Identifying Discriminating Features in Time Series Data for Diagnosis of Industrial Machines

Erik Olsson

Department of Computer Science and Electronics, Malardalen University
Erik.M.Olsson@mdh.se

Abstract. Reducing the inherent high dimensionality in time series data is a desirable goal. Algorithms used for classification can easily be misguided if presented with data of to high dimension. E.g. the k-nearest neighbor algorithm which is often used for classification performs best on smaller dimensions with less than 20 attributes. In this paper we address the problem using a time series case base containing previously classified time series measurements. Feature vectors for time series measurements is selected with respect to their discriminating power using an unsupervised feature discrimination approach incorporating statistical feature discrimination. For evaluation, previously classified current measurements from an electrical motor driving a gearbox on an industrial robot were used. The results were promising and we managed to correctly classify measurements from healthy and unhealthy gearboxes.

1 Introduction

Selecting adequate features for classification of time series data can be a time-consuming task that requires good domain knowledge and a tedious manual inspection of the data. Even if adequate domain knowledge is present it may not always be directly applicable due to a noisy sensor environment. Using the original high dimensional and presumably noisy data for classification may cause the "curse of the high dimensionality problem" [1] and result in a misguided matching process due to unwanted computation of similarities between irrelevant features. Individual weighting of important features [2] may be a solution to this problem but it often requires expert knowledge about the relevance of each feature and its impact in the matching process. In this paper we present an unsupervised feature selection algorithm which requires no expert knowledge and no individual weighting of features. It uses a time series case base and a feature discrimination approach incorporating an unsupervised function based on statistical feature discrimination finding features with maximum discriminating power. Feature vectors for time series data measurements is assembled from these features. For evaluation, previously classified current measurements from an electrical motor driving a gearbox on an industrial robot were used. The results were promising and we managed to correctly classify measurements from healthy and unhealthy gearboxes. The paper is organized as follows; section 2 gives some background and related work, section 3 and 4 presents our solution to the problem, section 5 presents an evaluation on real world time series data and section 6 concludes the paper with a discussion and a proposal for future work.

2 Background and Related Work

2.1 Feature Discrimination

Feature discrimination relies on the fact that certain features in time series data has a stronger discriminating power than others. By letting the features with the strongest discriminating power represent the time series we have hopefully achieved a great reduction in dimension and a more qualitative knowledge representation of the data. The reduced representation will thereby stand a better chance to perform well in applications for classification of time series data. E.g. the k-nearest neighbor algorithm which is often used for case-based classification [3] performs best on smaller dimensions with less than 20 attributes. Feature discrimination usually relies on on a criterion function and a search strategy. The search strategy is used to select features and the criterion function is used to evaluate whether a selected feature is better than another. Bayesian probability estimation has been successfully used for criterion [5] and key sequences in synthesized data was found with great accuracy. In [11] several approaches of feature discrimination is discussed. Also the use of a neural network for simultaneous clustering and feature discrimination has been proven useful [12].

3 Computing Feature Vectors for Time-Series Data

We address the dimensionality problem using a time series case base CB containing cases with previously classified time series measurements. Each measurement is first transformed into a time/frequency representation of the original time series data by computing a time FFT [9] transformation. A feature extraction function $FV = GetFeatures(CB_k, \delta)$ is then applied on each time/frequency representation that for a given case CB_k returns a subset of time/frequency elements in feature vector FV representing CB_k in a reduced dimension form. δ is a threshold value defining the criterion for discriminating power of CB_k. δ is found by search and criterion function $\delta = GlobalMaximun(N(CB_k))$. Function $N(CB_k)$ returns the number of fully discriminated cases by CB_k with respect to CB and function $GlobalMaximun(N(CB_k))$ returns the value of δ where $N(CB_k)$ has its global maximum thus representing a maximum of discriminating power of CB_k with respect to CB.

