Get Up!: Assessing Postural Activity & Transitions using Bi-Directional Gated Recurrent Units (Bi-GRUs) on Smartphone Motion Data

Kavin Chandrasekaran, Luke Buquicchio, Walter Gerych, Emmanuel Agu and Elke Rundensteiner
Dept. of Computer Science, Worcester Polytechnic Institute, Worcester, USA
Email: {kchandrasekaran, ljbuquicchio, wgerych, emmanuel, rundenst}@wpi.edu

***Abstract*—Many health conditions can affect a person's mobility. Consequently, a person's ability to perform transitions between activity states (e.g. sit-to-stand) are accurate measures of their mobility and general health. Mobility impairments can manifest either as discomfort while performing certain activity transitions or a complete inability to perform such transitions. The Timed up and Go (TUG) is an important clinical test that assesses patients' sit-to-stand abilities. Research into passive methods to assess the quality of patients activity transitions and thus conduct the Timed Up and Go autonomously as they live their lives, have recently become popular. Machine and deep learning analysis of smartphone accelerometer and gyroscope data have demonstrated promising activity and transition recognition results. In this paper, we present *Get Up!*, a novel deep learning-based method to detect whether a person is performing a certain postural activity or transitioning between activities. *Get Up!* analyzes data from the accelerometer and gyroscope of the patient's smartphone using Bi-Directional Gated Recurrent Units (Bi-GRU) neural networks with an attention mechanism. Our method outperforms TAHAR, the current state of the art machine learning method, achieving an error rate of 1.47% for activity classification and an accuracy of 97%. We also achieved an error rate of 0.17% with an accuracy of 93.3% when classifying postural transitions. As *Get Up!* segments activities and transitions, individual TUG sub-components can be timed to identify sub-components that patients find challenging.**

***Index Terms*—activity recognition, activity transitions, bi-directional GRU**

I. Introduction

A. The Problem

Many illnesses and disorders have symptoms that manifest physically, affecting a person's balance and range of motion. For example, a person with Parkinson's might have tremors and muscle rigidity that will affect their mobility. Patients' physical activities can be categorized either as postural activities or dynamic activities. Postural activities such as sitting, lying down and standing, and longer dynamic activities such as walking and running, have been widely studied using machine learning methods [1] [2] [3]. However, systems to passively detect and assess shorter dynamic activities such as falling, and transitions between different activities that provide valuable insights into a patient's condition, are much less studied.

The Timed Up and Go (TUG) is a popular test to evaluate postural transitions [4], which measures the time taken by a patient to stand up from sitting, walk around and return to sitting position. TUG is a well established clinical test that reliably quantifies a patient's balance and risk of falls. One disadvantages of the TUG is that it is currently administered manually, which introduces errors. Also, it only measures the total time taken for the test, but not the individual sub-components of the TUG test (Sit-to-Stand, Walking, Turning 180°, Walking, Turn and Stand-to-Sit). Patients can have varying levels of difficulty in performing each TUG sub-component and this information is lost in the traditional TUG test. For example, a patient might have trouble standing up from sitting but is fine with walking. This patient's total time taken for the test will not reveal their difficulty with standing up, especially as the duration of the transition is small.

A new instrumented TUG (iTUG) test has been introduced [5] which provides the sub-component times along with other parameters making the test more robust. Prior work by Milosevic *et al* has automated the iTUG test such that it can be performed using a smartphone attached to a person's sternum [6]. In order to better quantify the sub-components, we need to reliably identify the transitions indicating the beginning and end of sub-components in the data. Hence, detecting transitions is an important problem with ramifications in the performance of the automated iTUG tests.

Human Activity Recognition (HAR) is a widely studied problem that tries to detect the user's current activity, which plays an important role in various applications such as: health monitoring, security monitoring, and human computer interaction. HAR using smartphones has been proposed as a way to continuously assess and monitor the health of smartphone users. In such HAR systems, sensor data gathered from users smartphones is analyzed using Machine Learning (ML) or Deep Learning (DL) techniques. Modern smartphones come equipped with a plethora of sensors and are incorporated into most people's daily lives. Smartphone HAR is unobtrusive as it utilizes sensors built into the phone and thus requires no additional sensors to be worn or configured. Due to their ubiquity, smartphones are effective and efficient for passive data collection for HAR. Accelerometers are the most popular sensors to perform activity recognition [7], while using gyroscopes have been shown to improve the recognition accuracy [8]. A multitude of methods have shown that smartphone user activities can be classified reasonably accurately using

This work was funded by DARPA.

smartphone sensor data [9] [10].

