35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Using Decision Trees to Measure Activities in People with Stroke

Ting Zhang, George D. Fulk, Wenlong Tang, *Member, IEEE,* and Edward S. Sazonov, *Senior Member, IEEE*

Abstract—**Improving community mobility is a common goal for persons with stroke. Measuring daily physical activity is helpful to determine the effectiveness of rehabilitation interventions. In our previous studies, a novel wearable shoe-based sensor system (SmartShoe) was shown to be capable of accurately classify three major postures and activities (sitting, standing, and walking) from individuals with stroke by using Artificial Neural Network (ANN). In this study, we utilized decision tree algorithms to develop individual and group activity classification models for stroke patients. The data was acquired from 12 participants with stroke. For 3-class classification, the average accuracy was 99.1% with individual models and 91.5% with group models. Further, we extended the activities into 8 classes: sitting, standing, walking, cycling, stairs-up, stairs-down, wheel-chair-push, and wheel-chair-propel. The classification accuracy for individual models was 97.9%, and for group model was 80.2%, demonstrating feasibility of multi-class activity recognition by SmartShoe in stroke patients.**

I. INTRODUCTION

Stroke is a leading cause of disability among adults in the US [1]. More than 4 million people in the U.S. suffer from stroke and the nation spends more than $10 billion each year for visiting post-stroke rehabilitation experts [2]. The World Health Organization's International Classification of Functioning, Disability, and Health classified the effects of stroke into problems in the "body structure and function dimension" and in the "activity and participation dimension" [3]. Over 50% of the individuals who have experienced a stroke have difficulty in walking and around 75% of them have difficulty in performing basic activities of daily living [4]. Recovering walking ability and increasing levels of activity and mobility are important goals of the rehabilitation of post-stroke patients [5]. Therefore, it is very important to monitor physical activity for people with stroke. Such monitoring can help to determine the effectiveness of rehabilitation interventions as well as to provide behavior-enhancing feedback.

The performance of patients with stroke in clinics or in research labs may not truly reflect patients' real-life performance [6][7] . Therefore, there is a strong need for systems that can monitor physical activities in free-living conditions for stroke patients [8]. Modern sensor technologies and signal processing techniques make this possible. There are various sensor-based systems that can help with monitoring activities in patients with stroke. An accelerometer was worn on the wrist to monitor the amount of usage of the affected upper extremity (UE) in people after stroke in [9]. A Stroke Upper Limb Activity Monitor combines accelerometers placed on the lower extremities (LE), trunk, and UE with electrogoniometers on both elbows [10]. Saremi et al. examined the reliability and validity of the Intelligent Device for Energy Expenditure and Activity (IDEEA) with 5 bi-axial accelerometers [11] for hemiparetic subjects. However, most of these systems are too obtrusive for everyday use. Moreover, many people with stroke have limited walking ability [12]. Therefore, there is a need to better explore the use of sensors and to measure a wider range of activities of people with stroke.

We developed an unobtrusive wearable shoe-based sensor system (SmartShoe) for physical activity monitoring. It combines a tri-axial accelerometer mounted on the heel of each shoe and force sensitive resistors (FSR) embedded in insoles [13]. In our previous studies, the shoe sensor system with decision tree algorithm for activity classification was used on healthy subjects, but the performance for data from stroke patients was never explored [13]. Also, this shoe sensor system could accurately identify the 3-class postures (i.e. sitting, standing, and walking) in people with stroke by using artificial neural network (ANN) [6] with classification accuracy of the ANN for individual participants from 93.1% to 99.9% and the combined classification accuracy approximately 97.2%.

Although ANN could achieve high classification accuracy, the model to make decisions is a vast network of neurons, thus the computation is not easy to interpret. Also the algorithm needs floating point operation, therefore, it is not well suited for embedded devices with no floating point support. There is a need to utilize algorithms that can be simpler to interpret and easier to be embedded into a chip—such as a mobile phone or even Programmable System on Chip (PSOC). In this study, we utilized decision tree algorithms as a simpler alternative to ANN. Further, we extended the study to 8-classes (i.e. sitting, standing, walking, cycling, stairs-up, stairs-down, wheel-chair-push, and wheel-chair-propel).

