2013 IEEE 10th International Conference on Ubiquitous Intelligence & Computing and 2013 IEEE 10th International Conference on Autonomic & Trusted Computing

Towards Physical Activity Recognition Using Smartphone Sensors

Muhammad Shoaib, Hans Scholten, P. J. M. Havinga
Pervasive Systems, Department of Computer Science and Electrical Engineering,
University of Twente,
Enschede, The Netherlands
{m.shoaib, hans.scholten, p.j.m.havinga}@utwente.nl

Abstract—**In recent years, the use of a smartphone accelerometer in physical activity recognition has been well studied. However, the role of a gyroscope and a magnetometer is yet to be explored, both when used alone as well as in combination with an accelerometer. For this purpose, we investigate the role of these three smartphone sensors in activity recognition. We evaluate their roles on four body positions using seven classifiers while recognizing six physical activities. We show that in general an accelerometer and a gyroscope complement each other, thereby making the recognition process more reliable. Moreover, in most cases, a gyroscope does not only improve the recognition accuracy in combination with an accelerometer, but it also achieves a reasonable performance when used alone. The results for a magnetometer are not encouraging because it causes over-fitting in training classifiers due to its dependence on directions. Based on our evaluations, we show that it is difficult to make an exact general statement about which sensor performs better than the others in all situations because their recognition performance depends on the smartphone's position, the selected classifier, and the activity being recognized. However, statements about their roles in specific situations can be made. We report our observations and results in detail in this paper, while our data-set and data-collection app is publicly available, thereby making our experiments reproducible.**

Keywords— accelermeter; activity recognition; assisted living; gyroscope; health monitoring; magnetometer; sensor fusion; smartphone sensors; well-being applications.

I. INTRODUCTION

Physical activities play a very important role in people's physical and mental wellbeing[1]. The lack of physical activities can badly affect their wellbeing in general [14]. The use and capability of smartphones for motivating people to be physically active is discussed in detail in [1][2]. There is a need for a proper health and lifestyle coaching that detects user's activity and situation in real time and provides people with right motivational feedback. Such coaching mechanisms, if implemented practically, can help to reduce the overall cost for governments, employers and insurance companies caused by people's ill health and unhealthy lifestyle, and at the same time it can help to improve people's wellbeing [3]. To give people the right motivational feedback, their activities need to be recognized first. For this purpose, smartphones have been extensively studied for recognizing different physical activities in recent years [6]. Smartphones are being used because people already carry them and they are equipped with different sensors like an accelerometer that can be used in developing well-being coaching applications [14].

In the state of the art, the smartphone accelerometer has been mainly used for the activity recognition. Some people have used a gyroscope too. To the best of our knowledge, these two sensors are not individually analyzed, rather an accelerometer is considered as the main sensor and a gyroscope as an additional sensor in the activity recognition. For example, in [4], the authors use a gyroscope in combination with an accelerometer and report an increase in recognition accuracy by 3.1-13.4% for some activities. However, they evaluate it only using K-nearest neighbor (KNN) classifier on pocket position. On the other hand, in [5], the authors claim that the addition of a gyroscope to an accelerometer doesn't add any value to the recognition performance (accuracy). We extend this work by analyzing it from different angles and in a detailed way. We also find the possible reason why the authors in [4] and [5] show different results about the addition of a gyroscope (to be discussed in section VI).

To the best of our knowledge, we are the first to carry out a detailed analysis about the individual and combined roles of an accelerometer, a gyroscope and a magnetometer in physical activity recognition. In this work, we answer the question that in which situations an accelerometer performs better than a gyroscope and vice versa and why or when to combine the two sensors for better recognition performance? It is very important to answer this question because it enables us to use an appropriate sensor for the right activity, rather than just combining many sensors because more sensors mean more energy consumption for the smartphone battery. For this purpose, we evaluate each sensor individually as well as in combination with other sensors, to assess its individual and combined ability in activity recognition. We show the role of these sensors in overall recognition performance for all six activities as well as for individual activities. There are trends which give us a better insight about how these sensors perform in certain situations but it is difficult to make an exact generic statement about their role in all situations because our

[1] Wellbeing refers to how people experience the quality of their lives and includes both emotional reactions and cognitive judgments. (source: http://en.wikipedia.org/wiki/Subjective_well-being)

978-1-4799-2481-3/13 $31.00 © 2013 IEEE
DOI 10.1109/UIC-ATC.2013.43

evaluations demonstrate that their roles depend on the position of the smartphone, the activity being recognized, and the classifier being used. We discuss in detail these results in section VI. We list below our main contributions:

- We make our data set and data collection application publicly available so they can contribute to future research in this domain. The data collector will speed up the process of conducting such experiments rather than developing a new one. It can be accessed at [15].
- To the best of our knowledge, we for the first time evaluate the role of a smartphone accelerometer, gyroscope and magnetometer in activity recognition both when they are used alone as well as in combination with each other. By doing so, we show which in situations they perform better and which factors affect their performance.
- We evaluate these three sensors for four body position using seven commonly used classifiers. Moreover, we recognize six physical activities, commonly used in the state of the art.
- Unlike previous studies, we evaluate these sensors on a dataset with balanced class distribution in addition to a dataset with imbalanced class distribution (discussed in section V).

The rest of the paper is organized as follows. We describe related work in section II and data collection process in section III. The data preprocessing is described in section IV and our experiments' design in section V. We discuss the performance evaluation in section VI. Finally, we describe our cconclusions and future work in section VII.

