2015 The 5th International Workshop on Computer Science and Engineering
doi: 10.18178/wcse.2015.04.066

A Research on Feature Selection Method for Intelligent Landmine System

SUN Yujia, YU Jiyan, JIA Fangxiu, WANG Xiaoming, and ZHANG Qiang

ZNDY of Ministerial Key Laboratory, Nanjing University of Science and Technology, Jiangsu, 210094, China

Keywords: feature selection, algorithm, intelligent landmine.

Abstract. In this paper, we apply a mRMR method to select a feature subset for intelligent landmine movement classification(ILMC). We have compared mRMR with other three algorithms based on our ILMC dataset. And mRMR algorithm gives the best result for selecting features which gives very good classification accuracy.

Introduction

An intelligent landmine is comprised of a anti-removal system which can recognize the landmine's activity and execute necessary step to stop removal activity. There are several anti-removal method has been proposed based on different sensors such as acoustic, infrared and vibration. In this paper, we aim to apply a feature selection algorithm to find the optimal feature subset for landmine movement classification.

The feature of an intelligent landmine is an individual measurable property of the process being observed. Based on a set of features and a machine learning algorithm the intelligent landmine system can perform classification. In the past years in the applications of intelligent landmine movement classification, the domain of features have expanded to tens of variables or features used in this application with development of sensor technology. Several techniques are developed to address the problem of reducing irrelevant and redundant variables which are a burden on challenging tasks. An intelligent landmine system always have the problem concerned power consumption. Feature Selection (variable elimination) helps in understanding data, reducing computation requirement, reducing the effect of curse of dimensionality and improving the predictor performance. In this paper we look at some of the methods found in literature which use particular measurements to find a subset of variables (features) which improves the overall prediction performance.

Section 2 of this paper will describe about other works related to this research. Section 3 will describe the experiment to get raw dataset and the feature selection method. Section4 we will compare mRMR with other three methods and talk about the result. Section 4 will describe conclusion and further work.

Related Works

The focus of feature selection is to select a subset of variables from the input which can efficiently describe the input data while reducing effects from noise or irrelevant variables and still provide good prediction results [1]. One of the applications would be in gene microarray analysis [1–5] and these years several researchers have tried to imply some method to ILMC.

To remove an irrelevant feature, a feature selection criterion is required which can measure the relevance of each feature with the output class/labels. From a machine learning point if a system uses irrelevant variables, it will use this information for new data leading to poor generalization. Removing irrelevant variables must not be compared with other dimension reduction methods such as Principal Component Analysis (PCA) [6] since good features can be independent of the rest of the data [7]. Though a good subset should contain features that are highly relevant and

Supported by the National Natural Science Foundation of China(NO. 61201391)

nonredundant, weakly relevant (but nonredundant) features help the correlation-based feature selection algorithms [8], [9] and a tradeoff between relevancy and redundancy of features may be useful for classification [10]. The mRMR method does not allow a tradeoff between relevancy and redundancy of genes. Greedy algorithms and simulated annealing have been attempted to determine the optimal tradeoff between the relevancy and the redundancy of a set of genes [11], [12]. In another study, the relevancy–redundancy criterion was attempted in two stages [13] using Wilcoxon test or F-test, the relevant gene set was obtained from original microarray dataset, and subsequently, redundant genes were removed from the selected gene set by controlling the upper bound of Bayes error. Ooiet al. in [14] studied the tradeoff between relevancy and redundancy in multiclass gene selection problem by introducing a data-dependent tuning parameter called differential degree of prioritization. Recently, ReliefF and MRMR algorithms were combined in a two-stage strategy for large- scale gene selection [15]. In the first stage, a small subset of genes was selected using ReliefF, and then, MRMR method was applied to select non-redundant genes into the subset.

Method

1.1 Hardware

All the data was acquired from ZDY01 intelligent landmine system. Consisting of a Cortex-M3 microcontroller, a wireless transceiver and a tri-accelerometer, the system can sample the acceleration with a sampling rate 100Hz and transmitting the data to computer wirelessly. The diagram of the system was showed on figure 1.

