

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Sipola, Tuomo; Ristaniemi, Tapani; Averbuch, Amir

Title: Gear classification and fault detection using a diffusion map framework

Year: 2015

Version: Accepted version (Final draft)

Copyright: © 2015 Elsevier

Rights: CC BY-NC-ND 4.0

Rights url: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Please cite the original version:

Sipola, T., Ristaniemi, T., & Averbuch, A. (2015). Gear classification and fault detection using a diffusion map framework. *Pattern Recognition Letters*, 53(1 February 2015), 53-61.

<https://doi.org/10.1016/j.patrec.2014.10.019>

Gear classification and fault detection using a diffusion map framework

Tuomo Sipola^{a,1,*}, Tapani Ristaniemi^a, Amir Averbuch^b

^a*Department of Mathematical Information Technology, University of Jyväskylä, Finland*

^b*School of Computer Science, Tel Aviv University, Israel*

Abstract

This article proposes a system health monitoring approach that detects abnormal behavior of machines. Diffusion map is used to reduce the dimensionality of training data, which facilitates the classification of newly arriving measurements. The new measurements are handled with Nyström extension. The method is trained and tested with real gear monitoring data from several windmill parks. A machine health index is proposed, showing that data recordings can be classified as working or failing using dimensionality reduction and warning levels in the low dimensional space. The proposed approach can be used with any system that produces high-dimensional measurement data.

Keywords: system health monitoring, diffusion map, clustering, fault detection

1. Introduction

Modern industry monitoring systems produce high-dimensional data that are difficult to analyze as a whole without dimensionality reduction. The goal of the study is to estimate whether the proposed dimensionality reduction scheme effectively distinguishes working gears from broken ones. System health management has multiple sensors that measure vibration, temperature and oil properties. The early detection of anomalous gear behavior using this sensor data reduces the risk of severe damage. Sensor data are then used to monitor the health of the system, to detect anomalies and to predict problems (Chandola et al., 2009, pp. 15–16).

Anomaly detection methods try to find deviant or atypical measurements from a large data mass (Chandola et al., 2009). In this study known anomalies are in the training so that they can be contrasted with the normal behavior. An ideal indicator would tell with certainty that a machine works or is going to fail. However, in reality the non-working state is ambiguous and it can be difficult to classify.

*Corresponding author. Tel. +358 40 753 2169. Postal address: c/o Tapani Ristaniemi, Department of Mathematical Information Technology, P.O. Box 35 (Agora), FI-40014 University of Jyväskylä, Finland.

Email addresses: `tuomo.sipola@iki.fi` (Tuomo Sipola), `tapani.ristaniemi@jyu.fi` (Tapani Ristaniemi), `amir@math.tau.ac.il` (Amir Averbuch)

¹Current affiliation is with CAP Data Technologies and Koodilehto.

Spectral dimensionality reduction methods include principal component analysis (PCA), kernel PCA, multi-dimensional scaling (MDS), Laplacian eigenmaps, isomap and locally linear embedding (LLE). These methods facilitate the analysis of high-dimensional data by mapping the high-dimensional coordinates to a lower dimension. The spectral approach also leads to the concept of spectral clustering (Bengio et al., 2006; von Luxburg, 2007). Spectral methods have been used to analyze system operational states (Pylvänen et al., 2009), motor fault detection (Parra et al., 1996) and anomaly detection for spacecraft (Fujimaki et al., 2005).

Traditionally, frequency analysis has been used to determine the health status of gears or bearings. This study uses frequency information directly from sensors embedded to the gear. Analyzing the frequency using diffusion map and thresholding data produces a classification of working and failing gears.

Recently, wavelet decomposition has been used in a similar fashion to produce a frequency spectrum growth index with a specialized threshold function (Wang et al., 2009). Furthermore, lifting wavelet packet decomposition has been used to extract node energies at the third level of the decomposition as features. This way fuzzy c-means can be trained using normal and failure data. The state of incoming new data is then determined relative to the found clusters (Pan et al., 2010). Another health monitoring index is the average probability index (API), which utilizes empirical mode decomposition for time-frequency features and hidden Markov model for stochastic estimation (Miao et al., 2010). Support vector data description has been used for early fault detection and fault diagnosis (Wang et al., 2011). Wavelet-based filtering has also produced health condition indicator for cooling fan bearings (Miao et al., 2012). Finally, a study using diffusion map has been made concerning machine condition monitoring (Huang et al., 2013).