II. RELATED WORK

The use of a smartphone accelerometer in activity recognition has extensively been studied and has been summarized in [6]. However, the use of a gyroscope and a magnetometer in activity recognition is yet to be explored. Some researchers have evaluated the gyroscope's effect on an accelerometer. For example, the authors in [5] used an accelerometer and a gyroscope in combination and claimed that the gyroscope adds no value to the overall recognition performance. They used Naïve bays, Decision Tree (C.45) and K-nearest neighbor (KNN) for classification and collected data on multiple body positions [5].

The authors in [4] used an accelerometer and a gyroscope together, and reported 3.1-13.4% increase in recognition accuracy for some activities. However, they did these experiments only with KNN classifier. They used pocket position only except for jogging activity for which arm position was used. In [7], the authors used an accelerometer, a magnetometer, a gyroscope, linear acceleration, and the gravity in combination. Though this combination performed slightly better than the accelerometer alone, the paper does not discuss the role of these individual sensors. Therefore, it is not clear which sensors contributed (and how much) to the improvement in the activity recognition.

Our work is different from the existing work in certain aspects. In existing work, the role of a gyroscope and a magnetometer is seen as an additional sensor to an accelerometer [4][5][7] whereas we evaluate it individually as well as in combination with an accelerometer, thereby identifying their individual contributions in the activity recognition. This enables us to see where and when a sensor performs better than others and when they perform better together. Moreover, we evaluate these sensors using seven commonly used classifiers in activity recognition. This evaluation is done on four different body positions. This enables us to make more confident claims about our reported results. We do activity-wise analysis of each sensor that shows how they perform in recognizing individual activities. The importance of such activity-wise analysis is discussed in section VI. We also explain the possible reasons why the authors in [4] report an increase in performance by using a gyroscope in combination with an accelerometer while those in [5] report no improvement.

III. DATA COLLECTION

For the purpose of our experiments, we developed a data collection application for Android devices. This Android app currently collects data from the GPS (not used in our analysis), an accelerometer, a magnetometer and a gyroscope at a rate of 50 samples per second. However, more sensors can easily be added to it and its sampling rate can also be changed as needed. In order to keep the app's interface simple and to make it easy for use by the participants, we hard coded a sampling frequency of 50 samples per second in our app, rather than asking users to set a frequency at data collection time. This sampling rate (50 samples per second) is enough to recognize human physical activities as we show in section VI. Moreover, in the state of the art, frequencies lower than 50 samples per second have been shown to be sufficient for activity recognition [4][5].

We used four Samsung Galaxy S2 smartphones for data collection [8]. Using these smartphones, we collected data for six different physical activities. They are walking, running, sitting, standing, walking upstairs and downstairs. We asked four participants to perform these activities for few minutes. As these are repetitive activities, the amount of time for each activity was kept between 3-5 minutes per participant which gave enough examples for our evaluations. The activities were carried out indoor in one of our university buildings. For walking, and running, the department corridor was used. For sitting and standing activity, participants' offices were used. For walking upstairs and downstairs, stairs (5 floors) were used. It is important to mention that these stairs had short walks at each floor for switching between stairs (2-3 steps) but there were only four such switches in the whole walking upstairs and downstairs activities. Each of these participants was provided with four smartphones on four body positions: right jeans pocket, belt, right arm, and right wrist. The orientation of the smartphones was portrait for the arm, wrist, and pocket and was landscape for belt. The data was recorded for all four positions at the same time for each activity. All the four participants were male, between the ages of 25 and 30.

IV. PREPROCESSING DATA

In our data collection experiments, we first started the app on all four smartphones and then placed it on four body positions of a participant. After the completion of the activity, we removed them from participant's body and stopped the app. This caused some noise (abnormal spikes) at the start and end of each activity data. These noisy parts were removed before analyzing the data.

Then we divided our collected data into small segments for feature extraction using sliding window approach. The selection of an appropriate window size is important and different values can be set for it. However, we selected 2 second sliding window based on previous studies [4][9], because they have shown a window size of 2 seconds to be an effective and sufficient value for a reasonable activity recognition performance. We used sliding window approach with 50% overlap (1 second overlap here) based on previous studies [4][9][10]. Though different overlap values can be used but 50% overlap value has been shown to have produced reasonable results [4][9][10].

Each sensor reports values along its three dimensions, i.e. along x-axis, y-axis, and z-axis. For example, the accelerometer reports acceleration in meter per second squared (m/sec^2), the magnetometer reports the magnetic field in micro tesla (µT), and the gyroscope reports the rate of rotations in radians per second (rad/sec) along each axis. The orientation of a smartphone affects the performance (accuracies) of classification algorithms because sensors like an accelerometer changes its value with respect to smartphone's orientation [11. Most of the existing work assume fixed orientation while evaluating different classification algorithms [4]. That is why we added a fourth dimension to the existing three dimensions of each sensor, called magnitude of a sensor. The reason behind this addition is that magnitude is orientation-insensitive unlike other three axis of an accelerometer and a gyroscope. This choice was motivated by the work done in [11] about orientation-independence in activity recognition. The magnitude for each sensor was calculated using the following formula:

$$\text{Magnitude} = \sqrt{(x^2 + y^2 + z^2)}$$

Now we have four dimensions for each sensors i.e., (x, y, z, magnitude). For each sliding window with 50% overlap, we extracted two time domain features: mean, and standard deviation for all four dimensions of each sensor. So in total, we calculated 4 x 2 = 8 features per sensor and 8 x 3 = 24 features for all sensors. We selected these time domain features because they are computationally cheap as compared to frequency domain features due to the expensive Fourier transformation. The computationally cheapness of the time domain features has been argued in [5] as well. Moreover, we wanted to start from a simple possible situation. That is why we did not normalize our features as well. The process was kept as simple as possible for assessing the role of these sensors in the activity recognition process. Future studies can build on top of this study to explore this comparative analysis further in a detailed way. We describe some of the possible future studies in this regard in section VII.