II. METHODS

A. Wearable shoe sensors and data collection

The version of SmartShoe sensor system used in this study included pressure sensors and accelerometers. The pressure sensors contained five FSRs integrated in a flexible insole in each shoe, and the accelerometer was

Ting Zhang, and Wenlong Tang are with The Department of Electrical and Computer Engineering, The University of Alabama.

George D. Fulk is with The Physical Therapy Department, Clarkson University.

Edward S. Sazonov is with The Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL 35487 USA (Phone: 205-348-1981; Fax: 205-348-6959; e-mail: esazonov@eng.ua.edu)

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

positioned on the back of each shoe. The battery, power switch, and wireless board were installed on a rigid circuit board glued to the back of the shoe. Pressure and acceleration data were sampled at 25 Hz, which has been proved as an applicable rate, by a 12-bit A/D converter and then sent over Bluetooth to a smart phone for off-line data processing [14]. Each single sample of data from the shoe was represented by a vector $S = \{A_{AP}, A_{ML}, A_{SI}, P_H, P_{5M}, P_{3M}, P_{1M}, P_{HX}\}$. The meanings of the sensor readings and their corresponding representations in our decision tree algorithms are shown in Table I.

TABLE I. SENSOR REPRESENTATIONS

Sensor signals	Descriptions	Notations in decision trees
A_{AP}	Acceleration in the Anterior-posterior direction	acc1
A_{ML}	Acceleration in the Medial-lateral direction	acc2
A_{SI}	Acceleration in the Superior-inferior direction	acc3
P_H	Pressure applied on the FSR under the heel	pre1
P_{5M}	Pressure applied on the FSR under the fifth metatarsal head	pre2
P_{3M}	Pressure applied on the FSR under the third metatarsal head	pre3
P_{1M}	Pressure applied on the FSR under the first metatarsal head	pre4
P_{HX}	Pressure applied on the FSR under the hallux	pre5

B. Description of participants and activities

Data collection was performed on twelve participants with stroke. Eight participants had a middle cerebral artery stroke, three had brainstem stroke, and one had a cerebellar stroke. They were recruited from a local outpatient physical therapy clinic. Inclusion criteria were as follows: at least 3 months post-stroke, able to walk at home and/or community without physical assistance, able to stand without physical assistance for more than 60 seconds, able to transition from sitting to standing from a standard height chair without physical assistance [6]. Subjects were wearing SmartShoe for the duration of data collection. Pressure and acceleration data were collected in 8 posture and activity groups: sitting, standing, walking, ascending stairs, descending stairs, cycling on a stationary bike, being pushed in a wheelchair, and propelling a wheelchair.

TABLE II. DESCRIPTION OF THE SIX POSTURE/ACTIVITIES CLASSES

Activities	Description
Sitting ('sit') (4 ADLs)	Reading a magazine
	Working on a computer
	Simulated dating and drinking
	Sitting comfortably
Standing ('stand') (4 ADLs)	Folding laundry
	Donning/Doffing a coat
	Placing objects in a cupboard that is approximately head level
	Standing comfortably
Walking ('walk') (2 trials)	Walking at their self-selected pace
	Walking at their fastest safe pace
Ascending stairs ('stairs_up')	Subjects go upstairs
Descending stairs ('stairs_down')	Subjects go downstairs
Cycling	Subjects perform riding a stationary bike
Pushed in a wheelchair ('wc_push')	Someone pushing the subject when both their feet are up on a footplate, arms on the arm rests of the wheel chair
Propelling a wheelchair ('wc_propel')	The subject propelling the wheel chair using the unaffected arm and leg, with the affected leg up on a footplate

Data were collected for different activities in multiple positions to better mimic real-life conditions. For sitting and standing, participants performed activities of daily living (ADLs) while in these postures. Participants performed each activity in the sitting and standing 3 times for a minute. In addition to perform ADLs in sitting and standing activity, participants performed walking activity at two different paces, self-selected and fastest safe pace. They performed three 2-minute walk trials at each pace. During the data collection process, all subjects were supervised by a physical therapist for safety. The order in which each position trial was performed was randomized by means of a random number generator. The description of the eight major posture/activity classes is shown in the above table, Table II.

