Proceedings of the 12th SPIE: Health Monitoring and Smart Nondestructive Evaluation of Structural and Biological Systems IV, v. 5768, pp. 166-175, 2005.

Feature Selection for Partial Discharge Diagnosis

Weizhong Yan and Kai F. Goebel

GE Global Research Center
Niskayuna, NY 12309
{yan, goebelk} @ crd.ge.com

Abstract

In design of partial discharge (PD) diagnostic systems, finding a set of features corresponding to an optimal classification performance (accuracy and reliability) is critical. A diagnostic system designer typically does not have much difficulty to obtain a decent number of features by applying different feature extraction methods on PD measurements. However, the designer often faces challenges in finding a set of features that give optimal classification performance for the given PD diagnosis problem. The primary reasons for that are: a) features cannot be evaluated individually since feature interaction affects classification performance more significantly than features themselves; and b) optimal features cannot be obtained by simply combining all features from different feature extraction methods since there exist redundant and irrelevant features. This paper attempts to address the challenge by introducing feature selection to PD diagnosis. Through an example this paper demonstrates that feature selection can be an effective and efficient approach for systematically finding a small set of features that correspond to an optimal classification performance of PD diagnostic systems.

Keywords: feature selection; Partial discharge; PD diagnosis; Classification

1. Introduction

Partial discharges (PD) are a local, partial breakdown event that occurs for example, on the surface or inside insulation of electrical products due to possibly minute defects in insulation structure. These defects may be the result of the manufacturing process and/or the result of ageing and mechanical damage of the products. While normal (healthy) condition of insulation gives a baseline level of partial discharge activities, increase of partial discharge activities indicates insulation degradation or faults. Using partial discharge activities for diagnosing insulation defects/faults, generally known as "PD diagnosis" [Gulski (1995)], has played a critical role in condition-based maintenance of insulation systems, especially high voltage insulation systems, such as generators, transformers, and capacitors.

PD diagnosis, a typical classification problem, is to classify measured PD activities into the underlying insulation defects or sources that generate PDs. Like in any diagnostic/classification systems, the key to an accurate and reliable PD diagnosis is a set of high quality features/attributes. These features should represent/capture the characteristics of PD signals. More importantly, these features must posses strong discriminant power so that the classifier designed based on those features gives desired performance. Since PD is a stochastic process, namely, the occurrence of PD very much depends on many factors, such as temperature, pressure, applied voltage, and the test duration [Gulski (1995)], and since PD signals contain noise and interference, PD measurements/signals corresponding to different insulation conditions are almost indistinguishable, i.e., PD diagnosis is a complex classification problem. Thus finding a set of high quality features that give more accurate and reliable classification is even more critical in design of PD diagnostic systems.

In the attempt to find salient features for PD diagnosis, researchers have introduced several different feature extraction methods for deriving features out of PD measurements. These methods include the widely used statistical analysis of phase-resolved PD patterns and more modern methods, such as, fractal analysis [Gulski (1996)] and textual analysis [Rahman et al (2000)], among others. With increasing applications of PD diagnosis to new fields and greater desire for more salient features for higher performance of PD diagnosis, more and more new feature extraction methods continue to emerge.

However, not all of those feature extraction methods always yield effective features for *all* problems. In fact, effectiveness of features from those individual feature extraction methods on classification is highly problem-dependent. That is, features extracted using one method may perform very well for one problem, but may perform poorly for others. Therefore, the designer is responsible for choosing proper methods for the given problem at hand. Ultimately he needs to find a set of features based on those various feature extraction methods, which is optimal for the problem concerned in terms of classification performance. This task can be challenging due to the following two facts:

- The optimal feature set cannot be obtained through individually evaluating different feature extraction methods. Not only does individual evaluation take time and efforts, but also it fails to take into account "interaction" among features, which is typically more important for classification performance than individual features themselves. Features that work well individually may not work well when combined with other features, thus, may need not to be included into the final feature set. Conversely, features that do not perform well within one subset of features may work reasonably well when combined with another subset of features.