V. EXPERIMENT DESIGN

In order to analyze the preprocessed data, we used WEKA machine learning tool (Waikato Environment for Knowledge Analysis) [12]. There are different classification and preprocessing algorithms available in this tool. The preprocessed data (extracted features) was converted to ARFF (Attribute-Relation File Format), which is suitable for WEKA. Then we applied different classifiers on these data to see their performance. We used 10-fold stratified cross validation technique to evaluate different classifiers. In stratified cross validation, each fold or part of data contains all classes in equal proportion to ensure fairness [12]. We evaluated seven commonly used classifiers (in activity recognition's state of the art) as listed in Table I. We will use the short notations in Table I for these classifiers in all next sections.

TABLE I. CLASSIFIERS SUMMARY

Type of Classifier	WEKA-version	Notation
Naïve Bayes	Naïve Bayes	NB
Support vector machines	LibSVM	LSM
Neural Networks	MulitlayerPerceptron	MLP
Logistic Regression	Logistic Regression	LR
K Nearest Neighbor	IB1 (KNN with K=1)	IB1
Rule Based Classifiers	Decision Tables	DT
Decision Trees	J48	J48

We used all these classifiers in their default settings in WEKA 3.7.9. We did not do any optimizations because we are more interested in the relative roles of our three sensors in the classification process. This also means that reported absolute accuracies may not be the best possible ones and may be further improved.

We used the WEKA experimenter environment as it enables us to repeat different experiments multiple times with different random seeds to make it statistically reliable. Using this tool, we evaluated each classifier on our dataset ten times (each time using stratified 10-fold cross validation). We only report the average results of all the 10 repetitions here.

Unlike previous studies, we created two data sets. One was our original dataset with imbalance class distribution as shown in Table II. We created the second dataset with balanced class distribution where all classes have the same number of examples as shown in Table III. We created the latter dataset because many classifiers assume balanced class distribution and can therefore produce biased results with a dataset having an imbalanced class distribution [13]. We wanted to avoid this imbalance effect. So we removed examples from the majority classes and made it equal to the class having the smallest number of examples. In our case, walking downstairs activity has the smallest number of examples so other classes' sizes were set to the same number. The second reason is that WEKA reports its results by taking weighted (by class size) average of the individual classes' performance metrics, which can mislead the overall results if class sizes are not equal. As majority class has more effect on the overall performance when weighted average is taken, it can affect the overall results where achieving highest accuracy is the main goal.

We did evaluations on datasets with both balanced and imbalanced class distribution, but we report results only for dataset with balanced class distribution because they are fair and accurate. The results for dataset with imbalance class distribution also followed the same trends in many cases except that the absolute accuracy values for classifiers were different and there were few exceptions. This difference was observed both in overall as well as in individual accuracies for all six activities. As we are concerned about the relative roles of our three sensors, the absolute values are not important for us. We do not discuss these differences here because they are out of scope of this paper. Moreover, dataset with balanced class distribution gives more accurate and fair results, we only discuss them in this paper.

TABLE II. IMBALANCED CLASS DISTRIBUTION

Activity	Arm	Belt	Pocket	Wrist
Walk Downstairs	**368**	**374**	**381**	**364**
Running	569	588	586	573
Sitting	600	600	600	600
Standing	600	600	600	600
Walk Upstairs	435	438	446	433
Walking	600	600	600	600

TABLE III. BALANCED CLASS DISTRIBUTION

Activity	Arm	Belt	Pocket	Wrist
All six activities	368	374	381	364

VI. PERFORMANCE ANAYLSIS AND DISCUSSION

In this section, we discuss the performance analysis of an accelerometer, and a gyroscope, both when they are used standalone as well in combination with each other. Moreover, we discuss the role of a magnetometer in activity recognition in comparison with a gyroscope and an accelerometer.

A. The role of an accelerometer and a gyroscope

We evaluate the accelerometer and the gyroscope on four body positions using seven commonly used classifiers. We use accuracy or True Positive Rate (TPR) as our performance metric. The TPR of a classifier means the amount of correctly classified examples of a specific class out of its all examples. The overall TPR of a classifier is the weighted (by class size) average of individual TPR for all classes being recognized. We show overall TPR results in Fig. 1-4 respectively for all seven classifiers on the four body positions. These results give an overview of how different classifiers behave in terms of their classification accuracies or TPR with a gyroscope, an accelerometer and their combination. It is important to note that all these results are for the dataset with balanced class distribution. The trends in these results for each situation is summarized in the caption each figure as shown in Fig. 1-4.