This study uses diffusion map, which is another spectral dimensionality reduction method. Its mathematical foundation is random walk on Markov transition matrix of the graph of the data (Coifman and Lafon, 2006a). Diffusion map can be classified as a nonlinear distance-preserving dimensionality reduction method that preserves global properties (van der Maaten et al., 2009). Furthermore, the Nyström method is used to extend new points, although newer methods such as geometric harmonics exist (Fowlkes et al., 2004; Coifman and Lafon, 2006b).

This study presents a way to detect faults in windmill gears by devising an index to describe how close to the faulty state a gear is. Besides gear fault detection, this method can also be used with other collections of high-dimensional time series data.

2. Methodology

This method trains a diffusion map that describes the good and bad state of the gears in order to create a health management index that predicts faults in gears. It then extends newly arriving test measurements to the model and classifies the gear as good or bad. Most of the preprocessing is domain specific, but the dimensionality reduction and classification, that are more universally applicable, are presented here. Figure 1 introduces the overall data processing architecture. The equations are in matrix form. The details behind them are discussed elsewhere (Nadler et al., 2008; Fowlkes et al., 2004; Belongie et al., 2002).

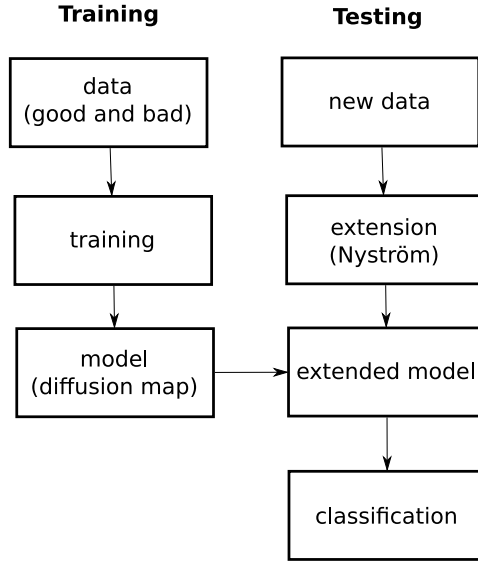


Figure 1: Data processing block diagram.

2.1. Training dimensionality reduction

The underlying assumption in manifold learning methods is that the data is situated on a lower-dimension manifold in the high-dimension measurement data (Chandola et al., 2009, p. 37). We try to create a function that maps the behavior of high-dimensional points to lower dimensions. Then new measurement points are mapped from high dimensions to this low-dimensional presentation (Coifman and Lafon, 2006a).

Let $\mathbf{x}_i \in \mathbb{R}^n, i = 1 \dots N$ be a measurement in n -dimensional space. The kernel matrix \mathbf{W} includes the pairwise distances of these points. The used kernel is the Gaussian kernel using Euclidean distance measure. This is the most computationally intensive step because each point is compared to other points.

$$\mathbf{W}_{ij} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\epsilon}\right) \quad (1)$$

Determining ϵ is a problem in itself. The chosen estimation is the median of the distances between the points, $\epsilon = \text{median}\{\|\mathbf{x}_i - \mathbf{x}_j\|\}_{\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^n}$ (Schclar et al., 2010). Depending on the problem, changing this parameter might give more meaningful results.

The diagonal matrix $\mathbf{D}_{ii} = \sum_{j=1}^N \mathbf{W}_{ij}$ has the degree of each point on its diagonal, the other elements are 0. The degree of a point is the sum of weights that connect to other points. This is equal to the sum of kernel matrix rows.

The rows are normalized by these sums. The result can also be understood as transition probabilities between points. These probabilities are collected in matrix \mathbf{P} .

$$\mathbf{P} = \mathbf{D}^{-1}\mathbf{W} \quad (2)$$

However, future calculations on \mathbf{P} become easier if a similarity transformation symmetrizes the matrix.

$$\tilde{\mathbf{P}} = \mathbf{D}^{\frac{1}{2}} \mathbf{P} \mathbf{D}^{-\frac{1}{2}} \quad (3)$$

These last two steps can be combined. Substituting \mathbf{P} with $\mathbf{D}^{-1} \mathbf{W}$ yields

$$\tilde{\mathbf{P}} = \mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{-\frac{1}{2}}. \quad (4)$$

Such normal matrix is decomposed as

$$\tilde{\mathbf{P}} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^*. \quad (5)$$

