A Comparative Study on Sensor Displacement Effect on Realistic Sensor Displacement Benchmark Dataset

Lumpapun Punchoojit and Nuttanont Hongwarittorrn

Department of Computer Science
Faculty of Science and Technology, Thammasat University, Thailand
l.punchoojit@gmail.com, nth@cs.tu.ac.th

Abstract. Activity Recognition (AR) research is growing and plays a major role in various fields. The approach of using wearable sensors for AR is well-accepted, as it compensates the need to install cameras in image processing approach which can lead to privacy violation. Using wearable sensors can suffer from one disadvantage – sensor displacement. There have been a number of research which studies sensor displacement problem. However, the conclusion cannot be made as which classifier is better than another in recognizing displacement data, as the prior experiments were performed under different conditions and focused on different parts of the body. This work aims to evaluate recognition performance of different algorithms – SVM, C4.5, and Naïve Bayes – on ideal-placement and displacement data on whole body activities, by adopting REALDISP dataset to make such evaluation. The accuracy of all algorithms on ideal placement data was above 90%, where SVM yielded the highest accuracy. Displacement data were tested against classification models constructed from ideal-placement data. The results shows that there was a dramatic drop in recognition performance. The accuracy of all algorithms on displacement data was between 50-60%, and C4.5 could handle displacement data the best.

Keywords: Activity Recognition, Sensor Displacement, Naïve Bayes, SVM, C4.5.

1 Introduction

Activity Recognition (AR) is an active research area in recent years. It plays a major role in diverse applications. In pervasive healthcare, AR supports preventive or chronic healthcare, cognitive assistance, and elderly monitoring [1, 2]. In security, AR supports the identification of terrorist actions and threats, or monitoring individual's activities in security sensitive areas, such as hospitals, and banks [3]. On mobile devices, AR supports user's activities monitoring, and enabling screen rotation [4].

According to Ugulino et. al, there are two approaches commonly used for activity recognition: image processing and the use of wearable sensors [1]. Image processing approach does not require users to put an equipment on their bodies; however, this approach encounters some limitations, including the requirement to install cameras, and image quality which can be affected by environmental conditions, such as lighting [1]. The installation of cameras may controversially violate user's privacy

H. Unger et al. (eds.), *Recent Advances in Information and Communication Technology 2015,*
Advances in Intelligent Systems and Computing 361, DOI: 10.1007/978-3-319-19024-2_10

[1]. The second approach for AR is the use of wearable sensors. This approach compensates the limitations of image processing approach [1]. Even though this approach requires users to wear the equipment through a period of time, general public is more likely to accept it [1], as the equipment can be easily turned-off or removed [2]. Wearable sensors also offer real-time activity information for AR [2].

Despite having many strengths, using wearable sensors for activity recognition can suffer from one major disadvantage – sensor displacement. The implementation of wearable sensors requires the equipment to be attached at predefined positions, and the classification model is built by assuming constant sensor positions [5]. However, this can hardly be maintained in real-life condition. As a result, the model may fail to classify the activities from the sensor data. Misclassification may lead to unwanted consequence; for instance, it could be very dangerous in elderly pervasive healthcare system if the classifier failed to recognize elderly falling. Displacement of sensors can be caused by either sensor loose fitting or displacing by the users themselves [5].

There have been a number of research which studies the sensor displacement; some of them proposes heuristics to improve classifier's robustness. The research may, however, focus on different parts of the body or a different set of activities. For instance, Chavarriaga et. al [6] study the effects of sensor displacement and propose an unsupervised adaptive classifier to tackle the problem. The experiment in this study focused on three activity scenarios: HCI gestures, fitness, and daily living scenarios. The sensors were installed on an participant's arm in HCI gesture scenario, on a leg in fitness scenario, and on both arms and the back in daily living scenario [6]. Kunze and Lukowicz investigate the effect of displacement on onbody activity recognition systems [7]. They presented a set of heuristics which increase the robustness of the recognition with the respect of sensor displacement [7]. In the experiment, they focused on locomotion activities (e.g. walking, running, walking up hill, biking, etc) and gym exercise activities (e.g. shoulder press, arm extension, arm curl, etc). The sensors were installed participant's legs in locomotive activity experiment, and the sensors were installed on the arms in gym exercise experiment [7].