In healthcare, Human Activity Recognition (HAR) systems can be used for monitoring and assessing the quality of patients' activities, detecting falls [11], symptoms of stroke [12] and changes in their gait patterns [13]. Identifying short-duration events and activity transitions can be modelled as a classification or an anomaly detection problem. Since detecting a transition is a result of identifying abrupt changes within patterns in a time series data, Change Point Detection (CPD) techniques have also been proposed to detect activity transition in smart home environments [14]. Detecting activity transitions have also been proposed to be leveraged to better classify various activities, as the transitions provide a better understanding of the activity boundaries [15].

Prior work by Reyes-Ortiz *et al* integrates both HAR and the detection of activity transitions in the Transition-Aware Human Activity Recognition (TAHAR) system using smartphones. TAHAR classifies the smartphone's accelerometer and gyroscope data in real-time. It combines outputting the sequence of probabilistic consecutive activity predictions of a Support Vector Machine (SVM) with a heuristic filtering approach.

In this paper, we propose *Get Up!*, a novel deep learning-based method to detect smartphone user activities and activity transitions using Bi-directional Gated Recurrent Neural Networks (Bi-GRUs). Neural networks have recently demonstrated superior performance to Machine Learning (ML) techniques on hand-crafted features in various tasks including image classification, audio analysis as well as HAR. The Bi-directional GRU (Bi-GRU) is suitable for analyzing future and previous time-series data such as the smartphone's accelerometer and gyroscope data. We also include an attention layer, which enables our classification model to identify specific regions in the time series to focus on. *Get Up!* outperformed the state-of-the-art TAHAR approach by achieving an error rate of 1.47% and accuracy of 97% for activity classification and an error rate of 0.17% with accuracy of 93.3% in classifying the postural transitions. As *Get Up!* segments activities and transitions accurately, individual sub-components can be timed to identify sub-components that patients find challenging. Our *Get Up!* approach is envisioned as an offline system that analyzes patients' smartphone accelerometer and gyroscope data to provide end-of-day evaluations of a patient to provide evidence for their treatment.

II. Our Approach

A. Architecture

Our *Get Up!* approach aims to perform classification on smartphone sensor data that is continuously gathered and analyzed while the user performs their daily activities.

Activity and Transitions Dataset: We utilize the same smartphone dataset utilized by Reyes-Ortiz *et al* to create the TAHAR architecture, which facilitates direct comparison of the efficacy of both approaches. The dataset contains 12 labels with 6 activities (Walking, Walking Upstairs, Walking Downstairs, Sitting, Standing and Lying Down) and 6 postural transitions (Stand-to-Sit, Sit-to-Stand, Sit-to-Lie, Lie-to-Sit, Stand-to-Lie and Lie-to-Stand). We take a two-step approach to classify both the activities and the transitions. The overview of the architecture can be found in Fig.1. We begin with sensor data collected from the smartphone accelerometer and gyroscope. The data is then sent to the feature extractor, which creates a feature vector of size 561. The featurized time-series data is then passed on to the next classifier.