- The optimal feature set cannot be obtained by simply combining all features from all different extraction methods. Features from one extraction method may be redundant to those from other methods. Also, not all features extracted are necessarily relevant. The redundant and irrelevant features increase the dimensionality of feature space, thus the complexity of classifier, and may degrade the performance of classification.

This paper attempts to address the problem by introducing feature selection for finding optimal features for PD diagnosis. Feature selection is a process of choosing small subset of features out of a set of candidate features based on certain criteria. Feature selection has been widely used in other fields, such as data mining and pattern recognition. However, feature selection has not been actively studied in PD diagnosis. This paper demonstrates that feature selection performed on a collection of features extracted via different feature extraction methods from PD measurements can reliably and effectively find features that give an optimal or near optimal classification performance of PD diagnosis. Since it is a systematic approach and can be performed in an automated fashion, feature selection also greatly reduces the designer's time on finding optimal features, thus the time on overall design cycle of PD diagnostic systems.

The rest of the paper is organized as follows. Section 2 discusses feature selection methods in general and GA-based wrapper feature selection in particular. A case study is conducted in Section 3, where we describe the lab setup for generating PD data and we extract features by using various methods. Results and discussion are given in Section 4. Section 5 concludes the paper.

2. Feature selection

2.1 Overview

Feature selection is a process of choosing a small subset of features out of a given set of candidate features. Feature selection is an important and indispensable step in classifier design for achieving high classification performance.

Feature selection has been widely studied and documented, for example, by Dash and Liu (1997) who carried out a comprehensive overview of feature selection techniques. Broadly speaking, feature selection methods can be categorized into filter (also called open-loop) methods and wrapper (closed-loop) methods [Langley (1994), Kohavi et al (1997)]. The filter approach selects features as a result of preprocessing based on properties of the data itself independently of the learning algorithm. The wrapper approach, on the other hand, uses the learning algorithm as part of the evaluation. Typically, the wrapper approach gives more accurate results, but is also computationally more expensive.

Mathematically, feature selection is a search problem. It consists of two essential elements that differentiate feature selection methods from one to another. They are 1) search method being employed and 2) evaluation criteria being used.

Depending on which search method is used, feature selection methods can be categorized into (a) <u>optimal search</u> (exhaustive search [Liu and Motoda (1998)] and branch & bound algorithms [Narendra and Fukunaga (1977)]), (b) <u>heuristic search</u> (sequential selection [Kitter (1978)] and floating selection [Pudil, et al (1994)]), (c) <u>random search</u> (genetic algorithms [Yang and Honavar (1998)] and simulated annealing [Doak (1992)]), and (d) <u>weight-based search</u> (fuzzy feature selection [Rezaee, et al (1999)] and neural networks [Setiono and Liu (1997)]). While random and weight-based searches are computationally more efficient, optimal search guarantees an optimal solution.

On the other hand, no matter what search method is used, feature selection has to use evaluation criteria to measure the "goodness" of a particular subset of features. Popular evaluation criteria include (a) distance-based measures, such as Mahalanobis distance [Duda et al (2000)] and Hausdorff distance [Piramuthu (1999)]; (b) entropy measures [Ben-Bassat (1982)]; (c) statistical measures [Ben-Bassat (1982)]; (d) correlation based heuristic measures [Hall (2000)]; e) accuracy measures [John, et al (1994)]; and (f) relevance measures [Wang, et al (1999)].

2.2 GA-based wrapper feature selection for PD diagnosis

To find optimal features for PD diagnosis, an initial feature pool is generated first, which is simply the collection of all features extracted from PD measurement using different feature extraction methods. Feature selection is then performed on this initial feature pool to arrive at an optimal subset of features. Figure 1 shows the overall process of finding optimal features. As can be seen from the figure, this process does not involve evaluating features from individual extraction methods. It is rather a systematic approach that takes full consideration of feature interactions and is more effective and reliable in finding optimal features. For feature selection function, any of the feature selection methods discussed in the previous section can be used. In this paper, GA-based wrapper feature selection is chosen because it provides (near) optimal solutions.

