Journal of
Information Technology & Software Engineering

Nabian, J Inform Tech Softw Eng 2017, 7:4
DOI: 10.4172/2165-7866.1000209

Research Article Open Access

A Comparative Study on Machine Learning Classification Models for Activity Recognition

Mohsen Nabian*

Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA, United States

Abstract

Activity Recognition (AR) systems are machine learning models developed for cell-phones and smart wearables to recognize various real-time human activities such as walking, standing, running and biking. In this paper, the performance (accuracy and computational time) of several well-known supervised and unsupervised learning models including Logistic Regression, Support Vector Machine, K-Nearest Neighbors', Naive Base, 'Decision Tree' and Random Forest are examined on a dataset. It is shown that Random Forest model outperforms other models with accuracy over 99 percent. It is shown that PCA significantly improved the performance of Artificial Neural Network with one hidden layer and SVM models in both accuracy and time, while PCA showed to have negative impacts on Random Forest or Decision Tree models by increasing the running time and decreasing the prediction accuracy.

Keywords: Machine learning; Activity recognition; Support vector machine; Neural network

Introduction

Wearable-based activity recognition (AR) systems are typically built to recognize a predefined set of common activities such as sitting, walking, and running [1-3]. AR has captured the attraction of computer science communities due to its capabilities on providing supporting personalized information and its many applications in human-computer interface [4], medicine [5], insurance businesses [6], sports [5] and sociology [7]. There are variety of studies on AR systems for data provision [8-11]. Moreover, there are many studies centred on developing novel nonlinear models for activity recognition problem [12,13]. However, there is no literature found comparatively investigation on the performance of the well-known classification models in Activity Recognition problem. In this paper, Logistic Regression, Support Vector Machine, K-Nearest Neighbors', 'Artificial Neural Network with one hidden layer' Naive Base, 'Decision Tree' and Random Forest models have been trained and tested. The performance of these models has been examined by both prediction accuracy and running time. Also, the effectiveness of principal component analysis (PCA) preprocessing on each model is investigated. More-over, the important role of hyper-parameters for each model is discussed and visualized. In this paper, it is showed that Random Forest provides the best performance among other models. Furthermore, while PCA has improved the learning process for models such as KNN, Artificial Neural Network with one hidden layer and SVM, it worsens the prediction accuracy of random forest or decision tree models (Figure 1).

Problem Statement

Research studies in Activity recognition have been focused on developing novel models for rare activities and for unbalanced training data and novel methods for data collections [2,12-15]. However, there is a gap for a comprehensive study of the performance of the conventional machine learning models on activity recognition. In this research, a comprehensive study is performed over several well-known supervised learning algorithms. Moreover, the effectiveness of dimensionality reduction using PCA [16] on these models was studied.

Methodology

A large AR dataset [1] composed of 346 K instances almost uniformly distributed over 12 activities is used in this study. The data is shuffled and then normalized by linear map-ping to the range (0,1). 80 percent of the data points were selected for training and the remaining 20 percent were selected for testing. All the aforementioned models were trained with training data and tested with the test data. The running time, as well as the accuracy of these models, were comparatively discussed. The effect of hyperparameters of four models (Logistic Regression, KNN, SVM and Neural Network with one hidden layer) were elaborated on the accuracy of the test data prediction. And finally, the effect of component analysis (PCA) on the performance of these models was studied.

Figure 1: Activity recognition process pipeline.

Theory

Table 1 summarizes the hypothesis, approach and cost functions of a few models used in this paper. Since the problem is multi-class classification, softmax algorithm is used for Ridge-Regression and softmax activation function is used for the last layer of Neural Network. Furthermore, using the method of One-vs-all [17], SVM coefficients were classified 12 times to address the multiclass classification problem. Notice that Ridge Regression is a linear classifier, while SVM, KNN and ANN are nonlinear. Figures 2 and 3 provides the roadmap of this machine learning process. It is expected to have a better prediction

***Corresponding author:** Mohsen Nabian, Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA, United States, Tel: +1 617-373-2000; E-mail: monabiyan@gmail.com

Received September 26, 2017; **Accepted** October 04, 2017; **Published** October 10, 2017

Citation: Nabian M (2017) A Comparative Study on Machine Learning Classification Models for Activity Recognition. J Inform Tech Softw Eng 7: 209. doi: 10.4172/2165-7866.1000209

