FEATURE SELECTION BASED ON MUTUAL INFORMATION FOR HUMAN ACTIVITY RECOGNITION

Benjamin Fish, Ammar Khan[*], *Nabil Hajj Chehade*[†], *Chieh Chien, Greg Pottie*

Center for Embedded Networked Sensing
University of California, Los Angeles

ABSTRACT

In this work, we consider a classification problem of 14 physical activities using a body sensor network (BSN) consisting of 14 tri-axial accelerometers. We use a tree-based classifier, and develop a feature selection algorithm based on mutual information to find the relevant features at every internal node of the tree. We evaluate our algorithm on 31 features per accelerometer (total of 434), and we present the results on 8 subjects with a 96% average accuracy.

Index Terms— Activity Classification, Feature Selection, Accelerometers.

1. INTRODUCTION

Human activity recognition is central to many fields such as neurological rehabilitation [1], context-aware computing, and athletic training [2]. For example, in neurological rehabilitation, doctors are interested in monitoring their stroke patients' activities at home and in the community. Traditional methods of motion monitoring are based on tedious manual techniques such as self-monitoring or constant monitoring by an observer. These techniques are prone to error due to forgetfulness and other kinds of misreporting. Recent advances in low-power and compact sensor technology made the automation of activity monitoring, using a body sensor network, feasible and low cost. In [3], multi-modal sensor systems were used to classify basic physical activities, including walking, jogging, and going up and down stairs. In [4] and [5], sensor systems using only accelerometers were used for activity classification; [4] used biaxial accelerometers to monitor both ambulatory and sedentary motions, while [5] used tri-axial accelerometers to monitor workspace activities. Smart phone-based accelerometers were also used for activity recognition as in [6]. A representative sampling of previous research is presented in Table 1.

Ref.	No. Activities	No. Sensors	No. Subjects	Accuracy
[4]	20	5	20	84%
[6]	5	1	10	85%
[7]	5 groups	6	11	N/A
[8]	5	2	5	89 %
[9]	8	7	12	90%
[10]	8	12	1	65% - 95%
[11]	8	1	7	95%
[3]	10	7	2	95%
Ours	14	14	8	96 %

Table 1. Summary of previous research

In our work, we aim at capturing the motions of all the parts of the body for a thorough study of the activity recognition problem. We over-instrument the subjects with 14 tri-axial accelerometers placed on various parts of the body, and we consider the classification of 14 common daily activities. We take a supervised learning approach, using a binary decision-tree with a naïve Bayes classifier at every internal node and a large feature set of 31 features per accelerometer (total of 434 features). This is a high-dimensional problem where brute force is not possible, and a feature selection algorithm is needed to find the best features for every naïve Bayes classifier (present at every internal node). Feature selection is a problem that has been studied many times before in other contexts. Different types include margin-based algorithms such as RELIEF [12] and mutual information-based algorithms such as MIFS [13]. We use a mutual information-based algorithm because it is computationally capable of handling the large amount of data captured by 14 accelerometers. Our contribution is that we describe an activity classification system that can handle a large set of activities (14 activities) representative of the common daily activities, and achieves a very high recognition accuracy. The system was tested on 8 different subjects.

2. METHODOLOGY

2.1. Training Data Collection

Accelerometers are placed on an individual at fourteen locations, as shown in figure 1. The accelerometers we used were tri-axial wireless Gulf Coast Data Concept X6-2mini accelerometers (± 6g) [14], which continually collected data at a rate of 160Hz. Fourteen different activities are performed,

[*]Benjamin Fish and Ammar Khan performed the work while interning at the Center of Embedded Networked Sensing of UCLA.

[†]Corresponding author: nabilhc@ucla.edu

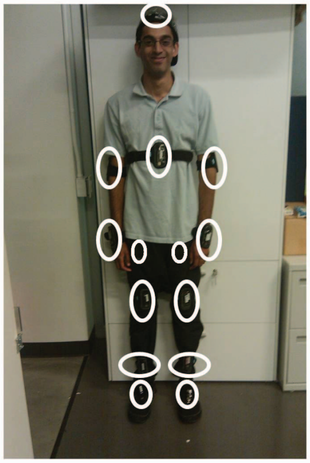

Fig. 1. Location of the 14 accelerometers

as described in table 2. To collect labels for ground truth, we used an Android phone application. The application has a list of the activities to choose from and a start/stop button to record the time the subject started the activity, and the time he/she stopped. Eight different data sets were collected from eight different healthy individuals for a length of five minutes per activity.

Active	Stationary
Slow walk	Stand
Fast walk	Sit (upright)
Walk (up-slope)	Sit(hunch)
Walk (down-slope)	Sit (slouch)
Walk (up stairs)	Lie down (on back)
Walk (down stairs)	Lie down (on stomach)
Run	Lie down (on side)

Table 2. The 14 activities that were classified

2.2. Features Computation

Features were computed on 4 second windows of acceleration data with 3 second overlapping between consecutive windows. We compute 31 different features for each sensor, shown in table 3. Since we used 14 different sensors, this meant a total of 434 features from which to choose.