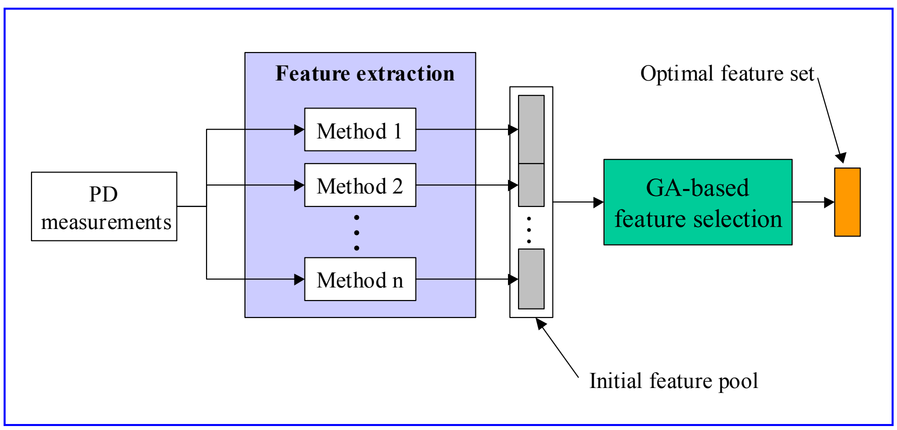

Figure 1: Overall process of finding optimal features

Genetic algorithms (GA) are a derivative-free, stochastic optimization method based loosely on the concepts of natural selection and evolutionary processes, which is well suitable for feature selection. GA-based feature selection was first introduced by Siedlecki & Sklansky in 1989. Since then, GA-based feature selection has been actively studied by numerous researchers, for example, [Yang and Honavar (1998)] and [Raymer et al (2000)]. In GA-based feature selection, a feature set is represented as a binary vector, where each bit is associated with a feature. A value of 1 at the i^{th} bit means the i^{th} feature is included into the feature set while a value of 0 at the i^{th} bit means the i^{th} feature is not included. In each iteration (generation) of the algorithm, a fixed number (population) of possible solutions are generated in a stochastic fashion. Each of the possible solutions is evaluated and modified following the defined genetic operators. Figure 2 illustrates the concept of GA-based feature selection process

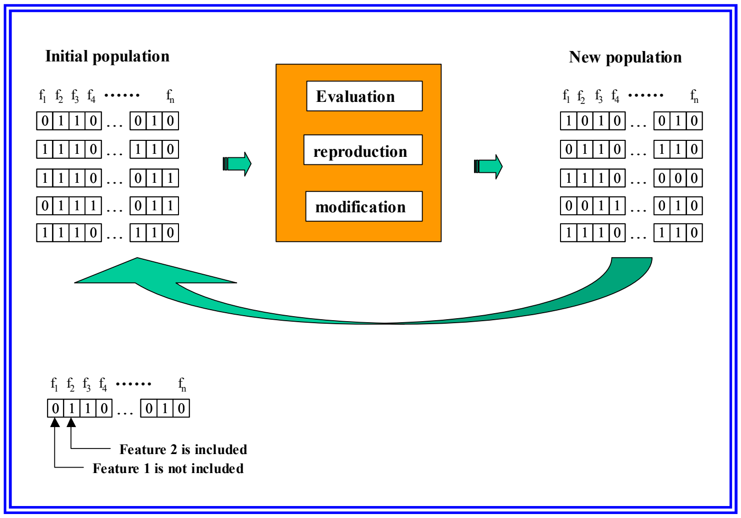

Figure 2: GA-based feature selection

One of important components of GAs is the fitness function that is required for evaluating possible solutions (subsets of features) in each search step. In this paper, the evaluation accuracy of a classifier is used as the fitness function of GAs. Hence, the GA-based feature selection concerned here belongs to a “wrapper approach” as defined in Section 2.1, which typically gives more accurate solutions than filter approach.