This decomposition is done using singular value decomposition (SVD). The columns of matrix \mathbf{U} contain eigenvectors \mathbf{u}_k of matrix $\tilde{\mathbf{P}}$. Likewise, the diagonal of $\mathbf{\Lambda}$ contains its corresponding eigenvalues. However, the real interest is in the eigenvectors of the transition matrix \mathbf{P} . The eigenvalues of \mathbf{P} are the same, but the eigenvectors are obtained from

$$\mathbf{V} = \mathbf{D}^{-\frac{1}{2}} \mathbf{U}. \quad (6)$$

Recall that the eigenvalues λ_k are in the diagonal of $\mathbf{\Lambda}$. The eigenvector \mathbf{v} are columns of \mathbf{V} . An original data point \mathbf{x}_i has a corresponding value on the i th row of the eigenvector. For example, $\mathbf{v}_2(\mathbf{x}_{236})$ would signify the second eigenvector and its 236th row, corresponding to the 236th sample \mathbf{x}_{236} of the original dataset.

The diffusion map itself is a function in the form $\Psi : \mathbb{R}^n \rightarrow \mathbb{R}^d$, when $d \ll n$. We multiply the eigenvectors and eigenvalues to get the diffusion coordinates of the training points in the corresponding matrix:

$$\Psi = \mathbf{V} \mathbf{\Lambda}. \quad (7)$$

The first eigenvector is constant, so only the following eigenvectors and eigenvalues are used. This way we get the following function that maps the original data points to a lower-dimensional space.

$$\Psi : \mathbf{x}_i \rightarrow \begin{pmatrix} \lambda_2 \mathbf{v}_2(\mathbf{x}_i) \\ \lambda_3 \mathbf{v}_3(\mathbf{x}_i) \\ \lambda_4 \mathbf{v}_4(\mathbf{x}_i) \\ \vdots \\ \lambda_{d+1} \mathbf{v}_{d+1}(\mathbf{x}_i) \end{pmatrix} \quad (8)$$

It has been shown that the diffusion distance in the original space equals to the Euclidean distance in the diffusion space (Coifman and Lafon, 2006a). Thus, the distance measurements in the diffusion space are actually meaningful and can be used in further analysis in this lower-dimensional space.

Later analysis uses only the first few diffusion coordinates. Fast decay of eigenvalues leaves most of the diffusion coordinates rather small compared to the first few. The overall reconstruction of \mathbf{P} does not differ much from a reconstruction that uses only the first coordinates. These coordinates capture most of the differences between the data points (Coifman and Lafon, 2006a; Nadler et al., 2006).

2.2. Extension of new measurements

New measurements that are not part of training are extended to the model with Nyström method (Fowlkes et al., 2004; Belongie et al., 2002). The features selected during training are the only ones needed. These new measurements are normalized using the same normalization as during training.

Let a new data point be $\mathbf{y}_j \in \mathbb{R}^n$. Then the distance between the new points and each training point are collected in a matrix $\bar{\mathbf{W}}$. This function uses the same ϵ as the one in training phase.

$$\bar{\mathbf{W}}_{ij} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{y}_j\|^2}{\epsilon}\right) \quad (9)$$

Diagonal matrix $\bar{\mathbf{D}}_{ii} = \sum_{i=1}^N \bar{\mathbf{W}}_{ij}$ contains the column sums of $\bar{\mathbf{W}}$. Now we can create the transition probability matrix \mathbf{B} .

$$\mathbf{B} = \bar{\mathbf{W}}\bar{\mathbf{D}}^{-1} \quad (10)$$

The following matrix multiplication produces new eigenvectors for the new point. The eigenvectors \mathbf{V} and eigenvalues $\mathbf{\Lambda}$ are the same as in training.

$$\bar{\mathbf{V}} = \mathbf{B}^T \mathbf{V} \mathbf{\Lambda}^{-1} \quad (11)$$

These new eigenvectors now extend the new point to the diffusion coordinates.

$$\bar{\mathbf{\Psi}} = \bar{\mathbf{V}} \mathbf{\Lambda} \quad (12)$$

The last two steps can be combined.

$$\bar{\mathbf{\Psi}} = \mathbf{B}^T \mathbf{V} \quad (13)$$

Matrix $\bar{\mathbf{\Psi}}$ now contains the extended eigenvectors in its columns for the new points \mathbf{y}_j .

2.3. Classification of new measurements

Low-dimensional presentation of the data facilitates clustering. The clustering approach here is spectral clustering and it reveals the normal and anomalous areas (von Luxburg, 2007; Kannan et al., 2004). Any other clustering, for example k -means, can be used if they provide better results (Ng et al., 2001; Meila and Shi, 2000; Shi and Malik, 2000). The used algorithm simply tests whether the sample is to the left or to the right of 0 on the dimension corresponding to the 2nd eigenvector. This provides a classifier that discriminates two states: working or broken.

