Activity Prediction for Improving Well-Being of Both The Elderly and Caregivers

Yuting Zhan, Hamed Haddadi
Imperial College London

ABSTRACT

The issue of ageing population is gaining significant attention across the world, while the caregivers' psychological burden caused by a variety of geriatric symptoms is often overlooked. Efficient collaboration between the elderly and caregivers has great potential to relieve the caregivers' psychological burden and improve the caregiving quality. For instance, activity prediction can provide a promising approach to cultivate this efficient collaboration. Given the ability to predict the elderly patients' activity and its timing, caregivers can provide timely and appropriate care, which not only can relieve caregiving stress for professional or family caregivers, but also can reduce the unwanted conflicts between both parties. In this paper, we train an activity predictor by integrating the activity temporal information into the *Long Short-Term Memory* (LSTM) networks. The approach leads to significant improvements in the prediction accuracy both in the next activity and its precise occurrence time.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; *Interactive systems and tools*; • **Applied computing** → Health care information systems.

KEYWORDS

Recurrent Neural Networks, Activity Recognition, Activity Prediction, Long-term Healthcare, Smart Home

ACM Reference Format:
Yuting Zhan, Hamed Haddadi. 2019. Activity Prediction for Improving Well-Being of Both The Elderly and Caregivers. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2019 International Symposium on Wearable Computers (UbiComp/ISWC '19 Adjunct), September 9–13, 2019, London, United Kingdom.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3341162.3344834

ACM ISBN 978-1-4503-6869-8/19/09…$15.00
https://doi.org/10.1145/3341162.3344834

1 INTRODUCTION

The topic of ageing population has emerged as one of the most formidable socioeconomic challenges faced by both developed and developing countries. The enormous quantity of elderly population increases the economic burden for the cosmopolitan well-being system, which also poses a critical challenge to long-term healthcare. In Europe, the number of people aged 65 and over is expected to grow to 28% of the population in 2060 [2]. The demographics of the elderly group afflicted by Alzheimer disease and related dementia is proportional to the growing trend of the ageing population. By 2030, $2 trillion dollars will be spent each year globally for the well-being of dementia group [6].

Integrating the assistive technologies into the smart home paradigm [4] is regarded as a potential trend for setting up care for different elderly groups, especially for the candidates with dementia, and other mild cognitively or functionally impaired elders. This approach can reduce time and cost expenditures of the long-term healthcare, while improving the quality of life for the target population and their caregivers. However, in current research, caregiver distress correlated with multifarious geriatric symptoms is overlooked, where there is a significant gap in relieving the psychological burden for caregivers [4]. For instance, the dementia group has a series of symptoms including but not limited to agitation, irritability, depression, delusion and etc.. Those symptoms torture and distress their formal or informal caregivers [4]. Caregiving with depression and frustration is detrimental to the wellness of both parties, which might escalate the tension and intensify potential conflicts between them [6].

One promising way to mitigate these negative effects is to implement efficient non-verbal communication, or collaboration, between the elderly group and their caregivers. Research by Koumakis et al. has emphasized the demand for cultivating an appropriate collaboration between those suffering from mild dementia and their caregivers [4].

The ability to model the elderly behaviors and predict their next activity is extremely valuable in cultivating this particular collaboration and smoothing their communication with each other. However, modelling human behaviors and

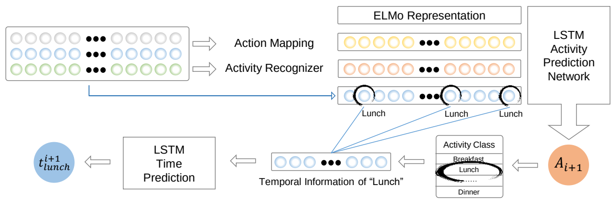

Figure 1: Overview of the proposed prediction architecture.

predicting human next activity are highly challenging due to the complexity of human behaviors. Though human behaviors and activities are hard to model, they can be represented by a time series of periodic, repetitive, and interdependent sequence data, as human is a creature of habit. These properties allow the sequence of human activity to be a predictable sequence. Especially for the elderly group and the candidates of dementia, their lives are much more monotonous and repetitive.