3.1 Extracting Discriminating Features for Case Indexing

Definitions

Definition 1 *We define a time series X of dimension n as a sequence of data points x ordered in time as $X = \{x_1, x_2, ..., x_i, ..., x_n\}$ where x_i refers to a data point at position i.*

Definition 2 *We define $Transform$ to be a function $f : \Re \rightarrow \Re^2$ mapping time series X to a time/frequency matrix A*

$$A = \begin{pmatrix} a_{11} & a_{12} \dots & a_{1j} \dots & a_{1n} \\ a_{21} & a_{22} \dots & a_{2j} \dots & a_{2n} \\ \vdots & \ddots & \vdots \ddots & \vdots \\ a_{i1} & a_{i2} \dots & a_{ij} \dots & a_{in} \\ \vdots & \ddots & \vdots \ddots & \vdots \\ a_{m1} & a_{m2} \dots & a_{mj} \dots & a_{mn} \end{pmatrix} \tag{1}$$

In A each element a_{ij} represents a discrete time/frequency element with time j and frequency i.

Definition 3 *A time series case base CB contains a number of cases where each case is represented by a vector of triplets $\{X, A, C\}$ where X is the original time series measurement, A is the time/frequency representation of the original time series data X and C represents its class. CB_k represents a case k in CB and $CB_{k_{a_ij}}$ represents a time/frequency element in case k.*

Definition 4 *We let the function $std(a, b)$ denote the standard deviation function returning the standard deviation of time/frequency elements a and b and we define the function $Threshold(a, b)$ to be:*

$$Threshold(a, b) = \begin{cases} 1 & \text{if } std(a, b) < \delta \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Definition 5 *We define case CB_k to be fully discriminated by case CB_l (and the opposite) if there exists a $\delta > 0$ such as*

$$\left[\sum_{i=1}^{m} \sum_{j=1}^{n} Threshold(CB_{k_{a_ij}}, CB_{l_{b_ij}}) \right] = 0, k \neq l \tag{3}$$

Definition 6 *We let the function $Discriminate(CB_k, CB_l)$ to be defined as*

$$Discriminate(CB_k, CB_l) = \begin{cases} 1 & \text{if } CB_k \text{ is fully discriminated by } CB_l \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Definition 7 *We define a measurement of discriminating power of a case CB_k to be the sum of all the cases in the case base it fully discriminates*

$$N(CB_k) = \sum_{l=1}^{n} Discriminate(CB_k, CB_l), k \neq l \tag{5}$$

where $N(CB_k)$ *denotes the number of fully discriminated cases by* CB_k *with respect to* CB.

We now want to extract the time/frequency elements from CB_k that represents the strongest discriminating power with respect to CB. The first step is to find δ where $N(CB_k)$ has its global maximum. By definition, a global maximum must be either a local maximum in the interior of the domain of $N(CB_k)$ or it must lie on the boundary of its domain [10]. The domain of $N(CB_k)$ is all positive real values $\delta > 0$ but we can limit the domain to $\delta = (c, d)$ where $c = 0, d \geq c, \delta = d \rightarrow N(CB_k) = 0$. We solve this with the search and criterion function $\delta = GlobalMaximun(N(CB_k))$.

```
SET max=0
SET ret=0
FOR δ = c TO d
      IF N(CB_k) > max
          SET max = N(CB_k)
          SET ret = δ
      END
END
RETURN ret
```

Fig. 1. Code for finding the global maximum of $N(CB_k)$

Definition 8 *If we let the function* $GetFeatures(CB_k, \delta)$ *be the function returning a set of time/frequency elements from case* CB_k *that for each case in* CB *satisfies (6).*

$$GetFeatures(CB_k, \delta) = \begin{cases} CB_{k_{a_{ij}}} & \textit{if } std(CB_{k_{a_{ij}}}, CB_{l_{b_{ij}}}) > \delta, k \neq l \\ 0 & \textit{otherwise} \end{cases} \quad (6)$$

then

$$FV = GetFeatures(CB_k, GlobalMaximun(N(CB_k))) \quad (7)$$

will produce a feature vector $FV = \{F_1, F_2, ..., F_i, ..., F_m\}$ *representing all time/frequency elements with the discriminating power* δ.