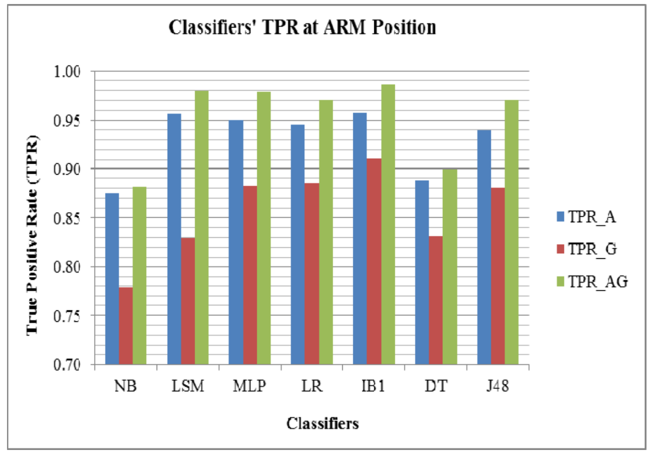

Fig. 1. Overal TPR results for an accelerometer (A), a gyroscope(G) and their combination (AG): Overall, an accelerometer performs better than a gyroscope. However, when both are used in combination, the results are further improved than their individual resuts.

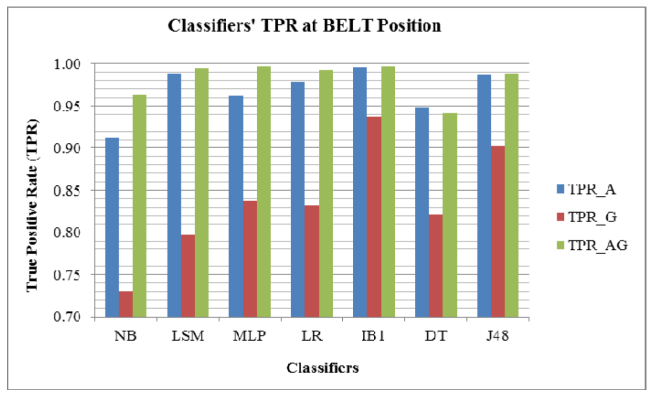

Fig. 2. Overall TPR results: the accelerometer performs better than the gyroscope whereas their combination perform better than their individual TPR. Their combination improves the gyroscope performance more than that of the accelerometer.

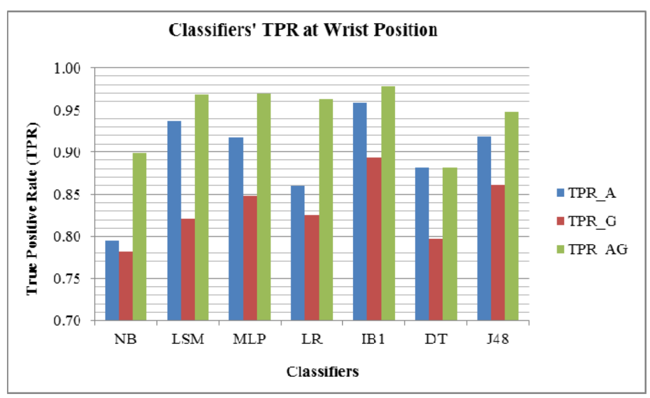

Fig. 3. Overall TPR results: it almost follows the same pattren as in Fig. 1 and Fig. 2 except that the improvement done by the combiniation the two sensors, gets relatively better compared to their individual TPR.

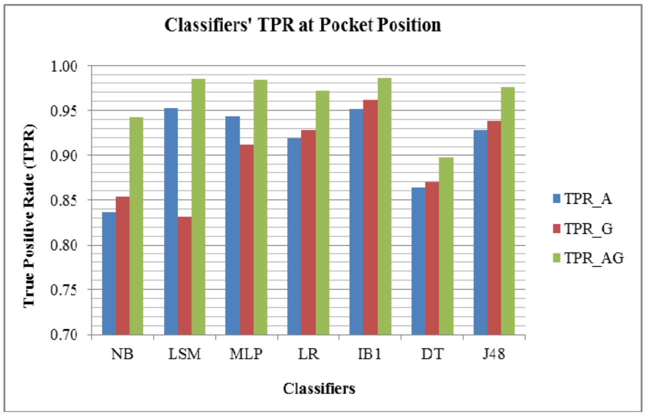

Fig. 4. Overall TPR results: the gyroscope perfoms slightly better than the accelerometer for all classifiers except for LSM and MLP. However, their combination performs better than their individual TPR for all classifiers.

It is clear from Fig. 1-4, that in terms of overall TPR, the accelerometer performs better than the gyroscope in recognizing our six activities for all positions except pocket. In pocket position, the gyroscope performs slightly better than the accelerometer for all classifiers except LSM and MLP. It is interesting to see that the gyroscope performs better with simple classifiers like DT, IB1 and NB.

The combination of the gyroscope and the acclerometer performs better than their individual performances in almost all cases except for DT at belt and pocket position. The improvements done by their combination are overall higher than the gyroscope's individual TPR compared to that of the accelerometer. Moreover, the amount of this improvement varies at different positions with being highest at pocket position and lowest at belt position. In order to further analyze these two sensors, we assess their roles in recognizing individual activities. We evaluate these sensors on all four body positions but report the results for pocket position only as shown in Fig. 5-10. For the other three positions, we only summarize the results due to the limited space.