2.3. Classification

We used the binary decision tree shown in figure 2, with a naïve Bayes classifier at each node. The naïve Bayes classifier is a probabilistic method given by the function (1), where $\mathcal{C}$ is the set of classes and $\mathcal{F}$ is the set of features.

$$\max_{C \in \mathcal{C}}\{p(C) \prod_{f \in \mathcal{F}} p(f|C)\} \tag{1}$$

Features
Standard deviation of x,y,z axes and m
Mean of x,y,z axes and m
Absolute mean of x,y,z axes and m
Energy ratio of x,y,z axes and m
Ratio of DC to sidelobe of x,y,z axes
First sidelobe location of x,y,z axes
Max value of x,y,z axes and m
Short time energy in x,y,z axes and m
Correlation between x and y axes

Table 3. Features used. (m refers to the magnitude of the 3D acceleration vector.)

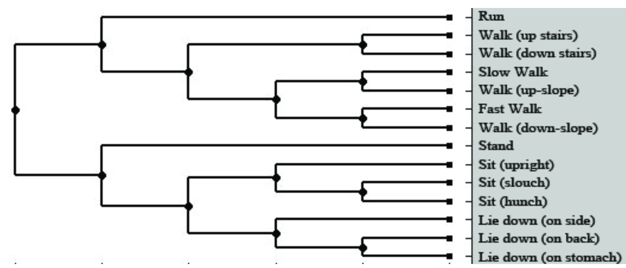

Fig. 2. Decision tree used.

This classification was performed offline. A tree was used so that the classifier would not have to distinguish between all 14 of the classes using the same set of features. Instead, classifiers are used to partition the data into smaller and smaller categories of classes until the categories consist of a single class, at which point the data is fully classified. In the probability calculations (given by Bayes' rule), the features were assumed to be independent with a Gaussian distribution, as required by the naïve Bayes classifier. For every subject, the naïve Bayes classifiers (at the internal nodes of the tree) were trained on his/her training data; this is often called user-dependent procedure. The feature selection was also personalized to every subject.

2.4. Feature Selection Algorithm

The high-dimensionality of the problem requires a good feature selection algorithm to find the best features for the naïve Bayes classifier at every internal node. In order to minimize computational complexity while maximizing accuracy, this algorithm employs a 'filter' solution first, and then a 'wrapper.' The algorithm works as follows:

1. We determine the Gaussianity of each feature by calculating the negentropy of each feature given each class using the approximation given in equation (2), where J is the negentropy, x is a random variable, E is the expected value, and $kurt$ is the kurtosis, the fourth central

moment of the distribution [15] [16].

$$J(x) \approx \frac{1}{12}E\{x^3\}^2 + \frac{1}{48}kurt(x)^2 \tag{2}$$

We remove all features with negentropy values that are higher than an apriori threshhold.

2. Using the mRMR algorithm, we ranked the features. [17] The term this algorithm wishes to maximize is given by formula (3).

$$I(C; f_i) - \frac{1}{|S|}\sum_{f_s \in S} I(f_s; f_i) \tag{3}$$

I is the mutual information, C is the class variable, f_i is the feature under consideration, and S is the set of features already selected and ranked. Calculating mutual information requires calculating the entropy of a feature or set of features, a computationally expensive process because each feature is a mixture of Gaussians. Hence a Taylor series approximation of the entropy was employed [18].

3. By now, there are a few parameters that can be changed: the threshhold for the negentropy values and the degree of the Taylor series approximation. In addition, there are really two different possible algorithms, using only the first term of (3) (Max-Relevance), or both (Max-Relevance and Min-Redundancy) [17]. Instead of choosing one algorithm, or just one set of parameters, a range of parameters are used over both algorithms, and the sets of features returned by these algorithms are captured. Because we wish to minimize the number of features, we use the first k features in each ranking, where k ranges from 1 to the full set.

4. This gives us a list of feature sets. We pick the feature set that minimizes the training error[1].

5. The above steps are repeated for each node in the tree. Then for each node, the highest ranking set of features are chosen, and the total number of sensors used so far is updated.

3. RESULTS

We collected sets of data from eight different individuals where the participant did five minutes of each of the 14 activities while wearing all 14 accelerometers. For every subject, we build a personalized classifier; we train on half of the data (2.5 minutes per activity) and tested on the other half, a time suggested by [5]. We got an average overall accuracy of 96.5%, as seen in Table 5. Table 5 also shows a comparison between our algorithm and a random selection of features (14 features) at every internal node of the decision tree. Our algorithm clearly outperforms the random selection of features. The number of features (14 features) was selected just for comparison reasons. Table 5 shows that a large number of sensors was used by our algorithm for every subject. This is due to the fact that the feature selection algorithm does not take into consideration from which sensor the features were selected. It would be interesting to change the feature selection algorithm to a sensor selection algorithm, while maintaining a relatively high accuracy. This could be done by adding a term to favor features from the same sensors. It is also worth noting that for different subjects, different features were selected. This is due to the variation in the acceleration data belonging to different subjects (e.g. different subjects walk differently, sit differently, and lie differently.).