In our proposed framework, by leveraging the indoor activity recognizer [3], an elder's activity can be recognized, recorded and further analyzed. The results of such activity recognition are then fed into the activity predictor, which can predict the next activity or intention of the group that is taken care of. The persuasive prediction itself can relive the tension resulted in discrete wills between two groups. For the long-term healthcare, successfully predicting the elder's next activity or intention can call caregivers' attention to helping them when needed, which not only can relive the caregiving burden and improve efficiency, but also improve the quality of life for both groups.

There is no doubt that activity prediction plays an important role in this particular communication between the elderly and their caregivers, which communication emphasizes the prediction performance. Hence, this work trains a more persuasive activity predictor by using the *deep contextualized word representation* (ELMo) [5] and integrating the temporal information of the data into the proposed LSTM framework. Our results highlight that an appropriate utility of the temporal information can have better prediction accuracy. While better predictions in our framework allow the caregivers to have more personal time and relive their psychological burden. With persuasive predictions, both groups' well-being are taken into consideration in their daily life.

2 PROPOSED ARCHITECTURE

Unlike some previous healthcare platforms driven by rule-based reminders, the proposed architecture in this work is sensor data-driven, autonomous and proactive. The overview of proposed prediction architecture is shown in Figure 1. Data is emanated from the ubiquitous sensors, and further labelled

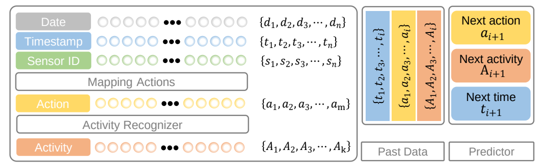

Figure 2: Dataset description. Each line in the dataset at least contains 'Date', Timestamp' and Sensor ID'.

as the sequences of activities. These labels are then represented as ELMo vectors [5] and fed into the LSTM model to train the activity predictor. At the same time, each activity class has its own LSTM-based time predictor. When the prediction of the next activity is generated, the corresponding time predictor will be triggered and predict the precise occurrence period, as shown in Figure 1. With this valuable information, caregivers can estimate when is the right time to provide appropriate help to the elderly.

2.1 Dataset Description

Our models use the sensor data instead of cutting-edge wireless signal data, as existing wireless signal-based activity recognition methods have lower accuracy and can only recognize one or several activities. However, the quality of prediction depends on the diversity of the recognized activities. We compare three widely-used datasets in activity recognition literature, that were published by Cook and Krishnan [3], Tapia et al. [7], and Van Kasteren et al. [8], respectively. All datasets chosen to validate our assumptions are single-person apartment monitoring data that emanated from multiple sensors. Each line of data should at least contain "date", "timestamp", "sensor ID", and "activity label", as shown in Figure 2. Formally, one timestamp t_i recorded as a result of one sensor trigger by resident, one sensor trigger is mapped as one action a_i. Activity recognizer will generate the sequence of activities by analyzing the sensor data.

The prediction task can be formulated as a variant of the sequence generation task: given a sequence of resident's past actions $\{a_1, a_2, \ldots, a_i\}$ or past activities $\{A_1, A_2, \ldots, A_i\}$ in concerted with a sequence of timestamps $\{t_1, t_2, \ldots, t_i\}$ until time t_i, to predict the next action a_{i+1} or next activity A_{i+1} and their occurring time t_{i+1}.

2.2 ELMo Representation

The prominent word representation in deep learning area is word embedding, and one of the most frequently-used word embeddings is *Word2Vec* embedding. However, the Word2Vec embedding and other similar word embeddings are now losing their dominance in *Natural Language Processing* (NLP) area due to the rise in availability of novel pre-trained language models. ELMo [5], the abbreviation of *Embedding from Language Models*, is a *deep contextualized word representation* which has shown the potential to improve the state-of-the-art performance of existing NLP tasks. Unlike other word vectors, the ELMo vector can be learned by a deep bidirectional language model pre-trained on a large text corpus [5]. Hence, ELMo representation can better model both semantic complexity and context-based polysemy. In our proposed model (Figure 1), we use ELMo representation to define each action's embedding matrix, a 128-dimension vector, in the embedding layer of the LSTM model.

2.3 Long Short-Term Memory Network

Long Short-Term Memory units, or LSTMs, proposed by Hochreiter and Schmidhuber in 1997, is a prominent variant of *recurrent neural networks* (RNN). LSTM has been shown to exhibit brilliant performance on modelling entire sequences of data, especially for linking remote causes and effects in time series data. LSTM can efficiently handle the difficulty of learning long-term dependencies with the gradient descent in a standard RNN. The capability of LSTM on remote dependencies empowers it to be one of the dominant networks in time-series-data analysis and sequence generation.