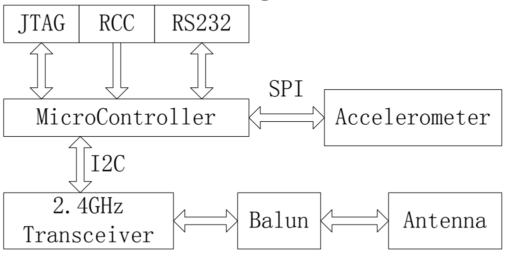

Fig.1: Hardware diagram Fig2: Experiment picture

The accelerometer ADXL345 from Analog Device is a small, low power, tri-axis digital accelerometer with signal conditioned voltage outputs. The ADXL345 accelerometer can sense acceleration in three axes with a full-scale range of ±2g, ±4g, ±8g,and ±and ±16g optional. In this data acquisition experiment ±16g was chose to meet the data requirement. The ADXL345 has a selectable output data rate ranging from 0.1Hz to 3200Hz. As 20Hz frequency is required to assess intelligent landmine physical activity[16] and the ADXL345 automatically modulates its power consumption in proportion to its output data rate, a output data rate 100Hz was chose. Four AAA batteries can power the ZDY01 for roughly 24 hours which is more than sufficient for the data collection sessions used in this study . A ZDY01 is shown in figure2 .

1.2 Experiment

In this study, firstly, the landmine model is on the floor with different postures such as stand, lying, updown. Then, subjects were asked to take the model up and hold the model waling a short way with two hands. At last the subjects will put the model down on the floor. We select 10 subjects and each one will take this procedure 10 times. All the sensing data from sensors are often noisy and ambiguity. The raw signals are filtered to remove noise and to fill out lost samples. In this step, the data filter performs both a low-pass filtering for removing abnormally low sample values and a high-pass filtering for removing abnormally high sample values, as showed in figure 3. After that, samples are grouped into sliding windows or frames.

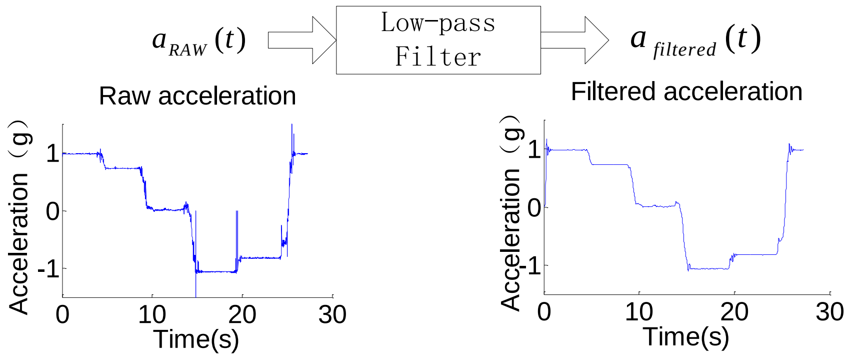

Fig3: Raw data and filtered data

Along with acceleration X, Y, Z, we compute pitch ,roll for each triplet:

$$Pitch = 2\arctan\left(\frac{y}{\sqrt{x^2 + z^2}}\right) \tag{1}$$

$$Roll = 2\arctan\left(\frac{x}{\sqrt{y^2 + z^2}}\right) \tag{2}$$

Where x, y, z are acceleration values of the three axes.

Previous studies showed that the length of sliding window has significantly impact on the performance of the pattern recognition algorithms[17]. In this study, we did a pilot study on the subset of collected dataset for selecting a reasonable length for sliding window. We varied the window length 1 second, 1.2 second, 1.5seconds, 1.8 seconds, 2 seconds and 2.5 seconds and we stick on the window length of 1.8 seconds. The reason for the choosing window length of 1.8 second is that this length allows avoiding delay from continuously real-time processing while providing a reasonable recognition rate.

For each frame of size n where n is number of time points, the following features are extracted:

$$Mean(x) = \frac{\sum_{i=1}^{n} x_i}{n} \tag{3}$$

$$Standard \quad deviation(x): \delta_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i^2) - [Mean(x)]^2} \tag{4}$$

$$Energy(x) = \frac{\sum_{i=1}^{n} x_i^2}{n} \tag{5}$$

$$Entropy(x) = -\sum_{i=1}^{n} p(x_i)\log(p(x_i)) \tag{6}$$

where x_i is an acceleration value; $p(x_i)$, a probability distribution of x_i within the sliding window, can be estimated as the number of xi in the window divided by n.

$$Correlation(X,Y) = \frac{\operatorname{cov}(x,y)}{\delta_x \delta_y} \tag{7}$$

in which $\operatorname{cov}(x,y)$ is covariance and δ_x , δ_y are standard deviation of x and y.

Peak/bottom acceleration: for each sliding window, we also extracted 3 peak values and 3 bottom values of acceleration.