Activity	Percent Correct
Run	100%
Walk (up stairs)	97.67 %
Walk (down stairs)	94.54 %
Slow Walk	92.77 %
Walk (up-slope)	95.95 %
Fast Walk	96.81 %
Walk (down-slope)	95.32 %
Stand	99.41 %
Sit (upright)	89.90 %
Sit (slouch)	94.62 %
Sit (hunch)	99.24 %
Lie down (on side)	100 %
Lie down (on back)	94.83 %
Lie down (on stomach)	99.66 %

Table 4. Average accuracy for each of the activities.

Subject	Random Features	Our algorithm	No. of Sensors
1	51 %	93 %	10
2	64 %	98 %	12
3	71 %	96 %	9
4	86 %	97 %	10
5	70 %	98 %	12
6	83 %	99 %	10
7	83 %	96 %	13
8	74 %	95 %	11
Average	72.75 %	96.5 %	10.875

Table 5. Average accuracy for each of the test subjects. We compare our algorithm to an algorithm that selects 14 random features at every internal node. The last column shows the number of sensors used by our algorithm.

4. CONCLUSION

This work presents a combination of a tree-based classification and a feature selection algorithm for human activity

[1]This corresponds to choosing the feature set that gives the highest discrimination between the two branches of the tree at the corresponding node. Training error is the percentage of misclassified training data

recognition, and shows that a high activity recognition rate is achievable for a large set of common daily activities. More than just a specific algorithm, this paper presents a framework that maximizes the accuracy that can be garnered from the results of specific algorithms, like the mRMR algorithm that we used. This work shows that different sensors (at different locations on the body) are the best for discriminating between subsets of the activities. The algorithm presented could be changed to minimize for the number of sensor used. This is a step forward towards understanding human activities and towards finding the best placements of sensors on the body for the recognition of a large set of activities.

5. ACKNOWLEDGEMENTS

We would like to thank all of the participants in the lab who helped us collect data and everybody at CENS who supported us. We also would like to thank Xiaoyu Xu for helping with setting up the sensors, and James Xu for providing the Android phone application to label the data. This work was funded by the National Science foundation as part of the Center for Embedded Networked Sensors (CENS), under grant CCR-0120778.

6. REFERENCES

[1] D.K. White, R.C. Wagenaar, M.E. Del Olmo, and T.D. Ellis, "Test-retest reliability of 24 hours of activity monitoring in individuals with parkinsons disease in home and community," *Neurorehabil Neural Repair*, 2007.

[2] F. M. Impellizzeri, S. M. Marcora, C. Castagna, T. Reilly, A. Sassi, F. M. Iaia, and E. Rampinini, "Physiological and performance effects of generic versus specific aerobic training in soccer players," *International Journal of Sports Medicine*, 2006.

[3] J. Lester, T. Choudhury, G. Borriello, S. Consolvo, J. Landay, K. Everitt, and I. Smith, "Sensing and modeling activities to support physical fitness," *Proceedings of UbiComp*, 2005.

[4] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data," *Pervasive Computing*, 2004.

[5] N. Ruchansky, C. Lochner, E. Do, T. Rawls, N. Hajj Chehade, J. Chien, G. Pottie, and W Kaiser, "Monitoring workspace activities using accelerometers," *Proceedings of ICASSP*, 2011.

[6] Zhongtang Zhao, Yiqiang Chen, Junfa Liu, Zhiqi Shen, and Mingjie Liu, "Cross-people mobile-phone based activity recognition," *International Joint Conference on Artificial Intelligence*, 2011.

[7] L. Atallah, B. Lo, R. King, and G. Yang, "Sensor placement for activity detection using wearable accelerometers," *International Conference on Body Sensor Networks (BSN)*, 2010.

[8] K. Aminian, Ph. Robert, E. Buchser, B. Rutschmann, D. Hayoz, and M. Depairon, "Physical activity monitoring based on accelerometry: validation and comparison with video observation," *Medical and Biological Engineering and Computing*, 1999.

[9] J. Lester, T. Choudhury, and G. Borriello, "A practical approach to recognizing physical activities," *Pervasive Computing*, 2006.

[10] N. Kern, B. Schiele, and A. Schmidt, "Multi-sensor activity context detection for wearable computing," in *Ambient Intelligence*.

[11] J.Y. Jhun-Ying Yang, J.S. Wang, and Y.P. Chen, "Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers," *Pattern Recognition Letters*, 2008.

[12] Kenji Kira and Larry A. Rendell, "A practical approach to feature selection," *Proceedings of the ninth international workshop on Machine learning*, 1992.

[13] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Transactions on Neural Networks*, 1994.

[14] Gulf Coast Data Concepts: http://www.gcdataconcepts.com.

[15] S. Roberts and R. Everson, *Independent component analysis: principles and practice*, Cambridge University Press, 2001.

[16] T.M. Cover and J.A. Thomas, *Elements of information theory*, Wiley Series in Telecommunications and Signal Processing. Wiley-Interscience, 2006.

[17] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005.

[18] M.F. Huber, T. Bailey, H. Durrant-Whyte, and U.D. Hanebeck, "On entropy approximation for gaussian mixture random vectors," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2008.