The classifier we used in the GA-based feature selection is a support vector machine (SVM) classifier. SVMs are a recently developed learning system originated from the statistical learning theory [Vapnik (1995). Compared to other classifiers, SVM classifiers have several unique properties. Two of these properties, namely, comparatively fast training and good average performance over a wide spectrum of different classification problems, in particular, make SVM classifiers a preferred tool for GA-based wrapper feature selection.

3. A case study

To demonstrate that feature selection is effective in finding optimal features for PD diagnosis, the GA-based wrapper feature selection is used for designing a PD diagnostic system for diagnosing aircraft wiring insulation faults.

3.1. Test setup and PD data generation

PD measurements for aircraft wiring are generated and recorded through laboratory tests. Damage to wiring insulation generally takes two forms: material degradation due to aging or thermal/electrical environment, and chafing that may occur during maintenance and mechanical abrasion during operation-induced vibration. The lab tests conducted in this paper focus on the later, i.e., wire chafing. For classifier design purpose, two wiring conditions are tested. One is for normal condition wires and another for wires with artificial chafing. Two different ways are considered in producing artificial defects. The first type represents defects occurring in twisted pair wires. For that a small piece of the upper insulation layer was removed from both wires in the twisted pair. This type of defect is the most reproducible as the thickness of the remaining insulator layer was not changed. Such samples were prepared from two aircraft grade wires, types: M22759/90-22-95 and M22759/81-22-52. The second type of defect simulated a chafed wire touching the shielding of the cable bundle or a metal part of the aircraft. The wire was chafed in a short length and a piece of tinned copper wire was twisted around the chafed area. The tinned copper wire was also pressed into the chafed section in order to touch the remained insulation layer. Such samples were prepared from three aircraft grade wires, types: M22759/90-22-95, M22759/81-22-52 and M81044/6-22-9.

The schematic of the setup is shown in Figure 3. The potential difference applied is 60-Hz AC voltage.

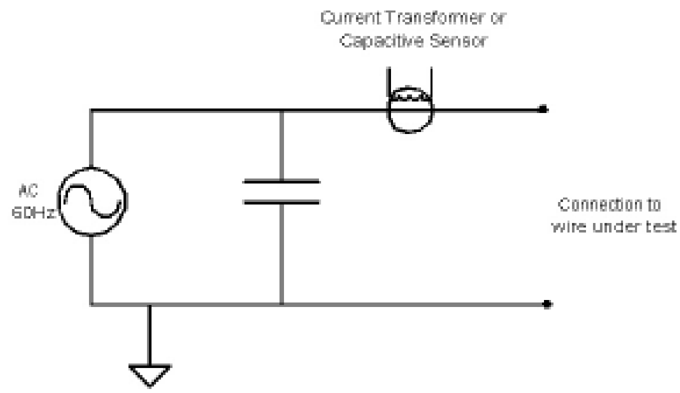

Figure 3: Setup for PD measurements

Unlike most conventional PD measurement systems where only magnitudes and location of PD pulses are recorded, this setup records and stores all PD waveforms and AC waveforms as well with a sampling rate of 2.5GS/s, which allows for a more thorough/advanced analysis of the signals and for extracting more characteristics as well. PD pulses are continuously collected for at least one full AC cycle to obtain complete phase data that aids in defect recognition. For each individual PD pulse, 1500 sampling points are taken. Figure 4 shows a typical PD pulse acquired from the wiring samples. Temperature and pressure of the test chamber are also recorded.

