Development of an algorithm to count steps from 24hr wrist accelerometry data

Introduction: the purpose of this work is to develop an algorithm to count steps from long term wrist accelerometer data.

Methods: Data from 30 subjects performing three different types of walking activities was obtained in conjunction with video data to allow for manual annotation of step occurrences. The three types of walking collected were regular, semi-regular and unstructured. A 24hr data-set was also included to test for performance on long term, remote monitoring data. Thresholds were tuned to find optimal results.

Results: Algorithm thresholds were tuned to prevent the proliferation of false positives in the 24hr data. Tuned thresholds resulted in F1 scores of 0.913, 0.755 and 0.575 on regular, semi-regular and unstructured walking conditions from the test-set hold out data. Precision scores on the same data-set were 0.997, 0.957 and 0.751 on regular, semi-regular and unstructured walking conditions.

Conclusion: On the annotated data-set, there is no indication of variance or bias, as the F1 and precision scores between the training and testing results are very similar. The steps algorithm presented in [2] with an additional magnitude threshold and adjusted coefficients is accurate for detecting steps in long term, remote monitoring scenarios.

Introduction

The purpose of this work is to develop an algorithm to count steps from long term wrist acceleration data.

Methods

30 participants performed three different types of walking tasks while wearing a tri-axial accelerometer (Shimmer3, Shimmer, Dublin, Ireland) on their wrist and being recorded with a video camera. Only the legs of the participants were recorded to ensure privacy. The walking tasks were regular, semi-regular and unstructured walking. The regular walking task consisted of walking two laps around a 350m path at the participants comfortable walking pace. The semi-regular walking task consisted of the participants performing a scavenger hunt in a building, thus the participant performed many stop and starts, shuffle steps, stair climbing and descending as well as turning around. The unstructured walking task was designed to replicate cooking a meal and consisted of the participants constructing a Lego toy for which they had to travel short distances to different buckets to get the required parts.

The video data was synchronized with the accelerometer data via the use of a custom manual annotation software tool that played the video and accelerometer data at the same time and allowed for an expert technician to manually annotate where each step occurred in the data. Further details of the data-set can be found in [1].

A secondary data-set was collected to test the algorithm in a long term monitoring scenario. This consisted of one subject wearing a Verisense (Shimmer, Dublin, Ireland) accelerometer on their nondominant wrist and an Apple Watch (Series 4, California, USA) on the same wrist, above the other sensor.

Statistics

For the annotated data set, a true positive was counted if an algorithm derived step occurred within plus or minus three samples of a manually annotated step location. A false positive occurred if an algorithm derived step occurred without a manually annotated step and a false negative was counted if there was no algorithm derived step and a manually annotated step. Precision (also known as positive predictive value – PPV) was calculated as follows;

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Recall was calculated as follows;

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

F1 score was then calculated as follows;

$$F1\ Score = 2\ \times \frac{Precision\ \times\ Recall}{Precision + Recall}$$

On the 24hr data, mean average error (MAE) was calculated by summing the absolute difference in the step count between the Apple Watch and the algorithm for each 24hr recording.

Step Detection Algorithm

A step detection algorithm was implemented based off a popular method presented in the literature [2]. The basis of the algorithm is designed around the common peak detection technique. First, the square root of the summed and squared acceleration channels are calculated to obtain the magnitude of acceleration measured. Peak values in a set range are found over the vector. Then, three features are quantified in relation to each peak; periodicity, similarity and continuity. Thresholds are set on each of these features to filter out artefacts. Table 1 shows the thresholds used in [2].

Threshold Description	Symbol	[2] value
Peak detection threshold	K	15
Minimum periodicity threshold	t_min	0.3
Maximum periodicity threshold	t_max	1
Similarity threshold	sim	.5
Window size of continuity	M	2
Number threshold of continuity	N	4
Variance threshold for motion recognition	var	0.7

Table 1 – Recommended thresholds table.