2.4. Health management index and warning levels

The 2nd eigenvector, corresponding to Ψ_1 , is the health management index. The working machines are above 0 and the failing machine below it. For more warning levels, different thresholds θ can be applied the coordinate Ψ_1 corresponding to the 2nd eigenvector. There are three warning levels: note, warning and damage. These describe the severity of the problem in the gear.

Note means that there is an unusual measurement in the data, but the gear is still in operational state. The sample is not inside the good cluster but is still closer to it than to the bad.

$$\theta_{note} = \min\{\Psi_{1,good}\} \quad (14)$$

Warning level is at $\theta_{warning} = 0$. It describes the border between good and bad clusters. The sample is closer to the bad cluster. This can be seen as a predictive sign that the gear has problems. If the bad cluster goes beyond 0, the middle point between the two clusters can be used.

Damage level is at $\theta_{damage} = \max\{\Psi_{1,bad}\}$. This level means that the sample is within the bad cluster.

3. Results

This study uses a dataset consisting of gear monitoring recordings of multiple features. These data are collected during normal operations from windmill gears from nine different windmill parks. Naturally, the operating conditions vary among the parks. The dataset consists of recordings of 18 good and 20 bad machines labeled by domain specialists. As stated, the gears come from different locations where the operational environment varies. However, each gear is of the same type and includes same features. Two of the gears are discarded because they contain empty data due to instrument failures. The dataset is divided to training and testing sets. The training set includes five good and five bad gears. The testing set includes the rest of the gears.

The sampling rate of the measurements is averaged measurement once in half an hour. Each of the windmill gears were monitored for months, which brings the sample size for each gear to about 1500 per month.

3.1. Preprocessing

The data are sampled at an approximate frequency of one sample per 30 minutes representing the average of measurements during that time. The recordings last for months. Because there were times when no data were available, linear interpolation is used. This data formed the samples \times features matrix.

Instrument failures give unrealistic or missing measurements. Because it is difficult to compare such measurements to ones that do not have unrealistic values, measurements containing missing values are discarded. However, this process might lose some usable information.

3.1.1. RPM filtering

Samples whose rotations per minute (RPM) value is too small are filtered out, because only higher values represent the actual working state of a gear. Lower values are associated with idle state, and those measurements are not interesting when monitoring actual working gears. The RPM values are clustered into two clusters using k -means clustering, resulting in clusters of samples named $RPM_{cluster\ 1}$ and $RPM_{cluster\ 2}$. The threshold value

$$threshold_{RPM} = \max\{\min\{RPM_{cluster\ 1}\}, \min\{RPM_{cluster\ 2}\}\}; \quad (15)$$

is calculated and all the samples whose RPM value is below this threshold are removed.

3.1.2. Data scaling

All the data are normalized with logarithm. Other normalizations, like dividing by maximum or dividing by norm, do not give as good separation for this dataset.

3.1.3. Feature selection

A preliminary feature selection in the original feature space is done. One feature is left out at a time. The average Mahalanobis distance between the good and bad machines shows how much that feature describes the difference. The features with smallest averaged Mahalanobis distances are most useful. Small distance reveals that leaving the feature out affects negatively the separation of good and bad. Thus, using the feature separates the groups well. The number of reduced features is determined by including the most separating features.

There are 136 features that describe oil pressure, oil particles, RPM, and the frequency of multiple vibration sensors in three dimensions. The initial feature selection reduced their number to 20. One feature is the oil pressure, five from the planetary stage of the gear, seven from the intermediate stage, six from the high speed stage, and one is the RPM. These features separate more clearly the two groups from each other.

3.2. Classification results

Five good and five bad gears were used in training. The original data has 136 features, 20 of which are used after preliminary feature selection. The diffusion map and the Nytsröm extension were performed using the 20 most separating features. All the gears, including training gears, were then tested as new incoming data. Table 1 shows that each of the failing test gears had alerts. Table 2 shows that no working gear had warnings, although some of them had notes. The alert percentage tells how many per cent of samples during the monitoring time gave alerts. The letter combinations AH, CA, ET, FE, LS, MB, OO, PH and QU refer to the nine windmill parks where the data were collected. The numbers following the letters identify the individual windmills. These tables show that all failing gear units had alerts during the monitoring time. None of the working gear units had alerts. This means that there were no false alarms in the working gears.