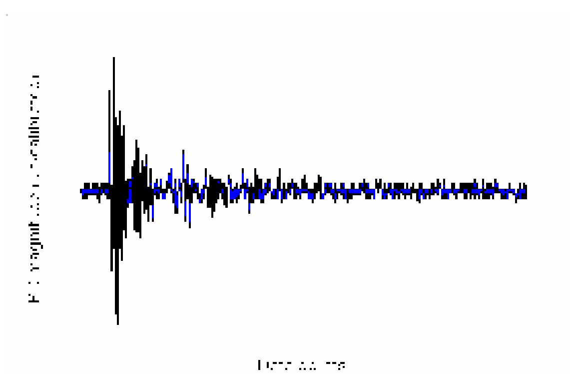

Figure 4: A typical PD pulse

A total of nine wire samples are tested with maximum of 10 repetitions for each sample. After data cleansing (including the removal of noise and incomplete data sets), 596 PD sequences are used for designing the PD diagnostic system. Out of the 596 PD sequences, 225 are for normal wires and 371 for chafed ones.

3.2. Feature extraction

For each of the 596 PD sequences obtained, features are first extracted. In this study, a total of 88 features are extracted using five feature extraction methods as briefly described below.

A) Features from statistical analysis of phase-resolved PD patterns. Phase-resolved PD pattern analysis is the most commonly used method for feature extraction in PD diagnosis. Given a sequence of PD pulses and the recorded voltage phase angles at corresponding pulse peaks, a 3D PD pattern is generated, where the number of pulses (pulse count) is plotted as a function of magnitude and phase of the PD pulses. A typical 3D PD pattern is shown in Figure 5. 3D PD patterns are a good representation/summary of all PD pulses recorded within a specified time window and should show different characteristics for different PD activities, thus different PD sources. For the convenience of statistical analysis, the 3D patterns are decomposed into two 2D distributions by projecting it into the two axes - phase and magnitude. Statistical analysis is performed separately for those two distributions. Also, statistical analysis is performed separately for phase angles from 0° to 180° ("positive" PDs), for phase angles from 180° to 360° ("negative" PDs), and on the difference between positive and negative PDs. For each of the distributions, two types of statistics, names amplitude statistics and shape statistics, are calculated. The statistical descriptors are mean, standard deviation, skewness and kurtosis. In addition, overall maximum magnitudes of positive and negative PDs and correlation between positive and negative PD patterns are also calculated as features.

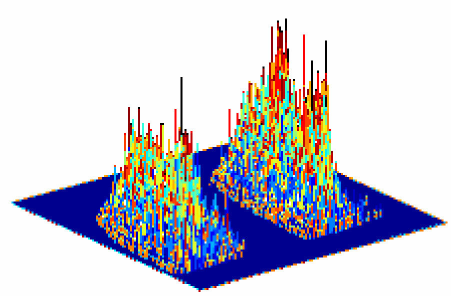

Figure 5: An example of 3D PD patterns

B) Features from PD height distribution analysis. Heights (peaks) of a sequence of PD pulses can be represented in a histogram that shows number of pulses as a function of their magnitude. According to Cacciari et al (2002), PD pulse height distribution tends to fit well with the two-parameter Weibull function defined as: $F(q) = 1 - e^{-(\frac{q}{\alpha})^{\beta}}$, where q is the pulse height, α and β are the scale and shape parameters of the Weibull function. They have found that the scale and shape (especially shape) parameters differ with different PD sources, thus can be used as features for PD identification or classification.

C) Features from "classification map". PD pulses are different in wave shape depending on the location and nature of the underlying defect that generates PD. One way to capture the different wave shapes is to use so-called "equivalent time-length", T^2, and "equivalent bandwidth", W^2 [Contin et al (2000)].

In the $T^2 - W^2$ plane (also called the "classification map" by [Contin et al (2000)]), each PD pulse is presented as a point and each sequence of PD pulses, which are similar in shape, fall into a well-defined area (cluster) in the $T^2 - W^2$ plane. The location and shape of the clusters in the $T^2 - W^2$ plane differ corresponding to different PD sources [Contin et al (2000 & 2002)]. In this paper, characteristics of the clusters are extracted through statistical analysis and used as features for classification purpose. The 15 features extracted include overall mean, means and standard deviations in both T^2 and W^2 directions, respectively, 1st through 4th orders of moments of distributions in both T^2 and W^2 directions, respectively, direction of the 1st eigenvector of the cluster, and ratio of the first two eigenvalues of the cluster.