Copyright: © 2017 Nabian M. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Algorithm	Hypothesis	Approach	Cost Function	Hyper Param.
Ridge Regression	$p(y = 1 \mid x) = \frac{1}{1 + e^{-w^T x}}$	Model should maximize likelihood of data and penalize magnitudes of W	$J = \sum \log(1 + e^{w^T x}) - yxw + \frac{\lambda}{2}\|w\|_2^2$	λ
SVM	$y = sgn(w^T x + b)$	Model should maximize geometric margin	$J = \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{N}\varepsilon_i$ $s.t. y^{(i)}(w^T x^{(i)} + b) > 1 - \varepsilon_i$	C
KNN	Majority vote of its k-neighbors	Model is majority vote of its k-neighbors	No cost Function	K
ANN	$y = h(x) = \frac{1}{1 + e^{-z}}$ Where z is vector of values of the last layer of the neural network	Model should minimize error and penalize magnitudes of W	$J = \frac{1}{N}\sum_{i=1}^{N} h\left(x^{(i)}\right) - y^{(i)}{}_2^2 + \frac{\lambda}{2}\sum_{l}^{n-1}\|w^{(l)}\|_2^2$	H

Table 1: Summary of the mathematical background of classification models.

Figure 2: Type of activities (labels) of the dataset.

Figure 3: Machine learning process road map.

for non-linear classifiers due to the nonlinearity of the sensor data for various activities. The hyperparameter for each classifier should be set to a value that maximizes the accuracy of the model over the test data. In the experiment section, hyperparameters for each classification will be examined. PCA convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. Eliminating components that contain low variation will decrease the dimension of the observation without losing a significant amount of information. Dimensionally reduced data is obtained by the following:

$$y(i) = uT(1) \qquad x(i)\,x \qquad (1)$$

where x and u are calculated as:

$$x = N\,X_{i\,x}(i) \qquad (2)$$

$$x = u_x\,X_x\,V_x{}^T \qquad (3)$$

$$u = u_x\,(:;\,1{:}\,d) \qquad (4)$$

and d is the reduced dimension of the data.

Results and Discussion

Data description

The dataset [1] comprises body motion and vital signs recordings for ten volunteers of diverse profile while per-forming several physical activities. Sensors placed on the subject's chest, right wrist and left ankle are used to measure the motion experienced by diverse body parts, namely, acceleration, the rate of turn and magnetic field orientation. The sensor positioned on the chest also provides 2-lead ECG measurements, which can be potentially used for basic heart monitoring, checking for various arrhythmia or looking at the effects of exercise on the ECG. Figure 4 illustrates the location of the sensors and the type of measurements they convey. The activities selected, are among the most common in people's daily life. The full list of activities is depicted. In average, there are about 30000 data points allocated for each activity except the last activity (jump front back) which has 10000 data point, which is still a very reasonable amount of data. In overall, the data set is quite well prepared for model fitting.

Experimental results

The performance of different learning algorithms on the same AR dataset is explored and compared. Moreover, the effect of hyperparameters on the accuracy of three classifiers are also investigated. Finally, it is shown whether PCA as preprocessing step can decrease the running time of each algorithm while maintaining the accuracy.

Performance measurement of supervised algorithms on AR dataset: Here, seven supervised learning classifiers are trained on the AR dataset and their prediction accuracy for the test data were measured. Figure 5 demonstrates a comparison of accuracies of the classifiers. It is shown in Figure 5 that KNN and Random Forest provided prediction accuracy of >99% while Decision Tree, as well as a simple Artificial Neural Network performance, were >98%. However, Naive Base and Ridge Logistic Regression, as well as SVM appeared to perform relatively poorly. Naive Base assumes independency between

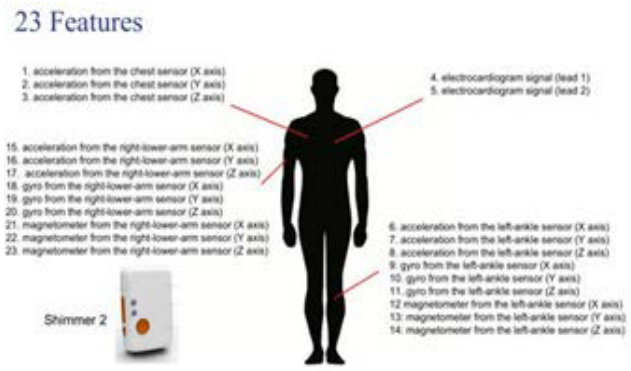

Figure 4: Dataset features for the activity recognition (AR) problem.