The following figures illustrate the behavior of failing gears. Normal state does not produce figures of interest because there are no alerts. Figure 2 shows how the newly incoming data is situated in low-dimensional space. This figure is mainly for

Table 1: Failing gear units (alert threshold 0). Asterisk marks training gears.

| gear | alerts |
|--------|----------|
| OO03 | 2.5703% |
| *OO06 | 14.4068% |
| OO08 | 42.6573% |
| OO09 | 6.6667% |
| AH01 | 16.835% |
| AH02 | 16.7431% |
| AH06 | 2.8777% |
| *AH11 | 14.916% |
| AH18 | 7.0941% |
| FE09 | 5.3495% |
| FE10 | 4.4068% |
| *FE12 | 16.0083% |
| CA03 | 22.7599% |
| CA04 | 7.7089% |
| QU23 | 6.0469% |
| *QU32 | 21.6535% |
| ET104 | 0.32841% |
| *ET403 | 3.3597% |
| PH05 | 9.3694% |

Table 2: Working gear units (alert threshold 0). Asterisk marks training gears.

| gear | alerts |
|-------|--------|
| *AH10 | 0% |
| AH16 | 0% |
| AH18 | 0% |
| FE01 | 0% |
| *FE02 | 0% |
| FE03 | 0% |
| CA01 | 0% |
| LS04 | 0% |
| *LS05 | 0% |
| MB08 | 0% |
| QU32 | 0% |
| *PH01 | 0% |
| PH03 | 0% |
| PH05 | 0% |
| PH08 | 0% |
| *PH09 | 0% |
| PH13 | 0% |

illustrative purposes, although it might be useful for monitoring. Figure 3 shows the alert index, while Figure 4 indicates the accumulating number of alerts. The images show the length of the measurement, from February to January. These two indicators should be of use for technicians monitoring the gears. The alerts themselves are in Figure 5. Figures 6, 7, 8, 9 show the same measurements for another gear. It breaks down more slowly but the high number of notes can be seen.

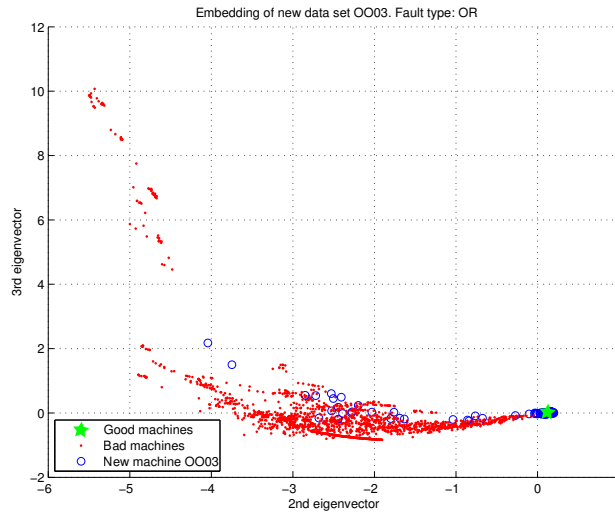


Figure 2: Samples of a failing gear in low-dimensional space.

The misclassification of good machine FE01 as bad is probably because of the data interpolation. Further domain analysis revealed that there actually had been a small problem with the gear, and thus raises the question whether it is labeled correctly. The misclassification of bad machine ET104 as good can be explained. Firstly, there are no training gears from this location. ET104 is too close to the good gears in diffusion space. Secondly, domain analysis reveals that this gear has only a small problem. Better training data and detailed labeling could prevent this kind of misclassification. Vastly different operating environment and behavior of gears in ET1 might also cause this misclassification.

4. Conclusion

The goal of this study is to estimate the usefulness of dimensionality reduction methods in windmill gear fault detection. This goal is met since almost all the gears are classified correctly according to their ground truth labels. The state of failing machines can be monitored using note, warning and damage indicators. This proves that the training is successful and separates the good gears from the bad. More importantly, measurements from new gears can be extended into the model.