Regarding the previous studies, the conclusion cannot be made as which classifier is better than others for recognizing sensor displacement data, as the experiments were performed under different conditions and were focusing on different parts of the body or different sets of activities. In addition, none of the studies has explicitly described performance degradation of the classifier on ideal-placement and displacement data. Therefore, this current work aims to examine performance of three popular recognition algorithms – SVM, C4.5, and Naïve Bayes – on ideal-placement and displacement data on whole body activities, by adopting REALDISP dataset to make such evaluation. The main objective is to evaluate which of the algorithms would outperform others on the displacement problem.

2 Related Works

2.1 REALDISP Dataset

REALDISP (REAListic sensor DISPlacement) dataset lends itself as a benchmark dataset for activity recognition, whether in ideal, real-life, or extreme displacement conditions.

The dataset was collected to investigate the effects of sensor displacement in activity recognition [8], which can be either caused by loose fitting of sensors or displacement by users themselves. This dataset was created by Banos et. al [5, 9]. The dataset was built upon three scenarios: ideal-placement, self-placement, and induced-displacement [8]. The dataset covers a wide range of physical activities and locations of wearable sensors [9].

Ideal-Placement data was generated when the sensors were positioned to predefined locations by the instructor (i.e. research team). The data from ideal-placement can be considered as the training set for the recognition model [10]. Self-Placement data was induced when the users place the sensors on their body parts specified by the research team. Data from self-placement may slightly differ from the ideal-placement one; however, the difference is considered to be too trivial [10]. Induced-Displacement occurred when the sensors were misplaced by rotations or translations with respect to the ideal setups. In REALDISP dataset, the induced-displacement data was generated by intentionally displacement of sensors by the research team [10].

Activity data were collected from 17 subjects. Thirty-three physical activities were included in the dataset, as listed in table 1.

Table 1. Activity Set

#	Activity	#	Activity	#	Activity
1	Walking	12	Waist rotation	23	Shoulders high amplitude rotation
2	Jogging	13	Waist bends	24	Shoulders low amplitude rotation
3	Running	14	Reach heels backwards	25	Arms inner rotation
4	Jump Up	15	Lateral bend	26	Knees to the breast
5	Jump Front & Back	16	Lateral bend arm up	27	Heels to the back side
6	Jump Sideways	17	Repetitive forward stretching	28	Knees bending
7	Jump legs/arms opened/closed	18	Upper trunk and lower body opposite twist	29	Knees bend forward
8	Jump rope	19	Arms lateral elevation	30	Rotation on the knees
9	Trunk twist (arms outstretched)	20	Arms frontal elevation	31	Rowing
10	Trunk twist (elbows bended)	21	Frontal hand claps	32	Elliptic bike
11	Waist bends forward	22	Arms frontal crossing	33	Cycling

The measurement of the whole body was measured in activities 1 – 3, 5 – 8, and 31 – 33 [10]. Activities focused on trunk were measured in activity 9 – 18, upper extremities in 19 – 25, and lower extremities in activity 26 – 29 [10]. All activities were measured by 9 sensors; each of which measured four sensor modalities: acceleration, rate of turn (gyroscope), magnetic field, and orientation [9]. Each sensor provided tri-directional measurements; except for orientation that was measured in quaternion format [10]. The sensors were installed on nine different parts of the subject's body:

1) left calf, 2) left thigh, 3) right calf, 4) right thigh, 5) back, 6) left lower arm, 7) left upper arm, 8) right lower arm, and 9) right upper arm [9]. Altogether, this makes up 117 attributes.

2.2 Recognition Algorithms

This section gives an overview of algorithms used for the comparison in this paper. The algorithms are selected based on the survey of activity recognition algorithms conducted by Ugulino et. al [1]. They are three algorithms which are commonly used for recognition tasks: SVM, C4.5, and Naïve Bayes.

SVM

SVM (Support Vector Machines) is a supervised learning algorithm used for binary classification of both linear and non-linear data [11]. When data are linearly separable, SVM would search for maximum marginal hyperplane, or a decision boundary that best separate the tuples of one class from another. Hyperplane with larger margin is expected to be more accurate in classification [11]. When data are not linearly separable, SVM uses a nonlinear mapping to transform the original data into a higher dimension feature space. Then, SVM searches for a linear separating hyperplane in the new feature space [11].