These features are combined into a 58-dimentional feature vector, composed of Mean X, Standard deviation X, Energy X, Entropy X, Mean Y, Standard deviation Y, Energy Y, Entropy Y, Mean Z, Standard deviation Z, Energy Z, Entropy Z, Mean Pitch, Standard deviation Pitch, Energy Pitch, Entropy Pitch, Mean Roll, Standard deviation Roll, Energy Roll, Entropy Roll, Correlation XY, Correlation YZ, Correlation ZX, peak value of X, peak of Y, peak of Z, bottom of X, bottom of Y, and bottom value of Z.

We need a feature selection algorithm to remove irrelevant and/or redundant features, which will be talking about in the following section.

1.3 Feature selection

A basic feature selection procedure diagram is showed in figure .To get a valid feature combination, there are always four steps: generation procedure, evaluation function, stopping criterion, validation procedure.

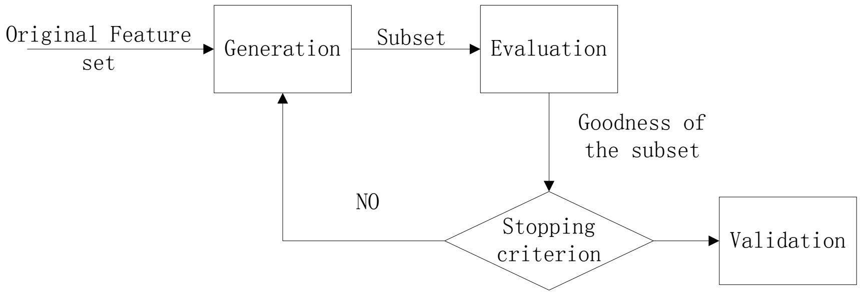

Fig4: Basic diagram of feature selection

mRMR Algorithm [18] is one of the feature selection which utilizes information-based theory approach with the concept of Max-Dependency. Max Dependency concept aims to find the set of features S with m number, which has the greatest dependence with the target class c.

mRMR algorithm is an improvement of Max-Relevance feature selection which implement Max-Dependency scheme. An example of Max-Relevance approach is mutual information based feature selection. Given two random variables x and y, mutual information is defined as a probabilistic function of $p(x)$, $p(y)$, and $p(x,y)$:

$$I(x,y)=\iint p(x,y)\log\frac{p(x,y)}{p(x)p(y)}dxdy \tag{8}$$

In Max-Relevance, the m selected features x_i has the largest mutual information value of $I(x,y)$ with the target class c.

In feature selection, the combination of these features does not always produce the best classification performance. This is because of the dependency between features that creates redundancy. mRMR algorithm is an algorithm that can be used to minimize redundancy by using a series of calculations of relevance and redundancy to select the features.

mRMR algorithm uses the Max-Relevance criterion as a basis to search for related features. It computes $D(S,x)$ that the average of all mutual information value of the individual features xiand class c:

$$\max D(S,x), D=\frac{1}{|S|}\sum_{x_i\in S} I(x_i,c) \tag{9}$$

Features selected based on Max-Relevance only can have a lot of redundancy because the dependency among these features is very large. Therefore, Min-Redundancy criterion added to select the features that mutually exclusive:

$$\min R(S), R=\frac{1}{|S|^2}\sum_{x_i,x_j\in S} I(x_i,x_j) \tag{10}$$

The criteria that contain both limits is called mRMR algorithm. Defined operator Φ (D, R) to combine D and R:

$$\max \Phi(D,R), \Phi=D-R \tag{11}$$

In practice, the incremental search methods can be used to find optimal or near optimal fit features defined in Φ. If we already have S_{m-1}, which is the set of m–1 features, next things to do is to select the m-th feature of the set $\{X-S_{m-1}\}$. This step is completed by selecting the features that have maximum Φ. Incremental search algorithm is performed to optimize the conditions below:

$$\max_{x_j \in X - S_{m-1}} [I(x_j, c) - \frac{1}{m-1} \sum_{x_i \in S_{m-1}} I(x_j, x_i)] \qquad (12)$$

The m-th feature can also be selected by maximizing the single-variable relevance divided by redundancy function. In several case with numeric data, mutual information is replaced by Pearson Correlation.

The complexity of this incremental search is O (|S|.m.r^2), where S is number of expected feature, m is number of features, and r is number of records.

Result

For above data, we perform a Branch and Bound[19] feature search algorithm with a mRMR evaluation function which was talked above. According to different feature subset size 5,10,15,20. After that, three classification methods: Decision Tree (ID3), SVM, and Naive Bayes, are applied to the features that have been selected. We use k-fold cross validation test with 10 fold for testing the model, Next, the accuracy results were compared with other feature selection algorithms: Max-Relevance (Mutual Information), Relief, and MIFS. The results are displayed in Table1, Table2, Table3.