B. Feature computation

Feature computation was performed on the data readings from every continuous two second disjoint processing windows. The features were mean, standard deviation, entropy [13], variance, maximum value, Number of Mean Crossings (NMC), and Mean Absolute Deviation (MAD). Therefore, the number of computed features in every two seconds will be 7 features / sensor * 8 sensors = 56 features. In our experiment, feature computation was performed in MATLAB. However, we always rounded the feature values to integers for potential further use in embedded microcontrollers.

C. The classifier—C5.0 decision tree

Decision Tree is a hierarchical model that recursively separates the input space into class regions. The result is a tree-like structure, which composes of decision nodes and leaves. Each node has a test function that determines which branch is taken for the next step. This process is repeated until one of the leaves is reached and therefore a decision is made. Iterative Dichotomiser 3 (ID3) algorithm, C4.5 algorithm and C5.0 are all implementations of decision tree algorithms [15].

In this study, we firstly split the dataset into two-second segments. Then we computed features for each data segment. We randomly selected half of the feature vectors for training and half for validation. After generating a decision tree from training feature vectors, we computed the class labels for each testing feature vector and got validation results. The program was written in C programming language. Figure 1 shows the procedure of how this experiment was performed.

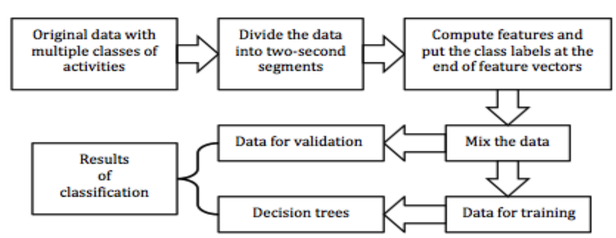

Figure 1. Data processing procedure with decision trees.

First, we built individual models and group models for the 3-class classification of posture and allocations. Then, we performed the same procedure for the 8-class of activities. Individual models performed training and validation on the same experimental subject. The individual models were the best fit to the individual traits and thus represented the baseline of accuracy. However, individual models are also more prone to over fitting. The group models were developed using Leave-One-Out cross-validation. The group models were trained on a dataset gathered from multiple subjects and therefore enabled a subject-independent classification.

III. RESULTS

For the training and validation of the 3-class classification, Figure 2 shows a decision tree structure obtained from the individual model built for Subject 1. The branches are the computed features from the sensor readings. The thresholds are obtained from the training process. The tree leaves are the classification results.

```
Decision tree:
entropy_acc2 > 10257: walk (196)
entropy_acc2 <= 10257:
: . . . mean_pre2 <= 125: sit (176)
        mean_pre2 >125:
        : . . . standard deviation_pre2 > 48: stand (153/2)
                standard deviation_pre2 <= 48:
                : . . . mean_pre4 <= 1266: stand (20)
                        mean_pre4 >1266:
                        : . . . mean absolute deviation_pre4 <= 54: sit (12)
                                mean absolute deviation_pre4 >54: stand (2)
```

Figure 2. Decision tree generated for the classification for 3-class individual model for Subject 1.

In Figure 2, the connection nodes (branches) are displayed as feature names followed by sensor names. The sensor names are shown in Table I. The first three lines of the decision tree can be read as: if the entropy of the acceleration in the medial-lateral direction is greater than 10,257, the activity is "walk". If not, while the mean of the pressure applied on the FSR under the fifth metatarsal head is not greater than 125, the activity is classified as "sit". Among all the feature vectors, there are 196 vectors being classified as "walk" and 176 vectors being classified as "sit".