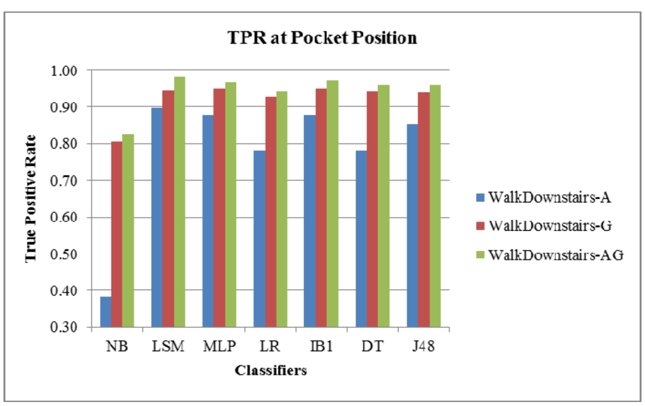

Fig. 5. The walking downstairs activity is better recognizd by the combination of an accelerometer and a gyroscope compared to their individual recognition. Morover, the gyroscope performs better than the accelerometer here.

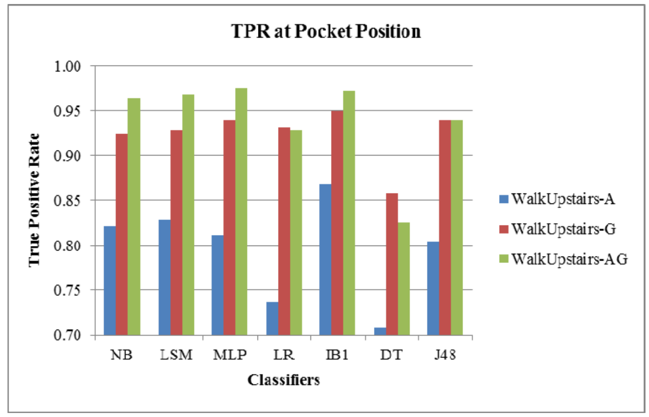

Fig. 6. For walking upstairs activity, the gyroscope performs better than the accelerometer. Moreover, their combination improves overall performance, especially with respect to the accelerometer.

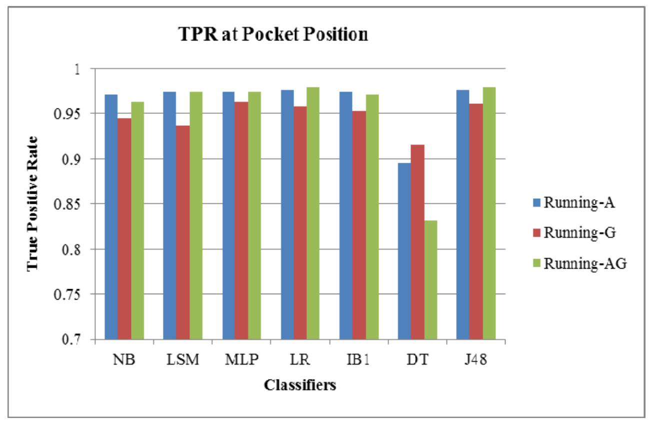

Fig. 7. For running activity, the acclerometer performs slightly better than the gyroscope except for the DT classifier. Their combination does not bring any significant improvents here with respect to the accelerometer but it does bring slight improvements with respect to the gyroscope TPR.

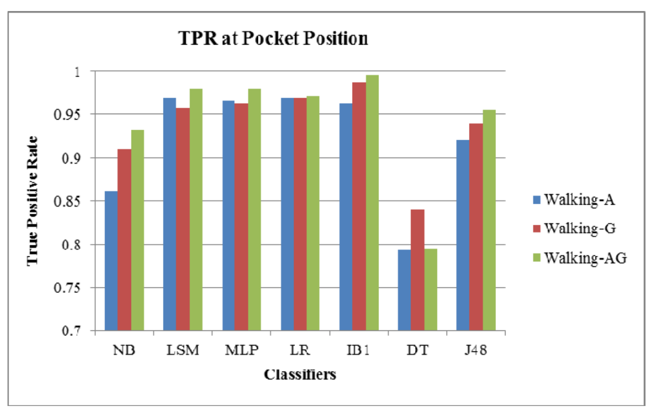

Fig. 8. For walking activity, the gyroscope performs slightly better than or equal to the acclerometer except for the LSM. Their combination improves overall results except for the DT classifier.

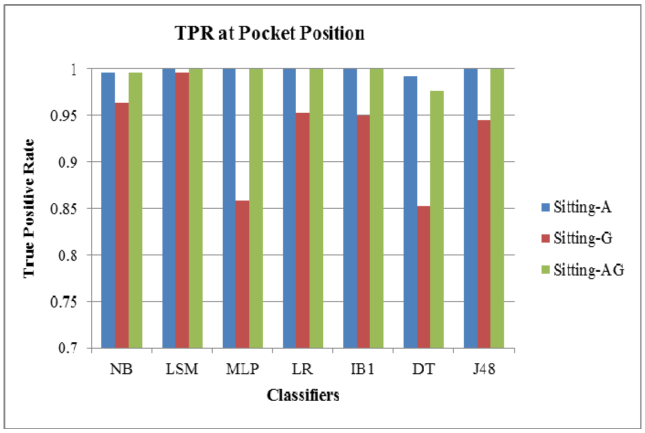

Fig. 9. The acclerometer performs better than a gyroscope for the sitting activity.