C4.5

C4.5 is an algorithm for decision tree induction, presented in the year 1993 [12]. C4.5 uses top-down approach to construct a classification model. It starts with all training tuples at the root node of the tree, then an attribute would be selected to partition the tuples [13]. The process would be repeated as the tree is being built [11].

An attribute selection measure is required to select the attribute that best split the tuples [11]. Attribute selection measure is a heuristic for selecting the splitting criterion that best separate the tuples of class-labeled training tuples into individual classes [11]. Specifically, the attribute that yields 'pure' partitions would be selected. A pure partition means that all tuples in that particular partition belong to the same class [11]. In other words, the selected attribute minimizes an information entropy applied to tuple partition [13]. For C4.5, it uses gain ratio as splitting criteria; the splitting would stop when the number of instances to be split is below the threshold.

Naïve Bayes

Naïve Bayes classification, or simple Bayesian Classifier, is in the family of Bayes Classification methods [11]. Classifiers in this family are statistical classifiers; meaning that they can predict membership probabilistic [11]. Naïve Bayes, like other Bayesian Classifiers, is based on Bayes' Theorem [11].

Naïve Bayes presumes conditionally independence of the classes. It determines the probability that an instance would belong to a particular class - posterior probability (i.e. $P(C_i|X)$). Posterior probability is calculated from another three prior probability values: $P(C_i)$, $P(X)$, and $P(X|C_i)$. $P(C_i)$ is the probability that an instance belongs to class C_i, regardless of X. P(X) is the probability that an instance has attribute values

X, regardless of C_i. $P(X|C_i)$ is the probability that an instance has attribute values X, given an instance belongs to class C_i [11]. $P(C_i|X)$ is calculated by the following equation:

$$P(Ci|X)=\frac{P(X|Ci)P(Ci)}{P(X)} \tag{1}$$

3 Comparative Experiments

3.1 Methodology

Although REALDISP contains activity data from 17 subjects, only some subjects had the data on both ideal-placement scenario and displacement scenario. In this paper, data from subject number 2 were selected for the analysis, as the data on both scenarios of this subject were available.

In this paper, only whole body activities were selected for the analysis. There were altogether 10 activities, including: 1) Walking, 2) Jogging, 3) Running, 4) Jump front and back (jump 1), 5) Jump Sideways (jump 2), 6) Jump legs/arms open/closed (jump 3), 7) Jump rope (jump 4), 8) Rowing, 9) Elliptic bike, and 10) Cycling.

Ideal-placement data were used to construct classification models by using classification algorithms described in section 2.2. Ten-fold cross validation was employed in every model construction. Sensor displacement dataset was used to test against the constructed models, to examine the effect of sensor displacement on activity recognition.

3.2 Results

This section describes the recognition performance of SVM, C4.5, and Naïve Bayes algorithms on ideal-placement data and displacement data. Confusion matrices for recognition performance under each condition for each algorithm are also provided.

Ideal Placement

The recognition performance of each algorithm on ideal-placement data was highly accurate. There were some differences; however, they were very small.

Table 2. Ideal-Placement Recognition Accuracy

SVM	C4.5	Naïve Bayes
99.95%	99.61%	91.73%

The accuracy of SVM is the highest among the three algorithms (99.95%). Accuracy of C4.5 is slightly lower than that of SVM (99.61%). The accuracy of Naïve Bayes is the lowest, compared to the other two algorithms (91.73%). Confusion matrix of each algorithm is elaborated as follows:

Table 3. Confusion Matrix (SVM)

Walking	Jogging	Running	Jump 1	Jump 2	Jump 3	Jump 4	Rowing	E. Bike	Cycling	
15976	0	0	0	0	0	0	0	0	0	Walking
0	**15124**	7	0	0	0	0	0	0	0	Jogging
0	13	**11656**	0	0	0	0	0	0	0	Running
0	0	0	**3603**	6	0	0	0	0	0	Jump 1
0	0	0	12	**3644**	0	0	0	0	0	Jump 2
0	0	0	0	0	**3785**	1	0	0	0	Jump 3
0	0	0	0	0	1	**2022**	0	0	0	Jump 4
0	0	0	0	0	0	0	**8203**	0	0	Rowing
0	0	0	0	0	0	0	0	**10040**	0	E. Bike
0	0	0	0	0	0	0	0	0	**11111**	Cycling

Table 3 describes the confusion matrix of classification results of SVM on ideal-placement data. Recognition accuracy on walking, rowing, elliptic bike, and cycling was 100% accurate. There was some instances in running and jogging activities that were misclassified as one another, and some particular jumping activities that were classified as other types of jumping.