Table III shows the attribute usages in percentage in the decision tree above. It is obvious that not all features are involved nor all sensors are used. Table IV shows the number of features and sensors that were used for each subject in individual models. There are 56 different features and 5 different sensor signals. The number of features and sensors that involved in the classification is much less than the total. Table V shows the classification accuracy for the 3-class activities in the individual models and the group models. A cumulative confusion matrix for the 3-class classification in group models is shown as Table VI.

TABLE III. ATTRIBUTE USAGE FROM THE DECISION TREE

Attribute name	Attribute usage
Entropy_acc 2	100%
Mean_pre 2	65%
Standard deviation_pre 2	33%
Mean_pre 4	6%
Mean absolute deviation_pre 4	3%

TABLE IV. FEATURES AND SENSORS USED IN CLASSIFICATION FOR EACH SUBJECT

Subject	S 1	S 2	S 3	S 4	S 5	Average	Total
N(features)	5	5	2	2	2	3	56
N(sensors)	3	2	2	2	2	2	5

TABLE V. CLASSIFICATION ACCURACY FOR 3-CLASS ACTIVITY

Subject	S 1	S 2	S 3	S 4	S 5	Average
Individual Model Accuracy	98.6%	98.2%	99.3%	99.8%	98.8%	99.1%
Group Model Accuracy	94.4%	97.9%	72.5%	98.8%	94.1%	91.5%

TABLE VI. CUMULATIVE CONFUSION MATRIX FOR THE 3-CLASS GROUP MODELS.

Predicted / Actual	sit	stand	walk	Recall
sit	1763	48	10	0.968
stand	10	1715	45	0.969
walk	40	307	1580	0.820
Precision	0.972	0.828	0.966	0.917

Further, we extended the 3-class activity models to the 8-class activity models. A sample decision tree for the individual model is shown in Figure 3.

```
Decision tree:

mean_acc2 <= 1596: walk (181)
mean_acc2 > 1596:
: . . . mean absolute deviation_acc2 > 48:
        : . . . mean_pre5 <= 627:
        :       : . . . num of mean crossing acc1 <= 7: sit (3)
        :       :       num of mean crossing acc1 >7: wc_propel (44)
        :       mean_pre5 >627:
        :       : . . . mean absolute deviation_pre5 <= 429: bike (44)
        :               mean absolute deviation_pre5 >429:
        :               : . . . max_acc3 <= 2653: stairs_down (9/1)
        :                       max_acc3 > 2653: stairs_up (4)
        mean absolute deviation_acc2 <= 48:
        : . . . mean_acc3 <= 1963: wc_push (46)
                mean_acc3 >1963:
                : . . . mean_pre2 <= 116: sit (188)
                        mean_pre2 > 116:
                        : . . . num of mean crossing_acc1 > 6 sit (5/1)
                                num of mean crossing_acc1 <= 6:
                                : . . .variance_pre2 > 571 stand (160/4)
                                       variance_pre2 <=571:
                                       : . . . max_pre4 <=1212: stand (6)
                                               max_pre4 >1212: sit (6)
```

Figure 3. Decision tree generated for classification for 8-class individual model for Subject 1.

The classification results for the 8-class are shown in Table VII. The cumulative confusion matrix for the group model is shown in Table VIII.

