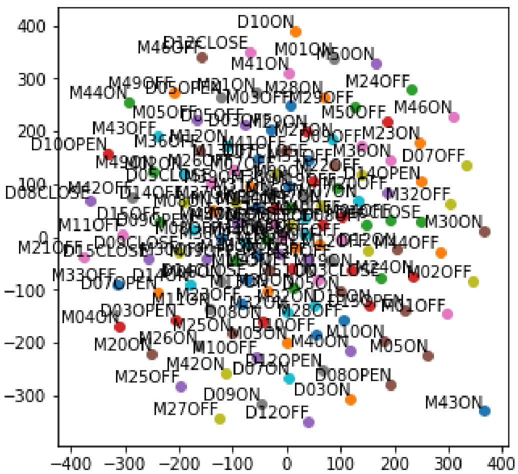

Fig. 4. The tSNE Visualisation of sensor2vec from Kyoto7

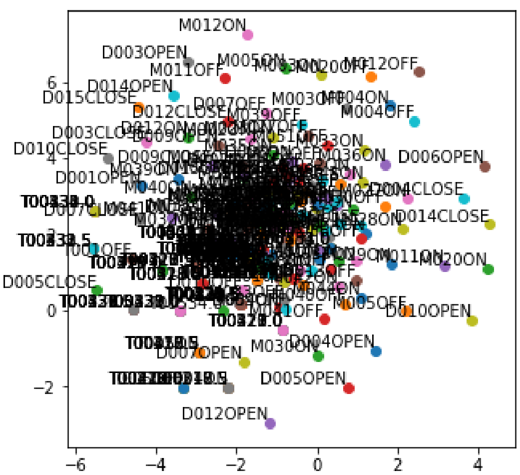

Fig. 5. The tSNE Visualisation of sensor2vec from Kyoto8

2 layers, and 700 hidden units. During training we use 15 batch size, and 0.8 as the drop-out rate. The Tab. III shows the classification accuracy by the two methods: the one-hot input and the sensor2vec inputs. As we can see, using the sensor2vec embedding inputs outperforms the one-hot inputs. But as the trade-off, the sensor2vec + LSTM training need more time than the one-hot input.

V. Conclusion & Future Work

In this paper, from the importance of the contextual information in the AAL, especially for the activity classification, we develop the idea of sensor2vec to deal with the heterogeneous sensory information from the ambient environment. Similar as word2vec in the NLP domain, the sensor2vec use the contextual information and encode the sensory blocks with similar context with the similarity in the embedding layer. Such contextual information facilitates the classification learning of activity using LSTM model.

TABLE I
The Closest Sensory Block with the Selected Sensory Block, trained with Kyoto7

Selected Sensory Block	Closest Sensory Block	Distance
M25ON	M24OFF	4.27
M09ON	M15OFF	7.85
T0126	M01OFF	2.09

TABLE II
The Closest Sensory Block with the Selected Sensory Block, trained with Kyoto8

Selected Sensory Block	Closest Sensory Block	Distance
M23ON	M01ON	1.27
M09ON	M07OFF	0.85
T0126	M23ON	0.44

This paper shows the preliminary result using the sensor2vec training. In the future work, a more systematic study should be conducted to investigate the embedding representation of the sensor2vec, and its relation with the physical meaning in the AAL setting. The implementation and testing in the real-world scenarios would be necessary as well.

References

[1] British Heart Foundation National Centre. Sedentary behaviour–evidence briefing, 2012.

[2] Kaare Christensen, Gabriele Doblhammer, Roland Rau, and James W Vaupel. Ageing populations: the challenges ahead. *The lancet*, 374(9696):1196–1208, 2009.

[3] Diane Cook, M Schmitter-Edgecombe, Aaron Crandall, Chad Sanders, and Brian Thomas. Collecting and disseminating smart home sensor data in the casas project. In *Proceedings of the CHI workshop on developing shared home behavior datasets to advance HCI and ubiquitous computing research*, pages 1–7, 2009.