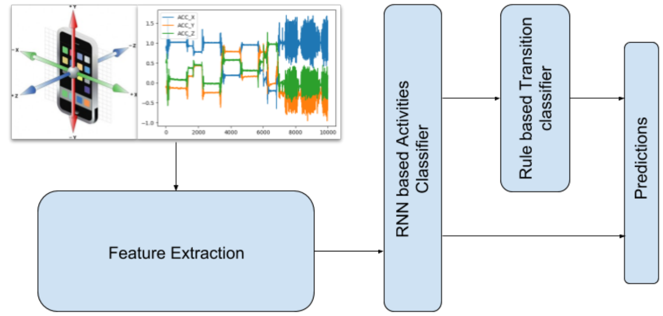

Fig. 1: Architecture Overview

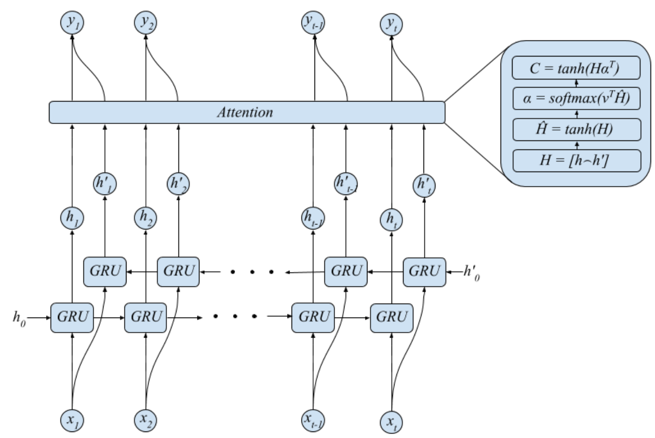

Fig. 2: Bi-GRU Architecture

Recurrent neural network architectures are suitable for classifying time series data. We designed a modified version of a GRU model combined with a self-attention layer to achieve better performance. Since we are proposing an offline method, we use a Bi-directional GRU(Bi-GRU), which can make use of the information from past instances and the future instances. The attention layer enables the model identify specific regions in the time series to focus on. The attention mechanism is a feed-foward layer that takes the hidden layer outputs from the Bi-GRU and outputs a transformed representation that can be used for classification [16]. The Bi-GRU architecture with attention mechanism is shown in Fig.2. The featurized time series input (x_1, x_2,...,x_T) is received from the feature extractor. x_i is fed into the GRU layer. Since we have a bi-directional GRU, there are essentially two GRU layers (forwards and backwards), with one layer processing the inputs from the beginning(x_1 *to* x_T) and the other layer processing the input backwards(x_T *to* x_1). Both the forward and backward GRUs outputs a hidden layer representation, h and h'. The hidden layer representations are passed on to the attention layer.

$$H_i = [h_i \frown h'_i] \tag{1}$$

For each time step i, the hidden representations from both forward and backward are concatenated (1) to generate a vector H_i. H_i across all time steps are stacked to yield a matrix representation H.

$$\widehat{H} = tanh(H) \tag{2}$$

The attention weight vector(v) is initialized with random weights that was learned through the back-propagation. Since the attention model is trying to find the best positions in the time-series to focus on, we need a function to quantify how much each hidden representation should be considered for the output. This is called *the alignment score function*, which is used to calculate the alignment score α (3). The context vector is a weighted sum of the hidden representation $\widehat{H}$ and the alignment scores (2) (3) (4).

$$\alpha = softmax(v^{\mathrm{T}}\widehat{H}) \tag{3}$$

$$C = H\alpha^T \tag{4}$$

We generate the attention vector a that is passed to the feed-forward layer for final classification using softmax (6). The prediction assigns each instance to one of seven classes (6 activity classes and one binary class to indicate transition).

$$a = tanh(C) \tag{5}$$

$$\widehat{y} = argmax(softmax(Wa + b)) \tag{6}$$

If the instance is classified as one of the activities, we keep the prediction but if it was predicted to be a transition, we pass the information to a rule-based transition classifier. Our architecture is modular, so that we could use any other classification model to replace the activity classifier and still continue with the transition classifier.

The rule-based classifier takes into account the predictions of the preceding and succeeding instances of the transition and assigns the instance to a specific activity transition class. For Example, if an instance at t is classified to be in transition by the GRU classifier, and the previous instance($t-1$) is classified as Sitting and the next instance($t+1$) is classified as Standing, then the instance t is assigned the class Sit-to-Stand.

III. Experiments and Results

A. TAHAR Activity and Transition Dataset

The data was collected from a controlled study with 30 participants. The participants had a smartphone mounted on their waist and performed six activities; 3 static postures(Sitting, Standing, Lying) and 3 dynamic activities(Walking, Walking upstairs and Walking downstairs). Data from the smartphone's 3-axial accelerometer and gyroscope were captured. The raw sensor data was sampled with a fixed-length sliding window and a feature vector of 561 values was generated by calculating numerous time and frequency domain variables. Each sample is accompanied by one of 12 labels (Walking, Walking Upstairs, Walking Downstairs, Sitting, Standing, Lying Down, Stand-to-Sit, Sit-to-Stand, Sit-to-Lie, Lie-to-Sit, Stand-to-Lie and Lie-to-Stand), which was labelled manually by going through the video recording of the study.

B. Experimental setup

We split the dataset into 3 parts: training set, validation set and test set consisting of 60%, 20%, 20% of the total number of users, respectively. We used the validation set to perform a grid search to find the best hyper-parameter combination. We tuned our model to find the best learning rate, hidden layer size and the number of epochs. We also evaluated the impact of the number of directions in the GRU model and the presence of the attention layer. After tuning the parameters, we ran the model on the test set and the results, averaged across 5-folds using k-fold cross-validation, are presented below. Since there are fewer instances of the transitions compared to the activities, data imbalance occurs. To address the class-imbalance, we define the accuracy as $Accuracy = (Sensitivity + Specificity)/2$. It is observed that the training loss with the attention mechanism helps to improve the performance of the model. The attention weights after learning to classify the activities can be seen in Fig. 3. It shows the regions on which the model has learned to focus, during the first 100 timesteps.