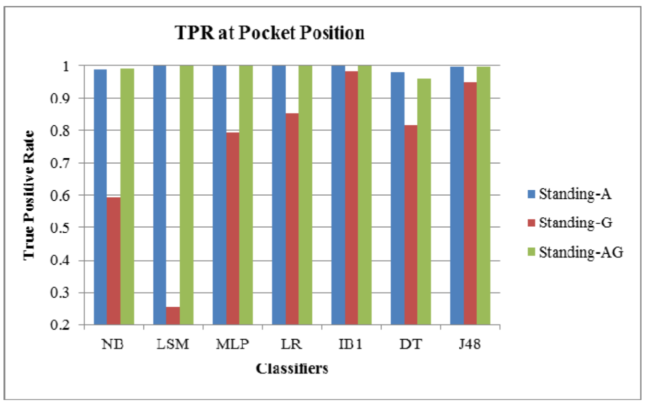

Fig. 10. For standing activity, the gyroscope performs poorly as compared to the accelerometer. For example for LSM, its TPR is just .25 which is too low. These low values contribute to the overall low TPR for a gyroscope in some situations. Moreover, the improvements by the combination is significant only with respect to the gyroscope.

It is clear from Fig. 5-10 that walking downstairs and upstairs activities are better recognized by the gyroscope than the accelerometer for all classifiers though the accelerometer also produces reasonable results. This is because the gyroscope produces better pattern in its raw data for these two activities compared to that of the accelerometer. The sitting and standing activities are better recognized by the accelerometer and the gyroscope perform very poorly for some classifiers (NB, LSM, MLP, and DT) because these classifiers confuse sitting and standing activities with each other. However, the root cause is the less recognizable patterns in the raw data of a gyroscope. These low individual TPR values for sitting and standing in turn contribute to low overall TPR values for a gyroscope as we noticed in Fig. 1-4. This becomes clearer from Table IV with low TPR values for standing activity on the remaining three positions (arm, wrist and belt). We also looked into the confusion matrices for these two activities and noticed that the standing activity was mainly confused with sitting activity by almost all classifiers on all four positions (for gyroscope only). The running activity is better recognized by the accelerometer but the gyroscope also produces reasonable results. The walking activity is sometime better recognized by the accelerometer and sometime by a gyroscope so it is mainly dependent on the choice of the classifier. The combination of a gyroscope and an accelerometer perform almost always better than their individual performances for all six activities. Their combined TPR values for an activity are always higher than or equal to the maximum of their individual TPR values except for DT classifier. This is an important result and supports the idea of using both sensors in combination with each other in order to better recognize these six activities. For example, the gyroscope can help the accelerometer in recognizing better the walking upstairs and downstairs activities. The accelerometer can help the gyroscope in differentiating between sitting and standing activities in a better way.

We evaluate these two sensors in the remaining three positions as well. For arm and wrist position, we observe almost the same trends in results as were observed for the pocket position. The results at belt position behave differently because the accelerometer performs better than the gyroscope for all six activities. We suspect that this low performance at belt position by the gyroscope might be due to the smartphone's relatively fixed position in a belt clipper or its landscape orientation at belt position unlike at other three positions (portrait orientation) but this needs further research.

Based on these results, we can now argue that the reported improvements in [4] are due to the pocket position and KNN classifier. The reason behind the claim of no additional value by a gyroscope in [5] could be due to the fact that they use many positions (handbag, hand, jeans pocket, and shirt pocket) in one dataset and by looking at overall TPR of their three classifiers. As we show that the performance of each sensor depend on the position of the smartphone, then using one dataset for all positions may not give a clear picture of these sensors' individual roles in the activity recognition. Moreover, for a clear picture of the individual sensors' roles activity-wise analysis in recognition process is required.

TABLE IV. TPR FOR STANDING ACTIVITY USING GYROSCOPE

Position	NB	LSM	MLP	LR	IB1	DT	J48
Arm	0.38	0.38	0.73	0.66	0.83	0.80	0.80
Belt	0.26	0.14	0.42	0.43	0.88	0.54	0.82
Wrist	0.22	0.28	0.57	0.45	0.74	0.61	0.71

B. The role of a magnetometer

We assess the magnetometer's role both individually as well as in combination with an accelerometer and a gyroscope. However, we show only the individual results here because they were not encouraging. We argue that its low performance is caused by overfitting the training process due to the lack of proper pattern in its raw data. For pocket position, we report its results in Fig. 11-14. These results are poor, especially knowing the fact that main contribution towards the overall perfomrnace is by two activities only: sitting and standing as shown in Table V. The magnetometer performs poorly for the remaining four activites: walking upsatairs and downstairs, walking, and running. We evaluated the magnetometer on arm, belt, and wrist as well. Its results were poor for all activities except sitting and standing.

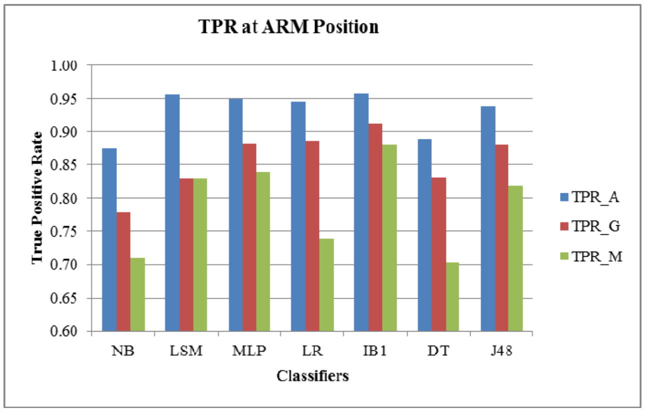

Fig. 11. Overall TPR values for the magnetometer are the lowest.