[4] Diane J Cook, Juan C Augusto, and Vikramaditya R Jakkula. Ambient intelligence: Technologies, applications, and opportunities. *Pervasive and Mobile Computing*, 5(4):277–298, 2009.

[5] Maria Cornacchia, Koray Ozcan, Yu Zheng, and Senem Velipasalar. A survey on activity detection and classification using wearable sensors. *IEEE Sensors Journal*, 17(2):386–403, 2016.

[6] Alessia D'Andrea, Arianna D'Ulizia, Fernando Ferri, and Patrizia Grifoni. A multimodal pervasive framework for ambient assisted living. In *Proceedings of the 2nd International Conference on PErvasive Technologies Related to Assistive Environments*, pages 1–8, 2009.

[7] Irfan A Essa. Ubiquitous sensing for smart and aware environments. *IEEE personal communications*, 7(5):47–49, 2000.

[8] Piyush Gupta and Tim Dallas. Feature selection and activity recognition system using a single triaxial accelerometer. *IEEE Transactions on Biomedical Engineering*, 61(6):1780–1786, 2014.

[9] Rida Ghafoor Hussain, Mustansar Ali Ghazanfar, Muhammad Awais Azam, Usman Naeem, and Shafiq Ur Rehman. A performance comparison of machine learning classification approaches for robust activity of daily living recognition. *Artificial Intelligence Review*, 52(1):357–379, 2019.

[10] Ahmad Jalal, Shaharyar Kamal, and Daijin Kim. Human depth sensors-based activity recognition using spatiotemporal features and hidden markov model for smart environments. *Journal of computer networks and communications*, 2016, 2016.

TABLE III
Accuracy of sensor2vec and one-hot inputs in LSTM classification

	Sensor2vec + LSTM	One-hot + LSTM
Accuracy	0.8723	0.7642

[11] Thomas B Kearns. Technology and the right to privacy: the convergence of surveillance and information privacy concerns. *Wm. & Mary Bill Rts. J.*, 7:975, 1998.

[12] Youngwook Kim and Taesup Moon. Human detection and activity classification based on micro-doppler signatures using deep convolutional neural networks. *IEEE geoscience and remote sensing letters*, 13(1):8–12, 2015.

[13] Mandira GA Komang, Michrandi N Surya, and Astuti N Ratna. Human activity recognition using skeleton data and support vector machine. In *Journal of Physics: Conference Series*, volume 1192, page 012044. IOP Publishing, 2019.

[14] Oscar D Lara and Miguel A Labrador. A survey on human activity recognition using wearable sensors. *IEEE communications surveys & tutorials*, 15(3):1192–1209, 2012.

[15] Jie Li, Chenguang Yang, Junpei Zhong, and Shilu Dai. Emotion-aroused human behaviors perception using rnnpb. In *2018 10th International Conference on Modelling, Identification and Control (ICMIC)*, pages 1–6. IEEE, 2018.

[16] Xiaomu Luo, Qiuju Guan, Huoyuan Tan, Liwen Gao, Zhengfei Wang, and Xiaoyan Luo. Simultaneous indoor tracking and activity recognition using pyroelectric infrared sensors. *Sensors*, 17(8):1738, 2017.

[17] Amira Ben Mabrouk and Ezzeddine Zagrouba. Abnormal behavior recognition for intelligent video surveillance systems: A review. *Expert Systems with Applications*, 91:480–491, 2018.

[18] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[19] Ram Nevatia, Tao Zhao, and Somboon Hongeng. Hierarchical language-based representation of events in video streams. In *2003 Conference on Computer Vision and Pattern Recognition Workshop*, volume 4, pages 39–39. IEEE, 2003.

[20] Liangying Peng, Ling Chen, Xiaojie Wu, Haodong Guo, and Gencai Chen. Hierarchical complex activity representation and recognition using topic model and classifier level fusion. *IEEE Transactions on Biomedical Engineering*, 64(6):1369–1379, 2016.