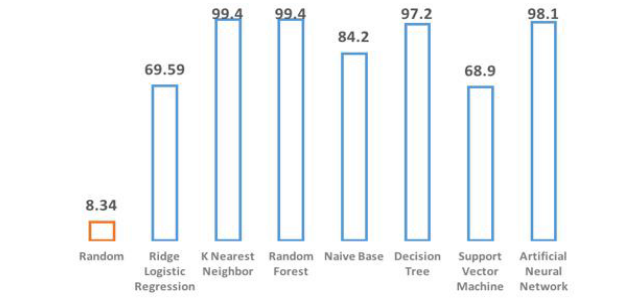

Figure 5: Prediction accuracy of classification models over AR dataset.

features of the observations. This assumption is not descent in the case of Activity Recognition since there are some significant correlations between the values of the sensors. For instance, in the case of lying down, all sensors may measure relatively small values, or in the case of running, the values of sensors may vary dramatically. Therefore, some dependencies exist between sensor values (features) and hence Naive Base does not provide very high accuracy. Moreover, due to the non-linearity of the data for different activities, poor predictions of linear models such as Ridge Logistic Regression and SVM with no kernel which are both linear classifiers.

Besides the accuracy, the running time of classifiers is an important factor for the performance of the machine learning classifiers. The importance of the running time will become tangible once the dimension of the data becomes large. The running times of classifiers on the AR dataset are provided in the Table 2.

Random Forest and Decision Tree appeared to be fast while KNN and SVM appeared to have large computational running time.

The effect of hyperparameters on the performance of the models: Different settings for hyperparameters in models result in different prediction accuracy of models. Therefore, it is essential to investigate the optimum hyperparameter for each classification model for the given training data.

Here, Ridge Logistic Regression, KNN, SVM and Artificial Neural Network with one hidden layer are investigated for the optimum corresponding hyperparameters (Figure 6). In Ridge Regression, is the hyperparameter that penalizes the optimization objective function over the magnitude of the model parameters. Large may prevent over-fitting

Classification Model	Time (sec)
Ridge Logistic Regression	64.3
K-nearest neighbor	149.2
Random Forest	19
Naïve Base	0.6
Decision Tree	15.2
Support Vector Machine	131.2
Artificial Neural Network	2052

Table 2: Running time for training and testing of classification models over AR dataset.

of the model to the training data. However, very large may cause bias in the model which is depicted in Figure 6. With up to around 100, there is no negative effect found on the prediction accuracy, however increasing above 100, a gradual decrease in the model accuracy is observed.

In KNN, K specifies the number of observations to be considered as neighbours for taking the majority vote. Figure 6 shows that increasing K in the case of AR dataset will decrease the accuracy. In SVM, C penalizes the observations that are not in the functional margin of 1. In other words, small Cs allows for some misclassification of data while large C strictly penalize misclassifications. C is particularly useful when the data is not linearly separable. According to Figure 6, the optimum C is between 0.1 to 1. That is certainly a hint that the data is not linearly separable and requires relaxation. However, high relaxation may have a negative impact on the accuracy. In Artificial Neural Net-work with one hidden layer, H is the hyperparameter which determines the number of neurons in the middle layer of the network. As H increases, the Neural Network model will have more parameters to learn. Despite other models, Artificial Neural Network accuracy monotonically improves by increasing the size its parameter. Based on Figure 6, low values of H in (1, 50) is not enough to encompass the complexity of the data. However, for H greater than 50 and the model can well capture the nonlinearity of the data.