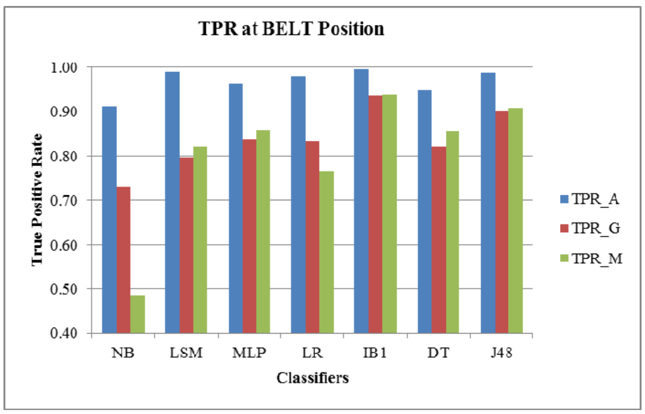

Fig. 12. Overall TPR values for the magnetometer are the low.

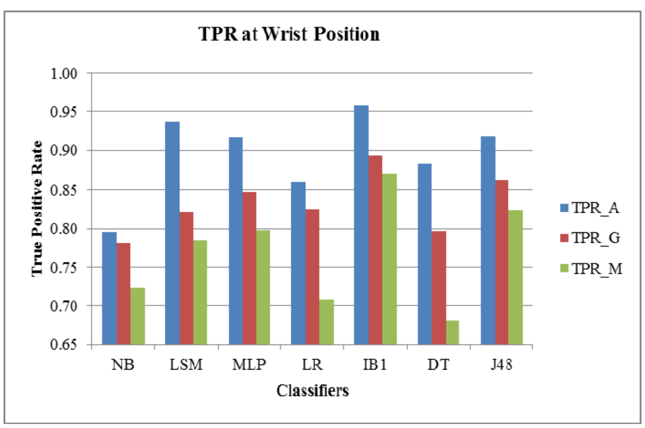

Fig. 13. Overall TPR values for the magnetometer are the lowest.

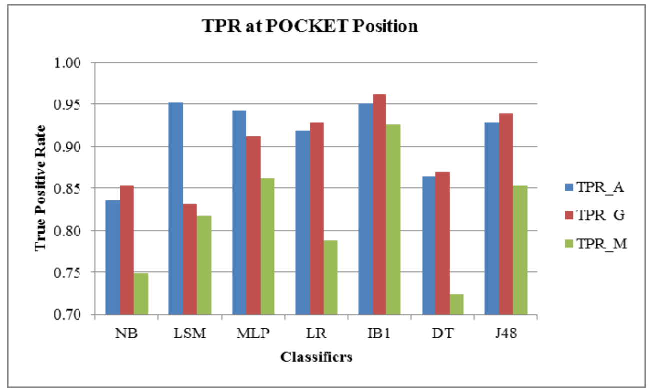

Fig. 14. Overall TPR values for the magnetometer are the lowest.

TABLE V. TPR FOR MAGNETEMETR AT POCKET POSITION

Activities	NB	LSM	MLP	LR	IB1	DT	J48
Walk Downstairs	0.69	0.98	0.74	0.74	0.87	0.69	0.75
Running	0.37	0.78	0.80	0.59	0.9	0.55	0.80
Sitting	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.98**	**0.99**
Standing	**0.97**	**0.97**	**0.98**	**0.99**	**0.99**	**0.92**	**0.99**
Walk Upstairs	0.58	0.24	0.70	0.56	0.87	0.49	0.71
Walking	0.85	0.80	0.91	0.79	0.92	0.74	0.88

Fig. 15 shows the magnetometer values for a walking activity. In this activity, we asked a participant to walk from a point A to a point B in a specific direction, and then come back towards the point A. The first half with positive M-z values represent walk from A to B, and the second half with negative M-z values represent the walk from B to A. The direction-dependence (shown in Fig. 15) needs to be taken into accout in the preprocessing to present smooth patterns to the classifiers. One solution could be to take absulte values of the raw magnetometer data before extracting features from it as shown in Fig. 16.

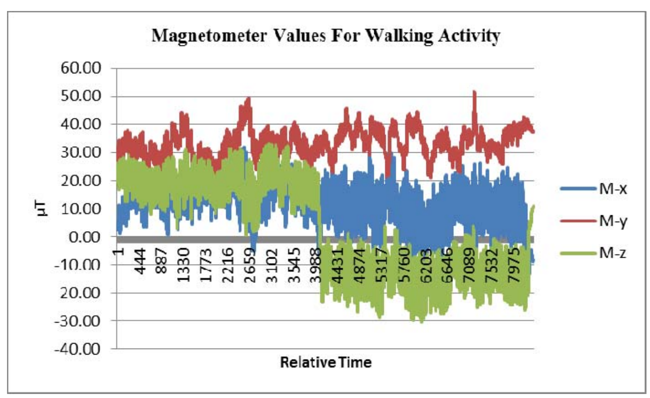

Fig. 15. All three axis values for magnetmeter while a participant walks in a corridor from point A to B and then returns from point B to A in a corridor.

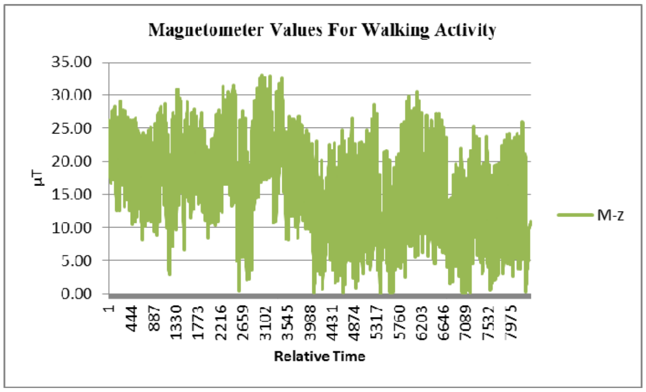

Fig. 16. Z-axis after taking its absolute value present a better data pattern.