[21] Minh Pham, Dan Yang, and Weihua Sheng. A sensor fusion approach to indoor human localization based on environmental and wearable sensors. *IEEE Transactions on Automation Science and Engineering*, 16(1):339–350, 2018.

[22] Martha E Pollack. Intelligent technology for an aging population: The use of ai to assist elders with cognitive impairment. *AI magazine*, 26(2):9–9, 2005.

[23] Sreenivasan Ramasamy Ramamurthy and Nirmalya Roy. Recent trends in machine learning for human activity recognition—a survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1254, 2018.

[24] Suneth Ranasinghe, Fadi Al Machot, and Heinrich C Mayr. A review on applications of activity recognition systems with regard to performance and evaluation. *International Journal of Distributed Sensor Networks*, 12(8):1550147716665520, 2016.

[25] Parisa Rashidi and Alex Mihailidis. A survey on ambient-assisted living tools for older adults. *IEEE journal of biomedical and health informatics*, 17(3):579–590, 2012.

[26] Daniele Riboni and Claudio Bettini. Cosar: hybrid reasoning for context-aware activity recognition. *Personal and Ubiquitous Computing*, 15(3):271–289, 2011.

[27] Xin Rong. word2vec parameter learning explained. *arXiv preprint arXiv:1411.2738*, 2014.

[28] Michael S Ryoo and Jake K Aggarwal. Recognition of composite human activities through context-free grammar based representation. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1709–1718. IEEE, 2006.

[29] Miriam Sander, Bjarke Oxlund, Astrid Jespersen, Allan Krasnik, Erik Lykke Mortensen, Rudi Gerardus Johannes Westendorp, and Lene Juel Rasmussen. The challenges of human population ageing. *Age and ageing*, 44(2):185–187, 2015.

[30] Muhammad Hameed Siddiqi, Madallah Alruwaili, Amjad Ali, Saad Alanazi, and Furkh Zeshan. Human activity recognition using gaussian mixture hidden conditional random fields. *Computational Intelligence and Neuroscience*, 2019, 2019.

[31] Odongo Steven Eyobu and Dong Seog Han. Feature representation and data augmentation for human activity classification based on wearable imu sensor data using a deep lstm neural network. *Sensors*, 18(9):2892, 2018.

[32] Marcin Stonawski, Vegard Skirbekk, Conrad Hackett, Michaela Potančoková, Phillip Connor, and Brian Grim. Global population projections by religion: 2010–2050. In *Yearbook of International Religious Demography 2015*, pages 99–116. Brill, 2015.

[33] Sheikh Badar ud din Tahir, Ahmad Jalal, and Mouazma Batool. Wearable sensors for activity analysis using smo-based random forest over smart home and sports datasets. In *2020 3rd International Conference on Advancements in Computational Sciences (ICACS)*, pages 1–6. IEEE, 2020.

[34] Eftim Zdravevski, Petre Lameski, Vladimir Trajkovik, Ivan Chorbev, Rossitza Goleva, Nuno Pombo, and Nuno M Garcia. Automation in systematic, scoping and rapid reviews by an nlp toolkit: a case study in enhanced living environments. In *Enhanced Living Environments*, pages 1–18. Springer, 2019.

[35] Licheng Zhang, Xihong Wu, and Dingsheng Luo. Improving activity recognition with context information. In *2015 IEEE international conference on mechatronics and automation (ICMA)*, pages 1241–1246. IEEE, 2015.

[36] Junpei Zhong and Lola Canamero. From continuous affective space to continuous expression space: Non-verbal behaviour recognition and generation. In *4th International Conference on Development and Learning and on Epigenetic Robotics*, pages 75–80. IEEE, 2014.

[37] Junpei Zhong, Rony Novianto, Mingjun Dai, Xinzheng Zhang, and Angelo Cangelosi. A hierarchical emotion regulated sensorimotor model: case studies. In *2016 Chinese Control and Decision Conference (CCDC)*, pages 4965–4970. IEEE, 2016.