Activity Recognition (AR): Study on the Performance of Machine Learning Classifiers, Principal Component Analysis and Sensor Ranking for Potential Reduction of Sensors in AR Systems

Mohsen Nabian

ABSTRACT :

The existing sensor technologies in cell-phones and smart wearables provide accurate, real-time information for the motion dynamics of individuals such as linear and radial accelerations. These information enables the recognition of the activities using machine learning classifier systems known as Activity Recognition System (AR). In this paper, several conventional classifiers will be developed and applied on the AR problem. The performance of the classifiers will be compared for the accuracy and time. Moreover, Principal Component Analysis will be performed to reduce the dimension of the data. The reduced sized data will be further used for classification using the same classifiers. The motivation is to measure how dimensionally reduced data will affect the accuracy and time for learning and prediction of classifiers. Lastly, Ridge regression will be applied on the problem to investigate the significance of each sensor to the AR classification problem. The result of Ridge regression can potentially reduce the number of sensors and lead to lower cost in system and its maintenance.

PROBLEM STATEMENT:

Activity Recognition (AR) systems comprises of machine learning algorithms developed for cellphones and smart watches applications to recognize real-time human activities such as walking, standing, sitting, running and biking. AR has captured the interest of many AI communities due to its wide variety of applications in Medicine, Insurance Companies, AI companies, Human-computer interface and Sociology [1] . The current sensor technologies which provides real-time information for the motion dynamics of individuals,enable the Machine Learning classifiers to recognize the real-time activity types of individuals. In this project, we investigate supervised learning classifiers and dimensional reduction techniques on AR datasets.

ALGORITHMS:

The project is composed of two main parts:

1) **Supervised Learning:** In this Project, **Softmax , Neural Network ,Support Vector Machine and KNN** classifiers will be examined on AR dataset. The accuracy performance and the running time of the classifiers will be compared.
2) **Dimensionality Reduction:** In this part, the following algorithms are applied: **a) Principal Component Analysis (PCA) b) Ridge Regression.** PCA is used to reduce the size of data and therefore decreasing the running time for the learning process while maintaining the accuracy. After training the classifiers with the output of PCA over the original data, the

classification performances will be compared with the classification of the original data. Moreover, Ridge regression will be applied on the problem to investigate the significance of each feature (sensor) and to address the possibility of reducing the number of sensors while maintaining the performance in the acceptable threshold. The reduction of sensor will decrease the cost of system and its maintenance.

NOVELTY:

There are many research studies in Activity Recognitions focusing on the development of novel models for the rare activities and unbalanced training data [1-5]. However, there is no literature found on the comprehensive study of the performance of the conventional Machine Learning Classifiers on the Activity Recognition problem. This paper will address the comparison study of several supervised learning models in both running time and accuracy. Moreover, the idea of "sensor reduction" using ridge regression is unique in the context of Activity Recognition Systems.

DATA DESCRIPTION:

The dataset is obtained from UCI Machine Learning Data Repository and it comprises of about 400K labeled data points containing body motion information. There are 23 features in each datapoint. The features include acceleration of chest,left ankle and right-lower arm in X, Y, Z direction. There are 12 activity labels including "standing still", "lying down","climbing stairs", "jumping front and back"," jogging"," bending knee","sitting", "walking", and "running".

EXPECTED RESULTS:

It is expected that non-linear classifiers such as SVM outperform linear classifiers like Softmax due to nonlinearity nature of the data. Also, KNN may result in a very high prediction performance while being timely expensive. However, by using PCA for dimensionality reduction, it is expected to have significant decrease in the running time for classifiers such as KNN. Furthermore, using ridge regression, it is expected to identify some non-significant sensors among the 23 sensors which lead to removing those sensors from the AR system to reduce the costs.

REFERENCES:

1. Lara, O. D., and Labrador, M. A. A survey on human activity recognition using wearable sensors. Communications Surveys & Tutorials, IEEE 15, 3 (2013), 1192–1209.
2. Nguyen, L. T., Zeng, M., Tague, P., Zhang, J. (2015). Recognizing New Activities with Limited Training Data. In IEEE International Symposium on Wearable Computers (ISWC).
3. Stikic, M., Van Laerhoven, K., and Schiele, B. Exploring semi-supervised and active learning for activity recognition. In IEEE International Symposium on Wearable Computers, IEEE (2008), 81
4. Stikic, M., and Schiele, B. Activity recognition from sparsely labeled data using multi-instance learning. In Location and Context Awareness. Springer, 2009, 156–173.
5. Nguyen, L. T., Zeng, M., Tague, P.: Supervised activity discovery. In International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, ACM (2015).