This study used diffusion map as the dimensionality reduction method. Non-linear methods, such as principal component analysis, could be used. However, diffusion

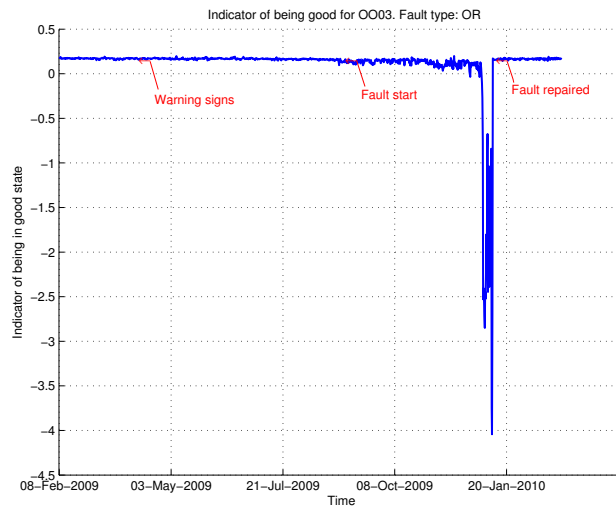


Figure 3: Alert level index of a failing gear over time. According to our method, above 0 is considered normal working state. The arrows pointing to warning signs, faults and repaired states, are a priori knowledge given by the domain experts.

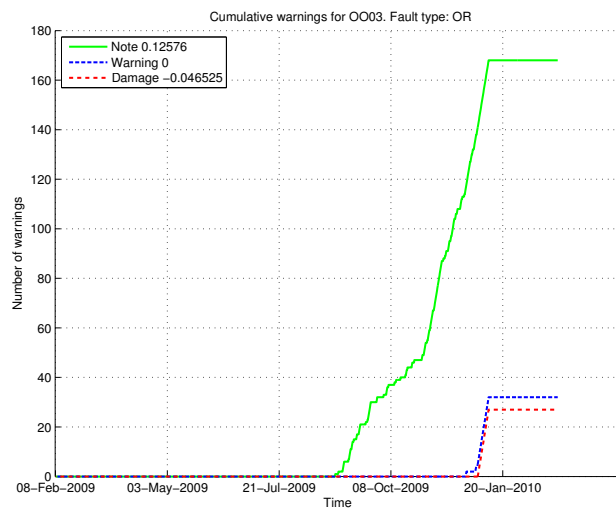


Figure 4: Number of alerts.

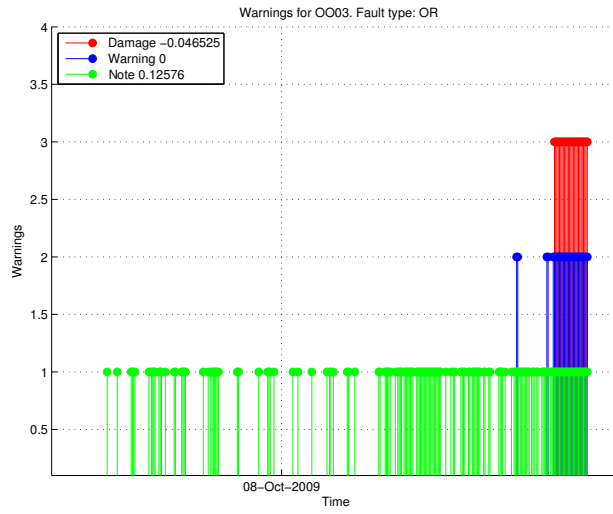


Figure 5: Alerts given by the method.

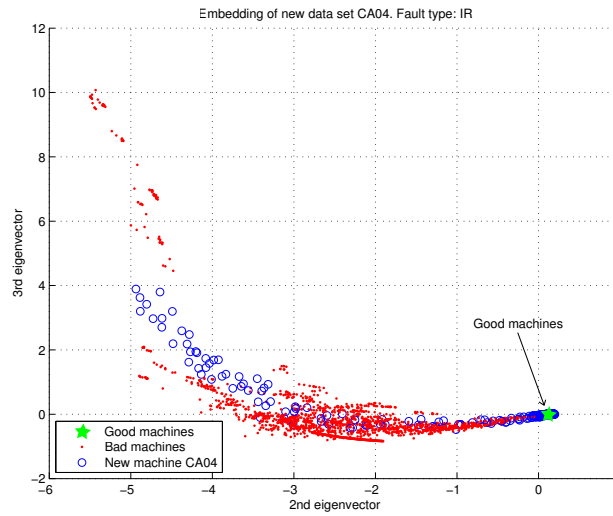


Figure 6: Samples of a failing gear in low-dimensional space.