TABLE VII. CLASSIFICATION ACCURACY FOR 8-CLASS ACTIVITIES

Subject	S 1	S 2	S 3	S 4	S 5	Average
Individual Model Accuracy	96.5%	97.4%	99.8%	97.2%	98.4%	97.9%
Group Model Accuracy	87.5%	91.1%	64.7%	82.2%	75.5%	80.2%

TABLE VIII. CUMULATIVE CONFUSION MATRIX FOR 8-CLASS GLOBAL MODELSSAME COMMENT AS PREVIOUS TABLE

Predicted / Actual	sit	stand	walk	bike	stairs_up	stairs_down	wc_push	wc_propel	Recall
sit	1701	36	9	13	0	1	1	60	0.93
stand	11	1574	19	71	11	3	2	0	0.93
walk	1	72	1326	4	67	319	1	16	0.73
bike	5	3	23	325	0	0	1	2	0.90
stairs_up	1	14	18	5	16	5	0	0	0.27
stairs_down	1	10	32	5	5	9	0	0	0.15
wc_push	10	3	0	0	0	0	323	88	0.76
wc_propel	121	24	21	0	0	2	204	78	0.17
Precision	0.92	0.91	0.92	0.77	0.16	0.03	0.60	0.32	0.80

For the 8-class individual models, the classification accuracy is still acceptable for real-life applications. For some activities in the group model, the classification accuracy is not high, and analysis is given in discussion.

IV. CONCLUSION AND DISCUSSION

In this study, we utilized decision tree algorithms for the classification of postures and activities captured by SmartShoe from stroke patients. For the main postures and activities, the subjects performed activities in different positions to mimic situations when the subjects are in their home and community. This demonstrates that a combination of the SmartShoe and the decision tree algorithms can accurately measure postures and activities for people with stroke when they perform common ADLs or perform mobility tasks.

Comparing with ANN, decision tree algorithms can be easily interpreted with real-life meaning, especially when it works with the feature computation. The algorithms are fast as well and also give a direct approach to simplify feature computation and sensor usage. From Figure 2 and Figure 3, it is found that the tree structures are not complicated for three class or eight class activity models. Table IV shows an example of simplification of features and sensors. The whole process of decision tree's decision-making process does not need any floating-point operation, which ANN cannot avoid. Since the tree structures are simple, and the computation does not need floating point operation, the decision tree algorithms in our study have a great potential to be embedded into mobile devices and even sensor chips.

The average classification accuracy for the 3-classes of activities (sit, stand, walk) is 99.1% for individual models and 91.5% for group models. The average classification accuracy for the 8-classes of activities is 97.9% for individual models and 80.2% for group models. The classification accuracy is lower than that for healthy subjects, but is still acceptable. Table VIII shows high misclassification rate for "stairs_up" and "stairs_down". A possible reason is that the sample sizes for these activities are not as big as the others. Thus, the training process may not provide enough information for the algorithms to learn the data. A large number of "wc_propel" segments are misclassified as "wc_push". One possible reason is that to keep consistency of our previous study with decision tree and to keep the features simple, we used only data from one shoe in the algorithm. Therefore, when the activity is not symmetrical, such as "wc_propel", misclassification may occur. We may consider using more features in our future studies.

In conclusion, decision tree algorithms and SmartShoe show a capability of measuring postures and activities of stroke patients as they perform common ADLs and mobility tasks. Future work includes exploring more effective features for distinguish particular activities, using pruning methods to simplify the tree structures, and embedding the algorithms into mobile devices and even sensor chips for real-time applications.

REFERENCES

[1] D. Lloyd-Jones, R. J. Adams, T. M. Brown, M. Carnethon, S. Dai, G. De Simone, T. B. Ferguson, E. Ford, K. Furie, C. Gillespie, A. Go, K. Greenlund, N. Haase, S. Hailpern, P. M. Ho, V. Howard, B. Kissela, S. Kittner, D. Lackland, L. Lisabeth, A. Marelli, M. M. McDermott, J. Meigs, D. Mozaffarian, M. Mussolino, G. Nichol, V. L. Roger, W. Rosamond, R. Sacco, P. Sorlie, V. L. Roger, R. Stafford, T. Thom, S. Wasserthiel-Smoller, N. D. Wong, and J. Wylie-Rosett, "Heart disease and stroke statistics--2010 update: a report from the American Heart Association," *Circulation*, vol. 121, no. 7, pp. e46–e215, Feb. 2010.