Table 4. Confusion Matrix (C4.5)

Walking	Jogging	Running	Jump 1	Jump 2	Jump 3	Jump 4	Rowing	E. Bike	Cycling	
15975	0	1	0	0	0	0	0	0	0	Walking
2	**15003**	125	0	0	0	0	0	0	0	Jogging
0	131	**11538**	0	0	0	0	0	0	0	Running
0	0	0	**3594**	7	7	1	0	0	0	Jump 1
0	0	2	16	**3636**	2	0	0	0	0	Jump 2
0	0	1	8	4	**3763**	10	0	0	0	Jump 3
0	0	0	2	1	9	**2011**	0	0	0	Jump 4
0	0	0	0	0	0	0	**8203**	0	0	Rowing
0	0	0	0	0	0	0	0	**10040**	0	E. Bike
0	0	0	0	0	0	0	0	0	**11111**	Cycling

Table 4 describes the confusion matrix of classification results of C4.5 on ideal-placement data. Recognition accuracy on rowing, elliptic bike, and cycling was 100% accurate. In walking activity, only one instance was misclassified as running. Similar to the results of SVM, some instances of jogging were classified as running, while some instances of running were classified as jogging. Some instances in a particular jumping activity was misclassified for other jumping activities.

Table 5. Confusion Matrix (Naïve Bayes)

Walking	Jogging	Running	Jump 1	Jump 2	Jump 3	Jump 4	Rowing	E. Bike	Cycling	
15840	136	0	0	0	0	0	0	0	0	Walking
2	**11612**	3519	0	0	0	0	0	0	0	Jogging
0	3191	**8477**	0	1	0	0	0	0	0	Running
0	0	0	**3541**	65	0	3	0	0	0	Jump 1
0	0	0	62	**3588**	6	0	0	0	0	Jump 2
8	0	0	24	7	**3747**	0	0	0	0	Jump 3
0	0	0	0	9	8	**2006**	0	0	0	Jump 4
0	0	0	0	0	0	0	**8194**	9	0	Rowing
0	0	0	0	0	0	0	0	**10040**	0	E. Bike
0	0	0	0	0	0	0	0	0	**11111**	Cycling

Table 5 describes the confusion matrix of classification results of Naïve Bayes on ideal-placement data. Recognition accuracy on elliptic bike, and cycling was 100% accurate. In walking activity, 136 instances were misclassified as running. In Naïve Bayes, many instances of jogging were classified as running, while many instances of running were classified as jogging. Similar to the results from the other two algorithms, some instances in a particular jumping activity was misclassified for other jumping activity. Misclassification in jogging, running, and all jumping activities was similar to that of C4.5.

Displacement

The recognition performance on displacement data dramatically dropped on every algorithm. The recognition accuracy is described in table 6.

Table 6. Sensor Displacement Recognition Accuracy

SVM	C4.5	Naïve Bayes
52.10%	60.78%	56.43%

When the constructed models were tested against sensor displacement data, C4.5 was most accurate (60.78%). The accuracy of Naïve Bayes was at 56.43%. The accuracy of SVM was the lowest (52.10%), even it was the highest when recognizing the ideal-placement one.

Table 7 describes the confusion matrix of classification results of SVM on displacement data. There was no 100% recognition accuracy on any of the activities. Although the misclassification percentage was high, the misclassified instances were not widely spread.

Table 8 describes the confusion matrix of classification results of C4.5 on displacement data. There was no 100% recognition accuracy on any of the activities. Misclassification was very disperse on every activities; except for rowing, which was misclassified for only another two activities.