The effect of dimensionality reduction on the performance of classification models: PCA is the technique used for the dimensionality reduction of the dataset. This section intends to investigate how dimensionality reduction will improve the running time and to explore whether the classification accuracy will be affected after PCA or not. In PCA, we explore principal directions in the high dimensional data until we capture a large fraction of total variance f in the original dataset. Here, the running time and accuracy of the classification models that were discussed earlier were examined under the range of f = (0.99, 0.90). Figure 7 demonstrates how classification accuracy may alter by reducing dimensions of the data using PCA and Figure 8 projects the effect of dimensionality reduction on the running time of the learning and prediction process of supervised learning models. It is very interesting that the accuracy of Neural Network with one hidden layer improves around 1 percent although we have lost some information in data by reducing the dimensions of the data with PCA step. This improvement still exists for f=0.90. This is consistent with the findings of Mohammad Saleh [18]. Accordingly, Neural Network works more efficiently on a dataset that has their features orthogonal or independent. Also, according to Figure 8, the running time of Neural Network drastically decreases by PCA up to 20 percent of the time required using the original data. Also, Neural Network is the only model that has its prediction accuracy improved by dimensionality reduction using PCA. The prediction accuracy of KNN and Random Forest drop only less than 1 percent for a wide range of f = (0.99, 0.90). According to Figure 8, there is around 60 percent running time drop for KNN. That implies that dimensionality reduction using PCA is highly

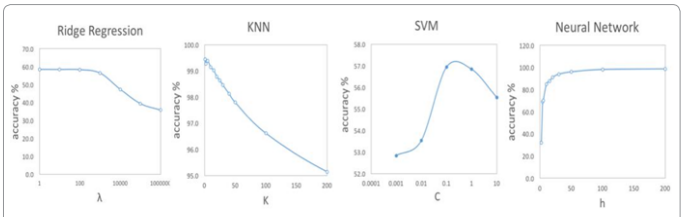

Figure 6: Effect of hyper parameter values on the prediction accuracy of classification models; from left to right: ridge logistic regression, knn, svm, neural network.

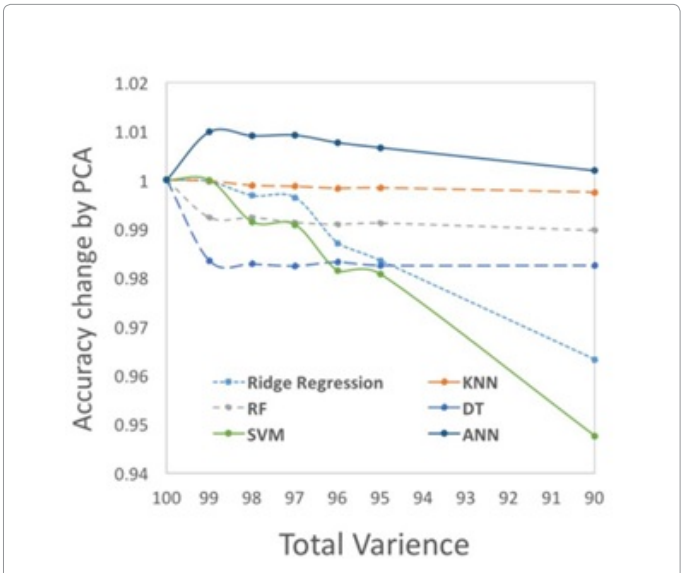

Figure 7: Effect of dimensionality reduction on the prediction accuracy of classifiers. The values of prediction accuracies are normalized with values in Figure 6.

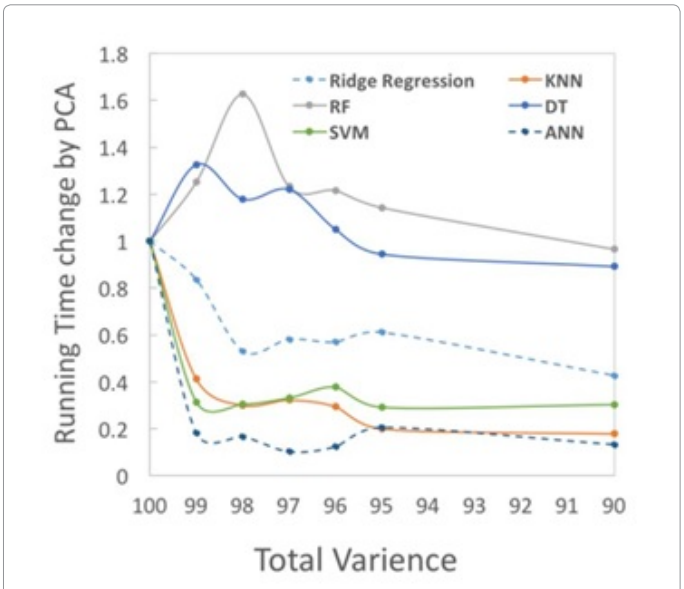

Figure 8: Effect of dimensionality reduction on the running time of classifiers. The values of running time are normalized with values in Table 2.

recommended for KNN. However, PCA is apparently increasing the running time of Random Forest which makes the PCA preprocessing ineffective and not recommended for Random Forest. In SVM, PCA with f=0.99 works effectively since the prediction accuracy doesn't change significantly while the running time may drop over 60 percent. For Decision Tree similar to Random Forest, PCA is not recommended since it has increased the running time.

