Feature Selection for Wireless Sensor Networks

S. Thiemjarus, B. P. L. Lo, K. V. Laerhoven*, G. Z. Yang
Department of Computing — Imperial College London
*UbiComp Group — Lancaster University

Abstract – In a wireless sensor network, appropriate use of resources can be achieved by choosing the most relevant sensors that are important to the current context. By applying feature selection to determine the optimal sensor locations, the number of sensors can be reduced and great savings (in term of power, hardware, and transmission channel) can be achieved without degrading the decision process. This paper presents the use of BFFS (Bayesian Approach for Feature Selection), a filter based feature selection method, for optimum sensor location selection. The virtue of the method is that the selection of features is purely based on the data distribution and thus is unbiased towards a specific model. The strength of this framework is demonstrated by experiments on activity recognition with the use of a Self-Organising-Map.

Key words – BFFS, feature selection, Wireless Sensor Networks, context awareness

I. INTRODUCTION

Wireless Sensor Networks (WSNs) are becoming a significant enabling technology for a wide range of applications. One of the promising applications of WSN is in the form of Body Sensor Networks (BSNs) for continuous monitoring of physiological signals for extending the current healthcare solutions [1]. Whilst the prospect of continuous sampling of physiological parameters for at-risk patients opens a whole range of new opportunities in medicine, it also imposes significant technical challenges. One of the key issues to be addressed is context awareness and multi-sensory data fusion. Reliable detection of patient activity under which the physiological signals are sampled is important to the capture of clinically relevant episodes. This normally requires the use of a large number of sensors around the body, thus imposing a significant burden on the overall power consumption and bandwidth requirement of the BSNs. Identifying sensors that have direct implication to the decision process is advantageous in that it can be used not only offline to determine optimal sensor locations, but also online to dynamically enable/disable the sensors depending on the current/predicted context. Previous work has shown that the power can be greatly saved simply by not transmitting the non-useful data [2]. The purpose of this paper is to present a new framework based on dimensionality reduction for optimum selection of sensor responses.

Feature selection [3] is a dimensionality reduction technique widely used for data mining and knowledge discovery. It allows elimination of (irrelevant/redundant) features, whilst retaining the underlying discriminatory information. In the context of BSNs, feature selection implies less data transmission and efficient data mining. It also brings potential communication advantages in terms of packet collisions, data rate, and storage.

II. MATERIAL AND METHODS

Bayesian Framework for Feature Selection (BFFS)

BFFS is a filter based feature selection method developed at Imperial College. The virtue of the method is that the selection of features is purely based on the data distribution and thus is unbiased towards a specific model. The criteria for feature selection are based on the expected AUC (the Area Under the ROC-Curve) and therefore features derived should yield the best classification performance in terms of sensitivity and specificity for an ideal classifier. Existing results have shown that with Bayes or C4.5 classifiers, the method has significant improvement over existing techniques both for artificial and real-world datasets.

In this study, the general activities was recorded with body-worn acceleration sensors. To evaluate the performance of BFFS, a multi-layer Self-Organising Map (SOM) [4, 5] with temporal information is employed as the classifier. The use of temporal information is important as a single SOM is not ideal for dealing with tasks with too high perplexity. The performance of the classifier is measured using the average mean and standard deviation of the accuracy over 30 runs.

Data Collection

The experiments were performed on the dataset obtained from ETH Zurich [6]. The dataset was obtained from six sets of sensors placed on each side of the body. There are two 2-axis accelerometers (*i.e.*, **x-y** and **x-z**) in each set. The sensors are located at all major joints of the human body, namely, on the shoulder (**a-d**), just above the elbow (**e-h**), on the wrist (**i-l**), on the hip (**m-p**), just above the knee (**q-t**), and just above ankle (**u-x**). The data was acquired in two separate streams, representing information from the left and right body, respectively, with the sample rate of 92Hz. For the training of the BFFS framework, only one data stream was used and we selected 2000 records for each activity to avoid bias. Since for each sensor the range of the acquired signal can be contrastingly different, the overall standard deviation was used as the quantisation scale for the BFFS algorithm. Experiments were performed on subsets of activities as follows: 1. {Sitting, Standing, Walking}, 2. {Going Upstairs, Going Downstairs}, 3. {Handshake, Writing on White Board, Typing}.

III. Results & Discussions

Figure 1 illustrates the results of BFFS for the subset related to sitting, standing, and walking. The selected features with BFFS are mainly related to sensor readings on the knee and hip. By using the first four features identified (**rspt**) the corresponding classification accuracy increased from 71.9% (all features) to 82.4% (**rspt**), as shown in Table 1.

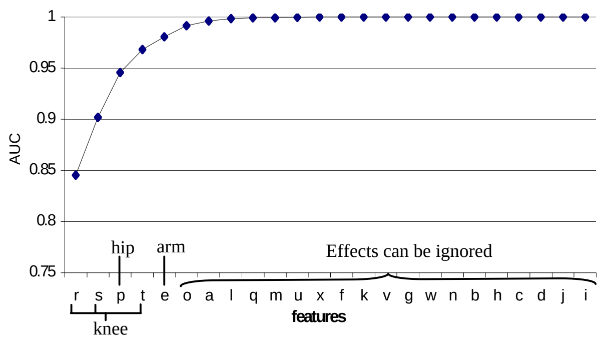

Figure 1. The Expected AUC for Sitting, Standing, and Walking

Features	Mean	SD
All features	0.719	0.051
rspte	0.773	0.02
rspt	**0.824**	**0.011**
rsp	0.792	0.025
rs	0.782	0.039
r	0.781	0.026

Table 1. Average Mean and Standard Deviation of the Accuracy for Sitting, Standing, and Walking (features are chosen based on the result in Figure 1).

Similar experiments were conducted for the remaining subsets. For Subset 2, where the activities were related to going up and down the stairs, the most important seven features selected by BFFS are 1 on the shoulder, 3 on the hip, 1 on the elbow and 2 on the ankle. This gave an average accuracy of 89% (**anp**). When all the features were used, the corresponding accuracy was 75%. For Subset 3 which involves activities such as handshaking, writing and typing, the selected features correspond to x, y-axes of the knee, z-axis wrist, and x-axis shoulder, respectively. The corresponding classification accuracy improved from 77.6% (all features) to 88.2% when the sensor readings determined from BFFS were used.

In conclusion, we have presented in this paper a framework for positioning the optimal locations of the sensors based on a feature reduction technique and demonstrated how it can be effectively used for activity recognition. Applying BFFS for feature selection, all relevant sensor positions have been revealed while the redundant and duplicated sensors discarded. Classification with SOM yields a significant improvement in accuracy when only the selected features are used. In SOM, irrelevant features can cause a considerably high confusion, each feature is allowed an equal voting weights towards the classification of each data record. An increase in accuracy can, therefore, be obtained when these features are eliminated.

V. References

1. H. Baldus, K. Klabunde, and G. Musch, “Reliable Set-Up of Medical Body-Sensor Network”, EWSN 2004, LNCS 2920, pp. 353-363, 2004.
2. http://www.cs.berkeley.edu/~polastre/listening.gif
3. M. Dash and H. Liu, “Feature Selection for Classification”, IDA, pp. 131-156, 1997
4. K. V. Laerhoven, “On-line Adaptive Context Awareness from Low-level sensors”, Master Thesis, Free University of Brussels, 1999.
5. T. Kohonen, Self-Organizing Maps, 2nd ed., Springer-Verlag Heidelberg, 1997.
6. N. Kern, B. Schiele and A. Schmidt, “Multi-Sensor Activity Context Detection for Wearable Computing”, EUSAI, 2003.