Table 7. Confusion Matrix (SVM)

Walking	Jogging	Running	Jump 1	Jump 2	Jump 3	Jump 4	Rowing	E. Bike	Cycling	
15999	0	0	0	3190	2263	3858	2672	31	4	Walking
0	**15126**	5	0	2945	1670	5203	3082	0	0	Jogging
0	12	**11699**	58	2870	2621	6438	1390	9	0	Running
2	70	165	**3613**	266	578	1849	536	0	256	Jump 1
73	46	494	9	**3968**	196	1942	329	0	457	Jump 2
0	0	517	6	255	**4509**	1778	885	0	5	Jump 3
0	0	0	0	0	953	**2586**	443	0	0	Jump 4
0	0	0	0	0	1721	0	**13405**	0	0	Rowing
0	0	27	77	4236	63	4533	6426	**10040**	0	E. Bike
0	40	0	1263	200	0	5090	10604	0	**11111**	Cycling

Table 8. Confusion Matrix (C4.5)

Walking	Jogging	Running	Jump 1	Jump 2	Jump 3	Jump 4	Rowing	E. Bike	Cycling	
18388	149	1757	963	287	5657	0	7	481	328	Walking
14	**15281**	3014	0	1318	7947	213	0	244	0	Jogging
9	599	**18012**	0	1198	4481	46	87	579	86	Running
287	451	40	**3757**	38	224	1282	13	804	439	Jump 1
293	396	19	4776	**3660**	353	1198	24	977	117	Jump 2
128	288	731	238	492	**5206**	726	39	0	77	Jump 3
598	11	33	89	0	555	**2412**	10	267	7	Jump 4
0	0	0	0	0	0	0	**8246**	4790	2090	Rowing
226	700	1169	0	29	3720	0	68	**19064**	426	E. Bike
3184	190	5832	0	3	50	5	823	4824	**13397**	Cycling

Table 9. Confusion Matrix (Naive Bayes)

Walking	Jogging	Running	Jump 1	Jump 2	Jump 3	Jump 4	Rowing	E. Bike	Cycling	
16225	3839	958	0	0	0	0	91	6904	0	Walking
0	**14249**	9820	0	4	189	8	12	3749	0	Jogging
0	4072	**17299**	0	9	1049	16	17	2635	0	Running
0	251	1457	**3553**	79	596	2	4	1393	0	Jump 1
0	332	1518	74	**3640**	598	0	1	1351	0	Jump 2
8	47	2927	23	15	**4709**	0	12	184	0	Jump 3
0	67	613	1	47	697	**2009**	1	547	0	Jump 4
0	181	2457	0	0	0	0	**11546**	941	1	Rowing
0	2857	7154	0	0	0	0	7	**15384**	0	E. Bike
0	864	7328	0	0	3	0	2726	6275	**11112**	Cycling

Table 9 describes the confusion matrix of classification results of Naïve Bayes on displacement data. There was no 100% recognition accuracy on any of the activities. Misclassification was very disperse on every activities.

4 Conclusion and Future Works

4.1 Conclusions

Activity Recognition (AR) plays a major role in various fields; for instance, pervasive health care, security, and wearable and mobile devices. The approach of using wearable sensors for AR is well-accepted, as it compensates the need to install cameras in image processing approach which can lead to privacy violation. Using wearable sensors for activity recognition can suffer from one disadvantage – sensor displacement. There have been a number of research that studies sensor displacement problem. However, the conclusion cannot be made as which classifier is better than another in recognizing displacement data, as the prior experiments were performed under different conditions and focused on different part of the body. This paper showed the recognition performance evaluation of SVM, C4.5, and Naïve Bayes – on ideal-placement and displacement data on whole body activities, by adopting REALDISP dataset to make such comparison.

Ideal placement data were employed to construct classification models. Displacement data were tested against the models, to examine the effects of sensor displacement on classification accuracy of each algorithm.

Recognition performance of all algorithms on ideal placement data was highly accurate. The accuracy of SVM on this dataset was at 99.95%, which was the highest among the three algorithms. Recognition accuracy of C4.5 was at 99.61%, while that of Naïve Bayes was at 91.73%. On ideal placement dataset, some misclassification patterns can be spotted. On jogging and running data, these two activities were misclassified for one another. There were also some misclassification occurred among the four jumping activities.