D) Features from spectrum analysis. Frequency spectrum of a PD pulse indicates frequency components of the PD pulse. Thus the shape or distribution of frequency spectra should be correlated with different PD sources. In this paper, the first 3 frequencies corresponding to the highest three magnitudes, and the three highest magnitudes themselves, the difference between the three frequencies, and the difference between the three magnitudes are used as features.

E) Features from raw PD signals. They are the maximum and minimum peaks of PD pulses, mean and standard deviation of peaks of PD pulses, inception voltage, and PD rate.

The total number of features extracted is 88 including temperature and pressure.

4. Results and discussion

Prior to performing feature selection, correlation analysis of the 88 extracted features is conducted. Figure 6 displays the heat map representation of correlation coefficients, where each entry represents correlation coefficients between a pair of features. The higher the entry, the higher the correlation between the feature pair. It can be seen that not only do features from one extraction method correlate highly with those from other feature extraction methods, but also there exists correlation among features within one feature extraction method.

Figure 7 shows correlation coefficients between each individual feature and the class labels. Features with low feature-class correlation have low predicting power, thus they are possibly the irrelevant features.

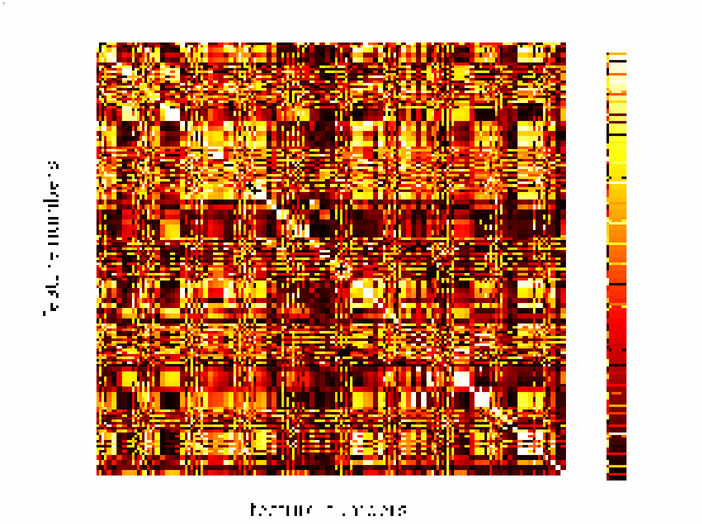

Figure 6: Correlation coefficients heat map

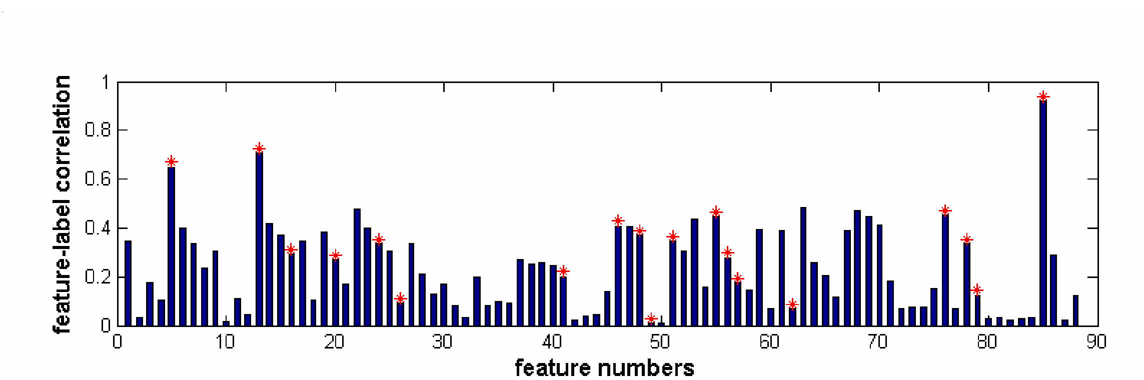

Figure 7: Feature-class correlation coefficients

Given the 88 features extracted from the previous section, we now perform feature selection using GA-based wrapper feature selection as describebed in Section 2. The Genetic Algorithm Optimization Toolbox [Houck et al (1995)] is used as GA engine. The fitness function used in GA is the evaluation accuracy of SVM classifier. That is, at each search step, the data set is randomly split into two disjoint subsets: one set is for training classifier and another for evaluation. To improve the robustness of evaluation, the SVM classifier is trained and evaluated 10 times and each time the data is randomly split into two disjoint subsets. The fitness function of GA is the average of the classifier accuracies of the 10 evaluations.

The population size is 50 and the number of generations is 10. Other GA parameters are: normalized geometric selection with the probability of selecting the best being 8%; simple single-point crossover; and binary mutation with probability of 5%.