In this work, the LSTM network is well-suited to our sensor data, which is sparse, time-varying and interdependent. In order to verify our hypothesis that the LSTM models take into account of the temporal information of sensor data will have better prediction performance than the one without this information, we have two LSTM models in comparison. The first model has only one LSTM layer for action embedding, while the other one has an extra LSTM layer to integrate time-stamps. These two models are referred to *One-LSTM* and *Temporal-LSTM*, respectively.

3 EVALUATION AND RESULTS

Three different experiments are set up to evaluate the performance of the proposed architectures with integrated temporal information. These architectures' performance are compared with a baseline, which used *Word2Vec* embedding for action representation and LSTM-based network for human behavior modelling [1]. Prediction accuracy is one of the most important features for assessing performance of the predictive model, thereby we use the prediction accuracy as the evaluation metric, from one-attempt to five-attempt, which would keep horizontal comparison with the baseline.

3.1 Varying Context Size

Different lengths of the context size exert an effect on prediction performance of a recurrent model [3]. By changing the length of the input action (context size) from 1 to 200, we observe that the length of input action has a high impact

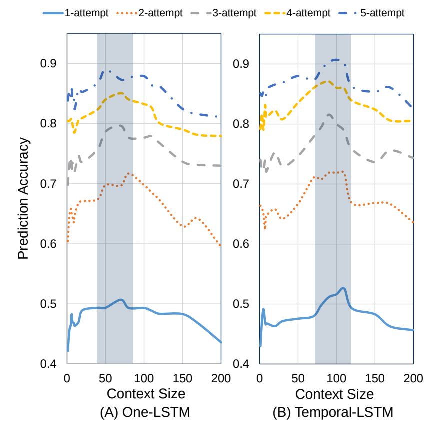

Figure 3: Prediction accuracy changes with the length of input action (context size) in two LSTM models. Grey bins illustrate the optimal area of each model.

on the accuracy in our two LSTM models, as shown in Figure 3. Note that the optimal context size of the first model is situated at the interval [40, 90], while the second model is at the interval [70, 120].

Compared with One-LSTM model, the Temporal-LSTM model has better performance from all 1-attempt to 5-attempt predictions, especially for 1-attempt and 5-attempt, where the accuracy is higher than 0.5 and 0.9, respectively. The comparison between these two models also illustrates that even with the same dataset, different architectures would have different optimal values. When considering the temporal information, sufficient context improves performance while insufficient context decreases the performance.

3.2 ELMo Embedding or Word2Vec Embedding

In the baseline experiment, Word2Vec embedding representation has been used in the embedding layer of LSTM to provide better performance than one-hot vectors[1]. In our proposed models, ELMo representation is used to define each action's embedding matrix. The pre-trained language model generates a 128-dimension vector for each action. Then these vectors are fed into the embedding layer of our LSTM models.

In this contrast experiment, the ELMo embedding-based models are compared with the Word2Vec embedding-based models, as shown in Figure 4. The results illustrate that the ELMo representation can improve the prediction accuracy, where the increment is 5.93% averagely for the One-LSTM(1LSTM), and 8.57% for the Temporal-LSTM(2LSTM). With ELMo representation, the integration of time information (2LSTM) can also improve the performance, which demonstrates that ELMo representation has more potential to illustrate the contextual-temporal dynamics in activity prediction.

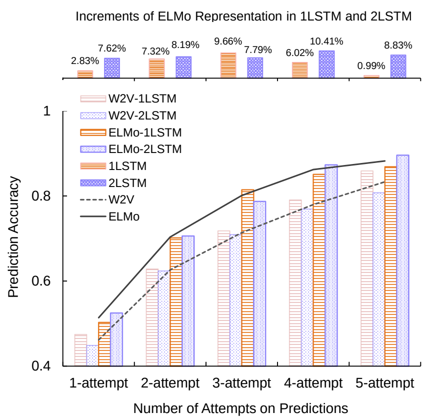

Figure 4: Improvement of ELMo representation when compared to Word2Vec embedding. The 2LSTM with ELMo representation outperforms the 1LSTM.