Clearly, it (Fig.16) has better pattern compared to the one in Fig. 15. However, improving the magnetometer's raw data for better classification is out of scope for this paper and we leave it as future work. We conclude that a magnetometer data does not help the recognition process in its raw form and causes over fitting for classifiers. However, if preprocessed correctly by taking into account its dependence on direction, it can be used in the activity recognition process.

VII. CONCLUSIONS AND FUTURE WORK

We evaluated the smartphone accelerometer, gyroscope, and magnetometer using seven commonly used classifiers on four body positions. We used time domain features with these classifiers. Based on our evaluations, we showed that in most cases, an accelerometer and a gyroscope complement each other in the activity recognition process. Moreover, a gyroscope not only improves the recognition performance in combination with an accelerometer but it also produces reasonable performance results when used alone. They perform better than each other for different activities in different situations. For example, the walking upstairs activity is better recognized by a gyroscope except at belt position. The standing activity is better recognized by an accelerometer at all four positions. Their combination improves the overall TPR or at least keep it equal to the maximum of their individual TPR in almost all situations with very few exceptions. The magnetometer's role in activity recognition was poor and may be improved if direction-insensitive features are extracted for it. Based on our evaluations, we conclude that it is difficult to make an exact generic statement about the role of these sensors in the activity recognition process for all situations because their roles depend on the position of the smartphone, the activity being recognized, and the choice of the classifiers. However, we can make statements about their roles in a defined situation.

This work was part of our planned work [14] which we intend to explore further in future. Moreover, these results can be verified for an extended set of activities and with different sliding windows. Similarly, it can be validated on an extended set of classifiers and with a bigger data set. It can also be validated on a different set of features, for example, frequency domain features only. Moreover, the feasibility of a magnetometer in activity recognition can be studied further.

ACKNOWLEDGMENT

This work is partly supported by Dutch National Program COMMIT in the context of SWELL project and partly by an InterReg project, WHM (wireless health monitoring). We would also like to thank our participants for contributing to the science by taking part in our data collection experiments voluntarily.

REFERENCES

[1] Klasnja, Predrag, and Wanda Pratt. "Healthcare in the pocket: Mapping the space of mobile-phone health interventions." *Journal of biomedical informatics* 45, no. 1 (2012): 184-198.

[2] Lane, Nicholas D., Mashfiqui Mohammod, Mu Lin, Xiaochao Yang, Hong Lu, Shahid Ali, Afsaneh Doryab, Ethan Berke, Tanzeem Choudhury, and Andrew Campbell. "BeWell: A smartphone application to monitor, model and promote wellbeing." In *5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth2011)*. 2011.

[3] Kraaij, W. COMMIT Projectplan, Workpackages, Deliverables, Budget User Centric Reasoning for Well-working (P07). TNO, Project planning, 2011.

[4] Wu, Wanmin, Sanjoy Dasgupta, Ernesto E. Ramirez, Carlyn Peterson, and Gregory J. Norman. "Classification accuracies of physical activities using smartphone motion sensors." *Journal of medical Internet research* 14, no. 5 (2012): e130.

[5] Anjum, Alvina, and Muhammad U. Ilyas. "Activity recognition using smartphone sensors." In *Consumer Communications and Networking Conference (CCNC), 2013 IEEE*, pp. 914-919. IEEE, 2013.

[6] Incel, Ozlem Durmaz, Mustafa Kose, and Cem Ersoy. "A review and taxonomy of activity recognition on mobile phones." *BioNanoScience* (2013): 1-27.

[7] Martín, Henar, Ana M. Bernardos, Josué Iglesias, and José R. Casar. "Activity logging using lightweight classification techniques in mobile devices." *Personal and Ubiquitous Computing* 17, no. 4 (2013): 675-695.

[8] Samsung Galaxy S2 specifications available at http://www.samsung.com/global/microsite/galaxys2/html/ [Last accessed on 24th July 2013].

[9] Preece, Stephen J., John Yannis Goulermas, Laurence PJ Kenney, and David Howard. "A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data." *Biomedical Engineering, IEEE Transactions on* 56, no. 3 (2009): 871-879.

[10] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," *Pers Comput.,* Lecture Notes in Computer Science, vol. 3001, pp. 1–17, 2004.

[11] Sun, Lin, Daqing Zhang, Bin Li, Bin Guo, and Shijian Li. "Activity recognition on an accelerometer embedded mobile phone with varying positions and orientations." In*Ubiquitous intelligence and computing*, pp. 548-562. Springer Berlin Heidelberg, 2010.

[12] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); "The WEKA Data Mining Software: An Update"; *SIGKDD Explorations*, Volume 11, Issue 1.

[13] Japkowicz, Nathalie, and Shaju Stephen. "The class imbalance problem: A systematic study." *Intelligent data analysis* 6, no. 5 (2002): 429-449.

[14] Shoaib, Muhammad, "Human Activity Recognition Using Hetrogenious Sensors" "*Appears in the Adjunct Proceedings of UbiComp (Doctoral School)* Sep 8-12, 2013.

[15] Our activity recognition dataset and data collection app is available at http://ps.ewi.utwente.nl/Datasets.php.