Out of the 88 original features, GA-based wrapper feature selection chooses 19 features as the optimal feature set. The 19 selected features are marked in red stars in figure 7. As can be seen in Figure 7, the selected features are not necessarily exclusively the ones with high feature-class correlation, which supports our argument that optimal feature set cannot be obtained by individual feature evaluation alone. It is the combining effect that makes difference in classification.

To validate the superior performance of the selected features in classification, a separate SVM classifier with the 19 selected features as the input is trained and evaluated. Once again the training and evaluation of the classifier are performed 10 times, each of which use different training and evaluation data that is the result of randomly splitting the full data set. Realizing that only limited PD data is obtained in well-defined lab environment and that PD measurements in real aircraft wires will be much more noisy, 1-sigma random noise is added to the feature values. The mean, the standard deviation, the minimum, and the maximum of accuracies of the 10 evaluations are summarized in Table 1 (last row). For comparison purpose, the four accuracies of the classifier trained and evaluated based on subsets of features from individual feature extraction methods and using all 88 features are also shown in Table I, where “f-A” through “f-E” stand for subsets of features from the five feature extraction methods described in Section 3.2 and “f-All” for all extracted features combined. The classification accuracies for these different feature subsets are also summarized in error bar plot in Figure 8, where highest and lowest ends of the error bars represent the maximum and minimum accuracies, respectively, of corresponding feature subsets; and the line-connected circles represent the mean accuracy of each feature subset.

Table I: Comparison of classification accuracies

	Classification accuracy			
	mean	**std. deviation**	**minimum**	**maximum**
f-A	0.9440	0.0089	0.9245	0.9560
f-B	0.6342	0.0268	0.5765	0.6625
f-C	0.7836	0.0134	0.7589	0.8050
f-D	0.6857	0.0187	0.6604	0.7128
f-E	0.9268	0.0104	0.9119	0.9455
f-All	0.9566	0.0206	0.9224	0.9748
GAFS	0.9723	0.0055	0.9644	0.9811

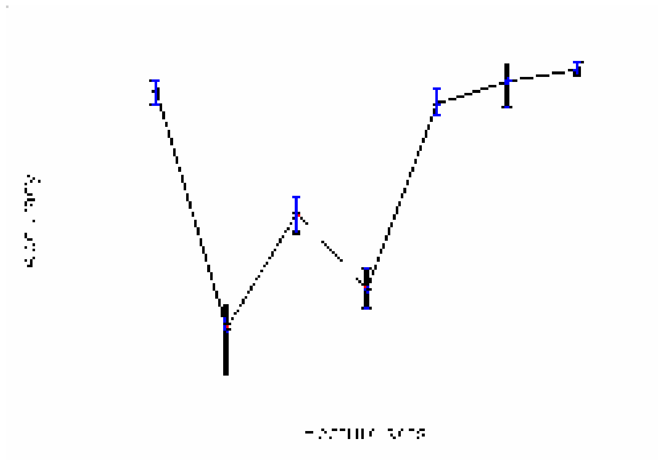

Figure 8: Comparison of classification accuracies

From Table I and Figure 8, we can see that the feature subset (of 19 features) selected by GA-based feature selection not only has the highest mean accuracy (0.9723), but also the smallest variation for the 10 runs. These results prove that feature selection improves both accuracy and robustness of classification in PD diagnosis.