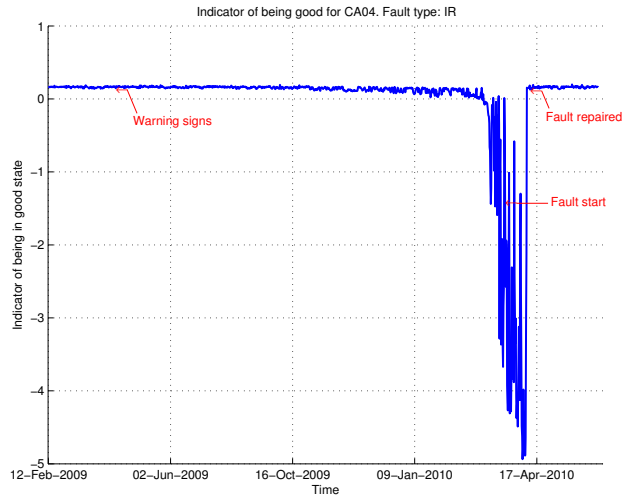


Figure 7: Alert level index of a failing gear over time. According to our method, above 0 is considered normal working state. The arrows pointing to warning signs, faults and repaired states, are a priori knowledge given by the domain experts.

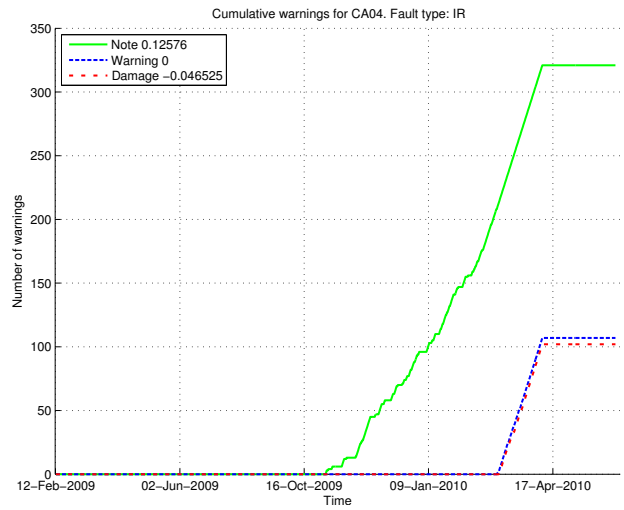


Figure 8: Number of alerts.

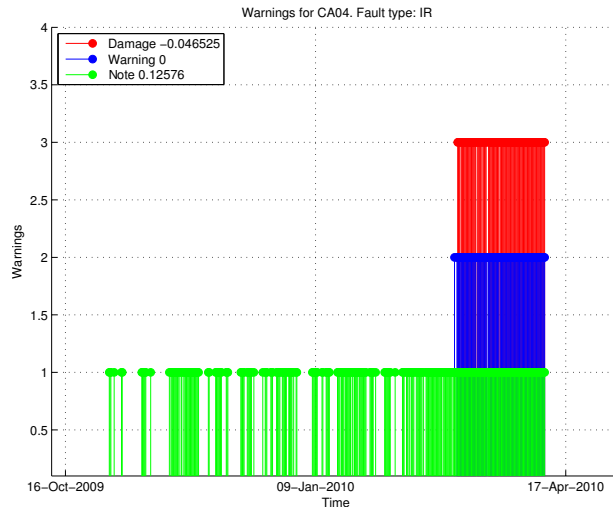


Figure 9: Alerts given by the method.

map has the additional kernel parameter to control the compactness of the clusters. The choice of the dimensionality reduction method is up to the users and their specific needs.

The training can be done within an hour. Memory usage is high during the training because it is dependent on the sample size. With such a low sampling rate (1 per 30 min) there is no problem with computational time when adding new data points.

The challenges caused by spectral methods in general need some addressing. The proposed method works because, after slight filtering, the good and bad gears are separable in the lower dimensions. However, the high computational cost could be a problem in a more real-time system. Moreover, the classification of a gear time series itself is an ambiguous concept. However, this study shows that gears in normal condition and gears that are going to break down behave differently and can be separated from each other.

Acknowledgements

This research has been funded by Tekes grants 40040/08 and 40264/12.

References

Belongie, S., Fowlkes, C., Chung, F., Malik, J., 2002. Spectral partitioning with indefinite kernels using the Nyström extension, in: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (Eds.), *Computer Vision – ECCV 2002*. Springer Berlin / Heidelberg, volume 2352 of *Lecture Notes in Computer Science*, pp. 51–57.