Table 1: ID3 Modeling accuracy result(%)

Algorithm	5	10	15	20
MR	60.52	69.32	67.43	70.02
Relief	70.32	71.4	69.32	68.44
MIFS	73.44	72.52	73.5	71.2
mRMR	82.34	80.42	79.2	78.32

Table 2: SVM Modeling accuracy result(%)

Algorithm	5	10	15	20
MR	60.42	68.67	67.02	70.43
Relief	71.02	70.52	69.2	68.41
MIFS	72.78	71.32	72.67	70.78
mRMR	81.34	82.32	79.2	79.12

Table 3: Naïve Bayes Modeling accuracy result(%)

Algorithm	5	10	15	20
MR	62.31	70.21	68.42	71.33
Relief	71.78	72.41	70.23	69.52
MIFS	74.33	73.42	74.52	73.11
mRMR	83.52	81.32	80.11	79.32

From the result, we found that mRMR algorithm has the best performance than other three algorithms. mRMR algorithm is an algorithm that is stable for all features retrievals classification models tested. This algorithm is recommended for selecting intelligent landmine movement features, although this algorithm has a relatively high complexity.

Conclusion and Discussion

In this paper we apply mRMR to select a feature subset for landmine movement classification. We have compared mRMR with other three algorithms. And mRMR algorithm gives the best result for selecting feature which gives very good classification accuracy.

For further work, we will test mRMR for other kinds of activities which intelligent landmine will perform such as shot.

References

[1] Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003;3:1157–82.

[2] Guyon I, Weston J, Barhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn 2002;46:389–422.

[3] Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. J BioinformComputBiol 2005;3:185–205.

[4] Chuang L-Y, Chang H-W, Tu C-J, Yang C-H. Improved binary PSO for feature selection using gene expression data. ComputBiolChem 2008;32:29–38.

[5] Lazar C, Taminau J, Meganck S, Steenhoff D, Coletta A, Molter C, et al. A survey on filter techniques for feature selection in gene expression microarray analysis. IEEE/ACM Trans ComputBiolBioinform 2012;9.

[6] Alpaydin E. Introduction to machine learning. The MIT Press; 2004.

[7] Law MH, Figueiredo M rio AT, Jain AK. Simultaneous feature selection and clustering using mixture models. IEEE Trans Pattern Anal Mach Intell 2004;26.

[8] R. Kohavi and G. John, “Wrappers for feature subset selection,” Artif. Intell., vol. 97, pp. 273–324, 1997.

[9] L. Yu and H. Liu, “Efficient feature selection via analysis of relevance and redundancy,” J. Mach. Learn. Res., vol. 5, pp. 1205–1224, 2004.

[10] I.GuyonandA.Elisseeff,“Anintroductiontovariableandfeatureselec- tion,” J. Mach. Learn. Res., vol. 3, pp. 1157–1182, 2003.

[11] X. Liu, A. Krishnan, and A. Mondry, “An entropy-based gene selection method for cancer classification using microarray data,,” BMC Bioinfor- matics, vol. 6, p. 76,,2005.

[12] R.Battiti,“Usingmutualinformationforselectingfeaturesinsupervised neural net learning,” IEEE Trans. Neural Netw., vol. 5, no. 4, pp. 537–550, Jul. 1994.].

[13] J. Zhang and H. Deng, “Gene selection for classification of microarray data based on Bayes error,” BMC Bioinformatics, vol. 8, p. 370, 2007.

[14] C. Ooi, M. Chetty, and S. Teng, “Differential prioritization between relevance and redundancy in correlation-based feature selection techniques for multiclass gene expression data,” BMC Bioinformatics, vol. 7, pp. 320– 339, 2006.

[15] Y. Zhang, C. Ding, and T. Li, “Gene selection algorithm by combining reliefF and mRMR,” BMC Genomics, vol. 9, no. 2, p. S27, 2008.

[16] Bouten C V C, Koekkoek K T M, Verduin M, et al. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity[J]. Biomedical Engineering, IEEE Transactions on, 1997, 44(3): 136-147.

[17] Bao L, Intille S S. Activity recognition from user-annotated acceleration data[M]//Pervasive computing. Springer Berlin Heidelberg, 2004: 1-17.

[18] Peng, Hanchuan. Feature Selection Based on Mutual Information: Criteria of Max-Dependecy, Max-Relevance, and Min-Redundancy. IEE Transaction on Pattern Analysys and Machine Intelligence Vol 27 August, 2005.

[19] Narendra P M, Fukunaga K. A branch and bound algorithm for feature subset selection[J]. Computers, IEEE Transactions on, 1977, 100(9): 917-922.