4 Case Indexing

In order to use FV for case indexing we want an appropriate representation of the features. We use a naive binary structure [5] in combination with sequence appearance numbers. In this case we reduce the time dimension and save only frequency information in our vector. We recalculate FV given in (7) and represents it as in (8)

$$FV = \{b_1 * f_1, b_2 * f_2, ..., b_i * f_i, ..., b_m * f_m\} \tag{8}$$

where b_i denotes the number of occurrences of similar frequency elements and f_i denotes the frequency.

For similarity measure between two time series we use the Euclidian distance function defined as

$$sim(FV_1, FV_2) = \sqrt{\sum_{i=1}^{m} (FV_{1i} - FV_{2i})^2} \tag{9}$$

5 Example Implementation and Evaluation

In order to evaluate our framework we tested it on pre-classified current time series data from an electrical motor driving the gearbox of axis 4 on an industrial robot. A total of 40 classified measurements where used in the evaluation. Our goal was to compute feature vectors that were able to discern a healthy gearbox from an unhealthy gearbox.

5.1 Measuring Current Time-Series

The robot control cabinet can log current signals from the electrical motors driving the gearboxes on the robot. We programmed the cabinet to log current signals from the electrical motor driving the gearbox of axis 4. The signal is passing through the robot control system and is an indirect measurement of current [4] derived from the motor torque. Because of its indirect nature, basic theory of feature selection usually applied to current measurements is difficult to apply here. We thereby find this time-series data especially suitable for evaluation of our framework.

Each time-series measurement is approximately 3.8 seconds long and involves a full rotation of the robot arm. The sampling rate is 2 kHz which result in a time-series measurement containing approximately 7600 samples (see Fig. 2).

5.2 Classification of Time-Series

During the end test of the industrial robots we logged current signals from 34 normal gearboxes and 6 noisy gearboxes. All robots where classified by experts. Table 1 presents the experts classifications.

Based on the information in table 1 we classified our measurements in five classes; C_{Normal}, $C_{RepeatedKnocks}$, C_{Knock}, C_{Burr} and C_{Noise} as for normal, repeated knocks, a single knock, burr and noise respectively. Classified cases where created for all measurements and inserted into the case library.

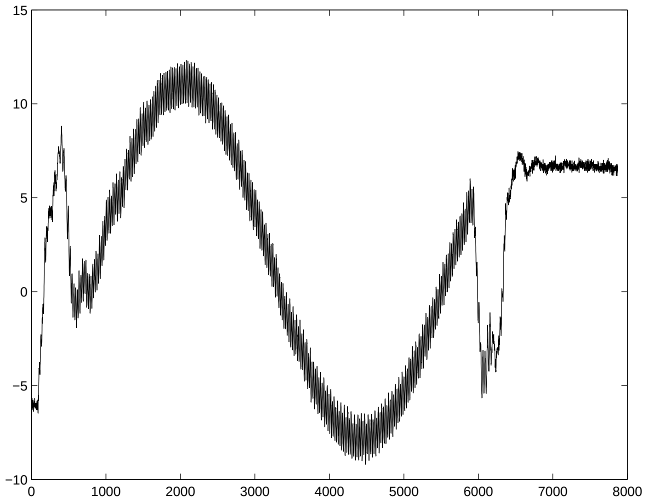

Fig. 2. An example of a time series of indirect measurements of motor current.

Table 1. Robots classified by human expert.

Class	Number of robots
No symptom (normal)	34
Repeated Knocks	3
Knock	1
Burr	1
Noise	1

5.3 Computing Feature Vectors

Before computing feature vectors for the classified cases we pre-processed the data in each case computing a time FFT [9] matrix A (see Defn. 2) on each time series current measurement X (see Defn. 1). The time FFT was computed with a precision of 46 time-segments and 169 frequency steps for each measurement resulting in an approximate time/frequency resolution of 83 ms/segment and 7 Hz/step.

After computing feature vectors for all cases as explained in section 3 and performing a leave-one-out k-nearest neighbor evaluation [8] with $k = 3$ on all cases as explained in section 4 (see Fig. 3). The result is presented in table 2.