3.3 One-LSTM or Temporal-LSTM

In the baseline study[1], Almeida and Azkune evaluated three different fusion strategies on LSTM network, and found that all fusion strategies that took into account of timestamps $\{t_1, t_2, \ldots, t_i\}$ were detrimental to the results. Their data is shown in the *w2v* line in the Table 1, where Temporal-LSTM (2LSTM) has worse performance.

On the contrary, in our proposed LSTM models, the results in Table 1 admit that the time layer can improve the accuracy of prediction, especially for the increments at 1-attempt(top1) and 5-attempt(top5) situations. In general, the Temporal-LSTM (2LSTM) has better prediction accuracy than One-LSTM (1LSTM). We demonstrate that if temporal information can be used in an appropriate way, the model can have better prediction results, especially for a small dataset.

4 DISCUSSIONS

The valuable information provided by activity prediction can build an efficient collaboration between the elderly and caregivers, which is a cornerstone of the better quality of life for both groups. In addition to effective communication between them, in the long-term healthcare, successfully predicting an elder's intention or next activity can provide a series of activity-aware services, including penalization of the intelligent environment, prompting-based intervention, anomaly detection, and etc. There is still a large potential improvement in prediction accuracy. A truly persuasive prediction can undoubtedly improve the elder's quality of life and well-being, so as their caregivers.

In real-time prediction settings, a well-trained activities recognizer that recognizes the activities as inputs for the predictor. At the same time, the recognizer would also provide a baseline for the predictor to verify each prediction. However,

Table 1: Prediction accuracy of two ELMo-embedded LSTM models when context size is 50 and 70, respectively, are compared with the baseline (w2v).

		top1	top2	top3	top4	top5
w2v	1LSTM	0.4744	0.6282	0.7179	0.7905	0.8589
	2LSTM	0.4487	0.6239	0.7094	0.7692	0.8076
50	1LSTM	0.4844	0.6933	0.7822	0.84	0.8577
	2LSTM	**0.5112**	0.6712	0.7734	**0.84**	**0.8712**
70	1LSTM	0.5027	0.7014	0.8145	0.8507	0.8688
	2LSTM	**0.5249**	0.6787	0.7828	**0.8733**	**0.8869**

there are three challenges: firstly, current indoor activity recognizer utilized in predictor, depends mostly on ambient sensors which cannot recognize micro actions and realize continuous tracking; secondly, activity predictor (LSTM-based, which has better performance than other baselines, like Markov Model) still has unexpected low accuracy and low time sensitivity; thirdly, the multi-person scenario is always challenging and need feasible solutions.

ACKNOWLEDGMENTS

The authors are grateful to Imperial College London and the China Scholarship Council (CSC) for financial support. Hamed Haddadi was supported partially by the UK-Dementia Research Institute.

REFERENCES

[1] Aitor Almeida and Gorka Azkune. 2018. Predicting human behaviour with recurrent neural networks. *Applied Sciences* 8, 2 (2018), 305.

[2] European Commission. 2015. The 2015 ageing report: Economic and budgetary projections for the 28 EU member states (2013–2060). *European Economy 3/2015 Financial Affairs, Brussels.* (2015).

[3] Diane J Cook and Narayanan C Krishnan. 2015. *Activity learning: discovering, recognizing, and predicting human behavior from sensor data.* John Wiley & Sons.

[4] Lefteris Koumakis, Charikleia Chatzaki, Eleni Kazantzaki, Evangelia Maniadi, and Manolis Tsiknakis. 2019. Dementia Care Frameworks and Assistive Technologies for Their Implementation: A Review. *IEEE reviews in biomedical engineering* 12 (2019), 4–18.

[5] Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365* (2018).

[6] Martin James Prince. 2015. *World Alzheimer Report 2015: the global impact of dementia: an analysis of prevalence, incidence, cost and trends.* Alzheimer's Disease International.

[7] Emmanuel Munguia Tapia, Natalia Marmasse, Stephen S Intille, and Kent Larson. 2004. MITes: Wireless portable sensors for studying behavior. In *Proceedings of Extended Abstracts Ubicomp 2004: Ubiquitous Computing.*

[8] Tim Van Kasteren, Athanasios Noulas, Gwenn Englebienne, and Ben Kröse. 2008. Accurate activity recognition in a home setting. In *Proceedings of the 10th international conference on Ubiquitous computing.* ACM, 1–9.