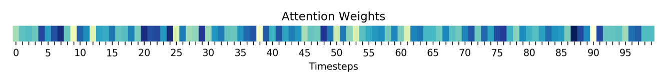

Fig. 3: Attention Weights

We used the rule-based classifier to classify activity transitions. Table I shows the rules we used.

Transition Classification rules		
Preceding Activity	Succeeding Activity	Transition Label
Sitting	Standing	Sit-to-Stand
Standing,	Sitting	Stand-to-Sit
Lying Down	Standing	Lie-to-Stand
Standing	Lying Down	Stand-to-Lie
Lying Down	Sitting	Lie-to-Sit
Sitting	Lying Down	Sit-to-Lie
[1]Lying down	Walking	Lie-to-Stand

TABLE I: Activity transitions classification rules

While using the rule-based classifier, we observed that our model was not predicting Lie-to-Stand as effectively as the other transitions. We noticed in the data that there were activity transitions labelled as Lie-to-Stand but the activities before and after were Lying Down and Walking respectively as shown in Fig. 4.

We added an exception in our rule-based to consider whenever the transition was from Lying to Walking, the transition to be labelled as Lie-to-Stand. We theorize that when a person stands up from lying down and starts walking, there is a very short period of time during which they are in the standing posture. We consider these activities to be complex and hierarchical in nature, consisting of multiple sub-activities

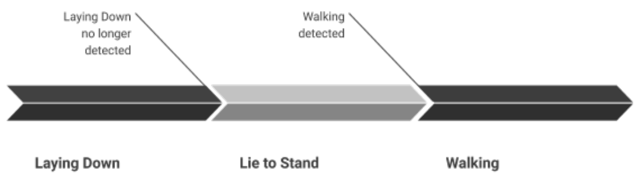

Fig. 4: Transition between Laying and Walking

[17] [18]. We propose to identify such complex activities automatically and use them to improve the transition classifier in future work.

Table II shows that our BiGRU with attention method outperforms the BiGRU and TAHAR for activities, transtions and overall. Figure 5 shows our accuracy for classifying individual activities and postural transitions. Table III is a confusion matrix of our activity and postural transition accuracies.

	Error %		
	Activities	Transitions	Overall
TAHAR [10]	$3.50\% \pm 4.7$	$0.24\% \pm 0.7$	$3.22\% \pm 4.3$
BiGRU	$1.74\% \pm 1.32$	0.19 ± 0.14	$0.96\% \pm 0.19$
BiGRU-Attn	$\mathbf{1.47\% \pm 0.96}$	$\mathbf{0.17\% \pm 0.16}$	$\mathbf{0.82\% \pm 0.28}$

TABLE II: Comparison of our Results with TAHAR

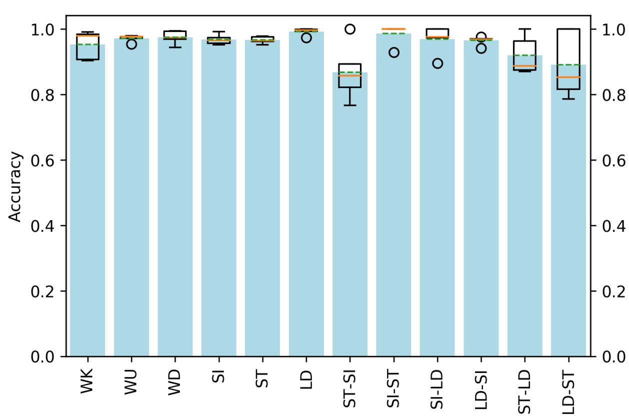

Fig. 5: Activity and Postural Transition Accuracies

	WK	WU	WD	SI	ST	LD	ST-SI	SI-ST	SI-LD	LD-SI	ST-LD	LD-ST	UK
WK	370	0	1	0	0	0	0	0	0	0	0	0	0
WU	31	270	0	0	0	0	0	0	0	0	0	0	3
WD	2	0	274	0	0	0	0	0	0	0	0	0	0
SI	0	1	0	288	55	15	0	0	0	0	0	0	8
ST	1	0	0	0	251	0	0	0	0	0	0	0	0
LD	0	0	0	0	0	347	0	0	0	0	0	0	0
ST-SI	0	0	0	0	0	0	11	0	0	0	0	0	3
SI-ST	0	0	0	0	0	0	0	7	0	0	0	0	0
SI-LD	0	0	0	0	0	0	0	0	19	0	4	0	0
LD-SI	0	0	0	0	0	0	0	0	0	15	0	1	0
ST-LD	0	0	0	0	0	0	0	0	0	0	22	0	0
LD-ST	0	0	0	0	0	0	0	0	0	0	0	16	0