Performance on the 24hr data-set revealed a high prevalence of false positives (Table 2). This is due to the fact that the algorithm thresholds were optimized to detect steps in data in which the majority of the data consisted of stepping. In 24h records, much of the data does not consist of stepping, thus the chance for a high amount of false positives is increased.

Trial	Apple Watch Steps	Accelerometer Steps
1	7104	25018
2	4158	18942
3	4689	17407
4	3978	18765
5	3742	17741
Mean	**4734**	**19574**

Table 2 – Accelerometer step count based on published thresholds.

An additional threshold was added to the algorithm to reject steps if the magnitude of the acceleration was under a certain magnitude. A range of thresholds was used and a tuning exercise was carried out on 20 of the 30 tests in the annotated walking data-set. As opposed to maximizing the F1 score on detections on each type of walking condition, precision on the unstructured walking was maximized in tests where the F1 score for regular, semi-regular and unstructured was over 0.75, 0.7 and 0.55, respectively and the precision for regular and semi-regular conditions was over 0.8. The threshold levels that were tested are listed in Table 3.

Threshold Description	Symbol	Values to test
Peak detection threshold	K	3, 4, 5
Minimum periodicity threshold	t_min	4, 5
Maximum periodicity threshold	t_max	15, 25, 35
Similarity threshold	sim	-1, -0.5, -0.1
Window size of continuity	M	2, 3, 4, 5
Number threshold of continuity	N	4, 6, 8, 10
Variance threshold for motion recognition	var	0.001, 0.07, 0.1
Magnitude threshold		1.0, 1.2, 1.3

Table 3 – Ranges of thresholds to use in algorithm tuning.

Results

7,776 threshold combinations were tested to determine step count on the training data-set. Of those results, 532 met the minimum threshold criteria (Figure 1). Precision on unstructured data was maximized in order to prevent the proliferation of false positives on long term, remote monitoring data.