[2] M. Kelly-Hayes, A. Beiser, C. S. Kase, A. Scaramucci, R. B. D'Agostino, and P. A. Wolf, "The influence of gender and age on disability following ischemic stroke: the Framingham study," *J Stroke Cerebrovasc Dis*, vol. 12, no. 3, pp. 119–126, Jun. 2003.

[3] A. Cieza, T. Ewert, T. B. Ustün, S. Chatterji, N. Kostanjsek, and G. Stucki, "Development of ICF Core Sets for patients with chronic conditions," *J Rehabil Med*, no. 44 Suppl, pp. 9–11, Jul. 2004.

[4] H. S. Jørgensen, H. Nakayama, H. O. Raaschou, and T. S. Olsen, "Stroke. Neurologic and functional recovery the Copenhagen Stroke Study," *Phys Med Rehabil Clin N Am*, vol. 10, no. 4, pp. 887–906, Nov. 1999.

[5] E. J. Harris JE, "Goal Priorities Identified through Client-Centred Measurement in Individuals with Chronic Stroke," *Physiotherapy Canada*, vol. 56, no. 3, pp. 171–176, 2004.

[6] G. D. Fulk, S. R. Edgar, R. Bierwirth, P. Hart, P. Lopez-Meyer, and E. Sazonov, "Identifying activity levels and steps of people with stroke using a novel shoe-based sensor," *J Neurol Phys Ther*, vol. 36, no. 2, pp. 100–107, Jun. 2012.

[7] G. D. Fulk, C. Reynolds, S. Mondal, and J. E. Deutsch, "Predicting home and community walking activity in people with stroke," *Arch Phys Med Rehabil*, vol. 91, no. 10, pp. 1582–1586, Oct. 2010.

[8] P. Lopez-Meyer, G. D. Fulk, and E. S. Sazonov, "Automatic detection of temporal gait parameters in poststroke individuals," *IEEE Trans Inf Technol Biomed*, vol. 15, no. 4, pp. 594–601, Jul. 2011.

[9] G. Uswatte, C. Giuliani, C. Winstein, A. Zeringue, L. Hobbs, and S. L. Wolf, "Validity of accelerometry for monitoring real-world arm activity in patients with subacute stroke: evidence from the extremity constraint-induced therapy evaluation trial," *Arch Phys Med Rehabil*, vol. 87, no. 10, pp. 1340–1345, Oct. 2006.

[10] M. de Niet, J. B. Bussmann, G. M. Ribbers, and H. J. Stam, "The stroke upper-limb activity monitor: its sensitivity to measure hemiplegic upper-limb activity during daily life," *Arch Phys Med Rehabil*, vol. 88, no. 9, pp. 1121–1126, Sep. 2007.

[11] K. Saremi, J. Marehbian, X. Yan, J.-P. Regnaux, R. Elashoff, B. Bussel, and B. H. Dobkin, "Reliability and validity of bilateral thigh and foot accelerometry measures of walking in healthy and hemiparetic subjects," *Neurorehabil Neural Repair*, vol. 20, no. 2, pp. 297–305, Jun. 2006.

[12] E. Haeuber, M. Shaughnessy, L. W. Forrester, K. L. Coleman, and R. F. Macko, "Accelerometer monitoring of home- and community-based ambulatory activity after stroke," *Arch Phys Med Rehabil*, vol. 85, no. 12, pp. 1997–2001, Dec. 2004.

[13] T. Zhang, W. Tang, and E. S. Sazonov, "Classification of posture and activities by using decision trees," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012, pp. 4353 –4356.

[14] E. S. Sazonov, G. Fulk, J. Hill, Y. Schutz, and R. Browning, "Monitoring of Posture Allocations and Activities by a Shoe-Based Wearable Sensor," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 4, pp. 983 –990, Apr. 2011.

[15] J. R. Quinlan, "Induction of Decision Trees," *Mach. Learn*, pp. 81–106, 1986.