When the classification models constructed from ideal placement data were tested against displacement data, recognition performance dropped dramatically. Accuracy of SVM, C4.5, and Naïve Bayes decreased to 52.10%, 60.78%, and 56.43%, respectively. Although the accuracy of SVM was the lowest, the misclassified instances were not widely spread like C4.5 and Naïve Bayes.

In sum, the results of the current work illustrates how recognition accuracy can suffer from sensor displacement, and also which algorithm is the most robust one in handling displacement data. The results can lead to further improvement on recognition algorithm in dealing with sensor displacement data.

4.2 Future Works

Some future works may include the investigation on activities related to specific parts of the body (e.g. trunk, upper extremities, and lower extremities), as this work focused only on whole-body activities. Some other recognition algorithms can be employed to examine the effect of sensor displacement on recognition accuracy. Further investigation can also be made on the activities that were likely to be classified for one another (e.g. jogging and running, and the set of jumping activities).

References

1. Ugulino, W., Cardador, D., Vega, K., Velloso, E., Milidiú, R., Fuks, H.: Wearable Computing: Accelerometers' Data Classification of Body Postures and Movements. In: Barros, L.N., Finger, M., Pozo, A.T., Gimenénez-Lugo, G.A., Castilho, M. (eds.) SBIA 2012. LNCS (LNAI), vol. 7589, pp. 52–61. Springer, Heidelberg (2012)
2. Choudhury, T., Consolvo, S., Harrison, B., Hightower, J., LaMarca, A., LeGrand, L., Rahimi, A., Rea, A., Borriello, G., Hemingway, B., Klasnja, P., Koscher, K., Landay, J.A., Lester, J., Wyatt, D., Haehnel, D.: The Mobile Sensing Platform: An Embedded Activity Recognition System. Journal of IEEE Pervasive Computing 7(2), 32–41 (2008)
3. Muncaster, J., Ma, Y.: Activity Recognition using Dynamic Bayesian Networks with Automatic State Selection. In: IEEE Workshop on Motion and Video Computing, pp. 30–37. IEEE Computer Society, Washington, D.C. (2007)
4. Kwapisz, J.R., Weiss, G.M., Moore, S.A.: Activity Recognition using Cell Phone Accelerometers. In: Sensor KDD 2010, pp. 74–82. ACM, New York (2011)
5. Banos, O., Damas, M., Pomares, H., Rojas, I., Toth, M.A., Amft, O.: A Benchmark Dataset to Evaluate Sensor Displacement in Activity Recognition. In: Proceeding of the 14th International Conference on Ubiquitous Computing, pp. 1026–1035. ACM, New York (2012)
6. Chavarriaga, R., Bayati, H., Millan, J.R.: Unsupervised Adaptation for Acceleration-Based Activity Recognition: Robustness to Sensor Displacement and Rotation. Pers. Ubiquit. Comput. 17, 479–490 (2014)
7. Kunze, K., Lukowicz, P.: Dealing with Sensor Displacement in Motion-Based Onbody Activity Recognition Systems. In: Proceeding of the 10th International Conference on Ubiquitous Computing, pp. 20–29. ACM, New York (2008)
8. Banos, O., Tóth, M.A., Amft, O.: REALDISP Activity Recognition Dataset, http://archive.ics.uci.edu/ml/datasets/REALDISP+Activity+Recognition+Dataset
9. Banos, O., Tóth, M.A., Damas, M., Pomares, H., Rojas, I.: Dealing with the Effects of Sensor Displacement in Wearable Activity Recognition. Sensors 14(6), 9995–10023 (2014)
10. Banos, O., Toth, M.A.: Realistic Sensor Displacement Benchmark Dataset. Dataset Manual (2014)
11. Han, J., Kamber, M., Pei, J.: Data Ming – Concepts and Techniques. Morgan Kaufmann, Massachusetts (2012)
12. Rokach, L., Maimon, O.: Data Mining with Decision Trees – Theory and Applications. World Scientific, Singapore (2008)
13. Kantardzic, M.: Data Mining – Concepts, Models, Methods, and Algorithms. IEEE Press, US (2003)