- Bengio, Y., Delalleau, O., Roux, N.L., Paiement, J.F., Vincent, P., Ouimet, M., 2006. Feature Extraction. Springer Berlin, Heidelberg. chapter Spectral Dimensionality Reduction. *Studies in Fuzziness and Soft Computing*, pp. 519–550.
- Chandola, V., Banerjee, A., Kumar, V., 2009. Anomaly detection: A survey. *ACM Comput. Surv.* 41, 15:1–15:58.
- Coifman, R.R., Lafon, S., 2006a. Diffusion maps. *Applied and Computational Harmonic Analysis* 21, 5–30.
- Coifman, R.R., Lafon, S., 2006b. Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions. *Applied and Computational Harmonic Analysis* 21, 31 – 52. *Diffusion Maps and Wavelets*.
- Fowlkes, C., Belongie, S., Chung, F., Malik, J., 2004. Spectral grouping using the Nystrom method. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 26, 214 –225.
- Fujimaki, R., Yairi, T., Machida, K., 2005. An approach to spacecraft anomaly detection problem using kernel feature space, in: *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, ACM, New York, NY, USA. pp. 401–410.
- Huang, Y., Zha, X.F., Lee, J., Liu, C., 2013. Discriminant diffusion maps analysis: A robust manifold learner for dimensionality reduction and its applications in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing* 34, 277–297.
- Kannan, R., Vempala, S., Vetta, A., 2004. On clusterings: Good, bad and spectral. *J. ACM* 51, 497–515.
- von Luxburg, U., 2007. A tutorial on spectral clustering. *Statistics and Computing* 17, 395–416.
- van der Maaten, L.J.P., Postma, E.O., van Den Herik, H.J., 2009. Dimensionality reduction: A comparative review. *Journal of Machine Learning Research* 10, 1–41.
- Meila, M., Shi, J., 2000. Learning segmentation by random walks, in: *NIPS*, pp. 873–879.
- Miao, Q., Tang, C., Liang, W., Pecht, M., 2012. Health assessment of cooling fan bearings using wavelet-based filtering. *Sensors* 13, 274–291.
- Miao, Q., Wang, D., Pecht, M., 2010. A probabilistic description scheme for rotating machinery health evaluation. *Journal of mechanical science and technology* 24, 2421–2430.
- Nadler, B., Lafon, S., Coifman, R., Kevrekidis, I., 2006. Diffusion maps, spectral clustering and eigenfunctions of Fokker-Planck operators, in: Weiss, Y., Schölkopf, B., Platt, J. (Eds.), *Advances in Neural Information Processing Systems* 18. MIT Press, Cambridge, MA, pp. 955–962.

- Nadler, B., Lafon, S., Coifman, R., Kevrekidis, I.G., 2008. Diffusion maps – a probabilistic interpretation for spectral embedding and clustering algorithms, in: Barth, T.J., Griebel, M., Keyes, D.E., Nieminen, R.M., Roose, D., Schlick, T., Gorban, A.N., Kégl, B., Wunsch, D.C., Zinovyev, A.Y. (Eds.), *Principal Manifolds for Data Visualization and Dimension Reduction*. Springer Berlin Heidelberg. volume 58 of *Lecture Notes in Computational Science and Engineering*, pp. 238–260.
- Ng, A.Y., Jordan, M.I., Weiss, Y., 2001. On spectral clustering: Analysis and an algorithm, in: *Advances in Neural Information Processing Systems 14*, MIT Press. pp. 849–856.
- Pan, Y., Chen, J., Li, X., 2010. Bearing performance degradation assessment based on lifting wavelet packet decomposition and fuzzy c-means. *Mechanical Systems and Signal Processing* 24, 559–566.
- Parra, L., Deco, G., Miesbach, S., 1996. Statistical independence and novelty detection with information preserving nonlinear maps. *Neural Computation* 8, 260–269.
- Pylvänen, M., Äyrämö, S., Kärkkäinen, T., 2009. Visualizing time series state changes with prototype based clustering, in: Kolehmainen, M., Toivanen, P., Beliczynski, B. (Eds.), *Adaptive and Natural Computing Algorithms*. Springer Berlin / Heidelberg. volume 5495 of *Lecture Notes in Computer Science*, pp. 619–628.
- Schlar, A., Averbuch, A., Rabin, N., Zheludev, V., Hochman, K., 2010. A diffusion framework for detection of moving vehicles. *Digital Signal Processing* 20, 111 – 122. doi:DOI : 10.1016/j.dsp.2009.02.002.
- Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22, 888 –905.
- Wang, D., Miao, Q., Kang, R., 2009. Robust health evaluation of gearbox subject to tooth failure with wavelet decomposition. *Journal of Sound and Vibration* 324, 1141–1157.
- Wang, D., Peter, W.T., Guo, W., Miao, Q., 2011. Support vector data description for fusion of multiple health indicators for enhancing gearbox fault diagnosis and prognosis. *Measurement Science and Technology* 22, 025102.