WK:Walking, WU:Walking Upstairs, WD: Walking Downstairs, SI: Sitting, ST: Standing, LD: Lying Down, ST-SI: Stand-To-Sit, SI-ST: Sit-To-Stand, SI-LD: Sit-To-Lie, LD-SI: Lie-To-Sit, ST-LD: Stand-To-Lie, LD-ST: Lie-To-Stand, UK: Unknown

TABLE III: Confusion Matrix

IV. Conclusion and Future Work

We proposed *Get Up!*, a modular two-tier recurrent neural network architecture to perform activity and postural transition classification from smartphone accelerometer and gyroscope data. We were able outperform TAHAR, the previous machine learning-based state-of-the-art method for activity and postural transition classification by achieving an error rate of 1.47% and accuracy of 97%. We were also able to achieve an error rate of 0.17% with accuracy of 93.3% in classifying the postural transitions. We encountered a unique transition scenario between the activities Lying Down and Walking, which we handled by considering Walking as a complex activity that consists of sub-activities, including standing. In future work, we will handle complex activities by learning to automatically identify the sub-activities instead of manually doing it on a case by case basis.

References

[1] Najafi, Bijan, et al. ”Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly.” IEEE Trans. on Biomedical Engineering 50.6 (2003): 711-723.

[2] Salarian, Arash, et al. ”Ambulatory monitoring of physical activities in patients with Parkinson’s disease.” IEEE Transactions on Biomedical Engineering 54.12 (2007): 2296-2299.

[3] Sam, Albert, et al. ”Dyskinesia and motor state detection in Parkinson’s disease patients with a single movement sensor.” 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2012.

[4] Podsiadlo, Diane, and Sandra Richardson. ”The timed Up & Go: a test of basic functional mobility for frail elderly persons.” Journal of the American geriatrics Society 39.2 (1991): 142-148.

[5] Salarian, Arash, et al. ”iTUG, a sensitive and reliable measure of mobility.” IEEE Transactions on Neural Systems and Rehabilitation Engineering 18.3 (2010): 303-310.

[6] Milosevic, Mladen, Emil Jovanov, and Aleksandar Milenkovi. ”Quantifying Timed-Up-and-Go test: A smartphone implementation.” IEEE Int’l Conf. Body Sensor Networks (BSN) 2013.

[7] A. Mannini, A. M. Sabatini, Machine learning methods for classifying human physical activity from on-body accelerometers, Sensors 10 (2010).

[8] W. Wu, S. Dasgupta, E. E. Ramirez, C. Peterson, G. J. Norman, Classification accuracies of physical activities using smartphone motion sensors, Journal of Medical Internet Research 14 (2012) 105130.

[9] Su, Xing, Hanghang Tong, and Ping Ji. ”Activity recognition with smartphone sensors.” Tsinghua Science & Tech. 19.3 (2014): 235-249.

[10] Reyes-Ortiz, Jorge-L., et al. ”Transition-aware human activity recognition using smartphones.” Neurocomputing 171 (2016): 754-767.

[11] Abbate, Stefano, et al. ”A smartphone-based fall detection system.” Pervasive and Mobile Computing 8.6 (2012): 883-899.

[12] Villar, Jos R., et al. ”Improving human activity recognition and its application in early stroke diagnosis.” International journal of neural systems 25.04 (2015): 1450036.

[13] K. M. Culhane, M. OConnor, D. Lyons, G. M. Lyons; Accelerometers in rehabilitation medicine for older adults, Age and Ageing, Volume 34, Issue 6, 1 November 2005, Pages 556560

[14] Aminikhanghahi, Samaneh, and Diane J. Cook. ”Using change point detection to automate daily activity segmentation.” Pervasive Computing and Communications Workshops (PerCom Workshops), 2017 IEEE International Conference on. IEEE, 2017.

[15] Aminikhanghahi, Samaneh, and Diane J. Cook. ”Using change point detection to automate daily activity segmentation.” Pervasive Computing and Communications Workshops (PerCom Workshops), 2017 IEEE International Conference on. IEEE, 2017.

[16] Zhou, Peng, et al. ”Attention-based bidirectional long short-term memory networks for relation classification.” Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Vol. 2. 2016.

[17] Okeyo G, Chen L, Wang H. Combining ontological and temporal formalisms for composite activity modelling and recognition in smart homes. Future Gener Comp Sy 2014; 39: 2943

[18] Ni Q, Garca Hernando AB, de la Cruz IP. The elderlys independent living in smart homes: a characterization of activities and sensing infrastructure survey to facilitate services development. Sensors 2015; 15(5): 1131211362