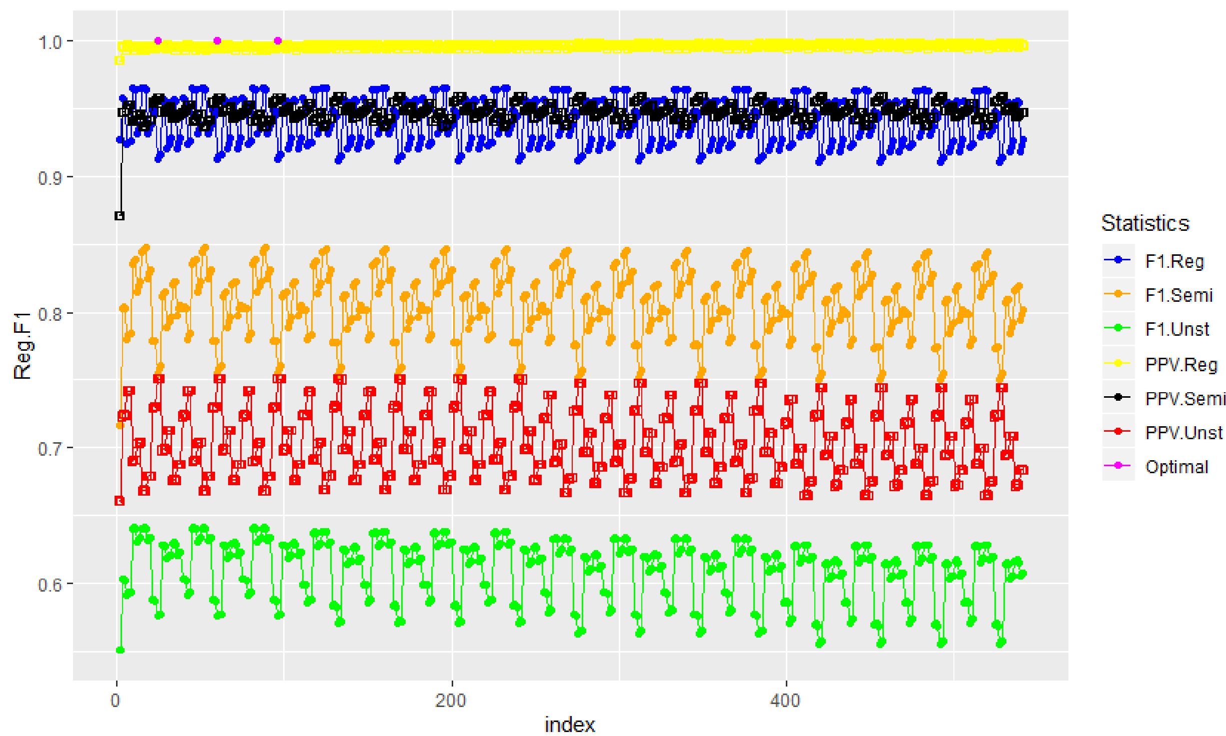

Figure 1 – threshold combinations that met the criteria of having regular walking F1-score over 0.75, semi-regular walking F1-score over 0.7, unstructured walking F1-score over 0.55, regular walking and semi-regular walking precision over 0.8.

The three top performing combinations of thresholds had the same F1 and precision scores on all walking conditions. They were tested on the 24hr data-set to assess their accuracy against the Apple Watch results. Absolute error was calculated for each trial and each trial was summed to arrive at a total MAE score for each combination of thresholds.

Trial	Apple Watch Step Count	Thresholds-1	Abs Err	Thresholds-2	Abs err	Thresholds-3	Abs err
1	7104	6737	367	6737	388	6651	453
2	4158	4556	398	4505	347	4365	207
3	4689	3866	823	3819	870	3658	1031
4	3978	4439	461	4391	413	4212	234
5	3742	4605	863	4563	821	4393	651
Total MAE			**2912**		**2839**		**2576**

Table 4 – Step count results with the top performing combinations of thresholds on the 24hr data-set. MAE – mean average error.

The thresholds that resulted in the lowest MAE score on the 24hr data are listed in Table 5.

Threshold Description	Symbol	Values to test
Peak detection threshold	K	3
Minimum periodicity threshold	t_min	5
Maximum periodicity threshold	t_max	15
Similarity threshold	sim	-0.5
Window size of continuity	M	4
Number threshold of continuity	N	4
Variance threshold for motion recognition	var	0.001
Magnitude threshold		1.2

Table 5 – Algorithm thresholds that results in the lowest MAE score on 24hr data.

The thresholds stated in Table 5 were used to calculate F1 and precision scores on the training and testing data in order to look for bias or variance in the algorithm. Results are shown in Table 6.

Statistic	Walking condition	Train data	Test data
n		20	10
F1	Regular	0.894	0.913
F1	Semi-Regular	0.715	0.755
F1	Unstructured	0.560	0.575
Precision	Regular	0.999	0.997
Precision	Semi-Regular	0.951	0.957
Precision	Unstructured	0.719	0.751

Table 6 – F1 and precision scores on regular, semi-regular and unstructured walking conditions for the optimal threshold combination.

Conclusion

On the annotated data-set, there is no indication of variance or bias, as the F1 and precision scores between the training and testing results are very similar. The steps algorithm presented in [2] with an additional magnitude threshold and adjusted coefficients is accurate for detecting steps in long term, remote monitoring scenarios.

References

[1] Mattfeld, R., Jesch, E. and Hoover, A., 2017, November. A new dataset for evaluating pedometer performance. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 865-869). IEEE.

[2] Gu, F., Khoshelham, K., Shang, J., Yu, F. and Wei, Z., 2017. Robust and accurate smartphone-based step counting for indoor localization. *IEEE Sensors Journal*, *17*(11), pp.3453-3460.