Conclusion

In this paper, Activity Recognition (AR) problem was investigated using several well-known classification models in machine learning. Due to non-linearity nature of the AR dataset, nonlinear classifiers outperforms linear models. Models including KNN, Neural Network with one hidden layer and Random Forest provided more than 99 percent prediction accuracy for this problem, however, by considering running time, random forest indicated the highest performance with respect to other models. Moreover, the effect of hyperparameters on the prediction accuracy of different models was investigated and lastly, dimensionality reduction is applied as a preprocessing step to these models. By investigating the running time and prediction accuracy of different models after the PCA. It turned out PCA may improve the performance of Neural Network with one hidden layer, KNN and SVM while it has negative effects on the Random Forest and Decision Tree models.

References

1. Banos O, Damas M, Pomares H, Rojas I, Toth MA, et al. (2012) A benchmark dataset to evaluate sensor displacement vi in activity recognition. Proceedings of the 2012 ACM Conference on Ubiquitous Computing. ACM, pp: 1026-1035.
2. Lara OD, Labrador MA (2013) A survey on human activity recognition using wearable sensors. IEEE Communications Surveys and Tutorials 15: 1192-1209.
3. Bao L, Intille S (2004) Activity recognition from user-annotated acceleration data. Pervasive computing, pp: 1-17.
4. Kuutti K (1995) Activity theory as a potential framework for human-computer interaction research. Context and consciousness: Activity theory and human-computer interaction, pp: 17-44.
5. Avci A, Bosch S, Marin-Perianu M, Marin-Perianu R, Havinga P (2010) Activity recognition using inertial sensing for health-care, wellbeing and sports applications: A survey. Architecture of computing systems (ARCS), 2010 23rd international conference on. VDE, pp: 1-10.
6. https://www.virginpulse.com
7. Ching-Hu L, Li Chen F (2009) Robust location-aware activity recogni-tion using wireless sensor network in an attentive home. IEEE Transactions on Automation Science and Engineering 6: 598-609.
8. Ravi N, Dandekar N, Mysore P, Littman ML (2005) Activity recognition from accelerometer data. Aaai 5: 1541-1546.
9. Tapia EM, Intille SS, Larson K (2004) Activity recognition in the home using simple and ubiquitous sensors. Pervasive 4: 158-175.
10. Kwapisz JR, Weiss GM, Moore SA (2010) Activity recogni-tion using cell phone accelerometers. ACM SigKDD Explorations Newsletter 12: 74-82.
11. Maurer U, Smailagic A, Siewiorek DP, Deisher M (2006) Activity recognition and monitoring using multiple sensors on different body positions. Wearable and Implantable Body Sensor Networks, 2006. BSN 2006 International Workshop.
12. Stikic M, Laerhoven KV, Schiele B (2008) Exploring semi-supervised and active learning for activity recognition. Wear-able computers, ISWC, 12th IEEE international symposium, pp: 81-88.
13. Nguyen LT, Zeng M, Tague P, Zhang J (2015) Recognizing new activities with limited training data. Proceedings of the 2015 ACM International Symposium on Wearable Computers. ACM, pp: 67-74.
14. Stikic M, Schiele B (2009) Activity recognition from sparsely labeled data using multi-instance learning. Location and context awareness, pp: 156-173.

15. Nguyen LT, Zeng M, Tague P, Zhang J (2015) I did not smoke 100 cigarettes today: avoiding false positives in real-world activity recognition. Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, pp: 1053-1063.

16. Wold S, Esbensen K, Geladi P (1987) Principal component analysis. Chemometrics and intelligent laboratory systems 2: 37-52.

17. Rifkin R, Klautau A (2004) In defense of one-vs-all classification. J Machine Learning Res 5: 101-141.

18. Mohamad-Saleh J, Hoyle BS (2008) Improved neural network performance using principal component analysis on matlab. Int J Comp, Internet & Manag 16: 1-8.