We managed, with 100 percent accuracy, to correctly classify all cases with class C_{Normal} and $C_{RepeatedKnocks}$. We failed to classify C_{Knock}, C_{Burr} and C_{Noise}. A reason for that is given in the next chapter.

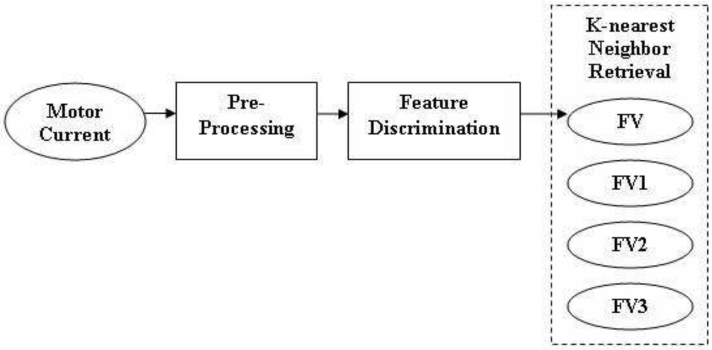

Fig. 3. System for k-nearest neighbor evaluation

Table 2. Robots classified by system.

Class	Percentage of correct hits
No symptom (normal)	100
Repeated Knocks	100
Knock	0
Burr	0
Noise	0

6 Conclusions and Future Work

Our approach of feature selection by feature discrimination proves to be useful for machine sensor time series. It shows that the method can be valuable on already classified time series which lacks of other useful domain knowledge or where domain knowledge is hard to apply. We managed to compute discriminating feature vectors and correctly classify the two dominating classes C_{Normal} and $C_{RepeatedKnocks}$. We failed to classify C_{Knock}, C_{Burr} and C_{Noise} because we had no similar cases in the case base. The leave-one-out k-nearest neighbor evaluation approach demands several cases of similar class in order to successfully classify all cases. A larger case base with more cases of class C_{Knock}, C_{Burr} and C_{Noise} would probably perform better in classification. Some suggestions for future work is stated below.

1. Expand the case base with more classified cases.
2. Evaluate the performance of the algorithm on other kinds of time series data.
3. Test other classification algorithms such as Self Organizing Maps (SOM) [6], the cosine matching function [7] etc.

References

1. Mitchell T.: Machine Learning. (1997) ISBN 978-0071154673
2. Wettschereck D., Aha, D.: Weighting features. In Proceedings of the 1st International Conference on Case-based Reasoning. Springer (1995) 347–358
3. Aamodt A., Plaza E.: Case-based reasoning: foundational issues, methodological variations, and system approaches. Artificial Intelligence Com **7** (1994) 39–59
4. Wallin, K.: Matdatorer ABB internal report (1997)
5. Funk P., Xiong N.: Discocvering Knowledge About Key Sequences for Indexing Time Series Cases in Medical Applications. Advances in Case-Based Reasoning. (2006)
6. Kohonen T.: Self-Organizing Maps. Springer Series in Information Sciences. **30** (2001) ISBN 3-540-67921-9
7. Plaza E., Lopez de Mantaras R.: A case-based apprentice that learns from fuzzy examples. Methodologies for Intelligent Systems. Elsevier **5** (1990) 420–427
8. Belur V.: Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques. (1991) ISBN 0-8186-8930-7
9. Bachman G., Lawrence N., Beckenstein E.:Fourier and Wavelet Analysis. ISBN 0-387-98899-8
10. Adams R.: Calculus: A Complete Course. (1995) ISBN 0-201-82823-5
11. Liakat M., A.:Feature Selection of DNA Microarray Data. Seminar and Computing Tools. (2005) http://web2.uwindsor.ca/courses/cs/aggarwal/cs60520/SeminarMaterial/FeatureSelectionDNAData.pdf
12. Hichem F., Nozha B., Soon-Ann L.:Unsupervised Clustering and Feature Discrimination with Application to Image Database Categorization. Digital Object Identifier **1** (2001) 401–406