Using all features gives a mean accuracy of 0.9566, which is higher than those from other designs, but lower than the GA-based feature selection design. The variation over the 10 runs for this design is also much higher than that for GA-based feature selection.

Out of the five feature extraction methods, statistical analysis of phase-resolved PD patterns seems to be relatively more effective for the problem under consideration in this paper since 11 of the 19 selected features are from this extraction method. On the other hand, the two features from PD height distribution analysis make insignificant contributions to PD classification.

5. Conclusions

In designing PD diagnostic systems, finding an optimal feature set is important for achieving the desired performance. This paper identifies/recognizes the challenge to the designer in finding optimal features even though a large number of candidate features can be readily extracted by using existing feature extraction methods. This paper further introduces feature selection into PD diagnosis, which enables the designer to systematically find optimal features for achieving higher classification performance. By designing a PD diagnostic system for aircraft wire fault diagnosis where GA-based wrapper feature selection is used for determining optimal features, this paper demonstrates that feature selection can reliably and effectively find a reasonably small subset of features that give the highest performance. Additionally, feature selection also reduces the designer's time in finding optimal features, thus the time of overall design cycle of PD diagnostic systems since it is a systematic approach and can be performed in an automated fashion.

References:

1. Gulski, E., Burger, H.P., Zielonka, A., and Brooks, R. (1996), "Classification of defects in HV components by fractal analysis of PD measurements", IEEE Annual Report Conference on Electrical Insulation and Dielectric Phenomena, San Francisco, October 20-23, 1996
2. Gulski, E. (1995), "Digital analysis of partial discharges", IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 2, No. 5, pp xxx
3. Contin, A., Cavallini, A., Montanari, G.C., Pasini, G., and Puletti, F. (2000), "Artificial intelligence methodology for separation and classification of partial discharge signals", Proceedings of IEEE Conference on Electrical Insulation and Dielectric Phenomena", Victoria, Canada, pp522-26.
4. Contin, A., Cavallini, A., Montanari, G.C., Pasini, G., and Puletti, F. (2002), "Digital detection and fuzzy classification of partial discharge signals", IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 9, No. 3, pp335-48
5. Abdul Rahman, M.K., Arora, R., and Srivastava, S.C. (2000), "Partial discharge classification using principal component transformation", IEE Proceedings of Scientific Measurement Technologies, Vol. 147, No. 1, pp7-13
6. Siedlecki, W. and Sklansky, J. (1989), "A note on genetic algorithms for large-scale feature selection", Pattern Recognition letter, Vol. 10, pp335-347, 1989
7. Vapnik, V. (1995), *The nature of statistical learning theory*, Springer-Verlag, New York, 1995
8. Bennett, K. P. and Campbell, C. (2000), "Support vector machines: Hype or Hallelujah?", Newsletter of the ACM Special Interest Group on Knowledge Discovery and Data Mining (ACM SIGKDD), Vol. 2, Issue 2.
9. Dash, M. and Liu, H. (1997), "Feature Selection for Classification", Intelligent Data Analysis, Vol.1, No.3
10. Langley, P. (1994), "Selection of Relevant Features in Machine Learning", Proceedings of 1994 AAAI Fall Symposium, pp127-131
11. Kohavi, R. and George, H.J. (1997), "Wrappers for feature subset selection", Artificial Intelligence, Vol. 97, No. 1-2, December 1997, pp273-324.
12. Liu, H. and Motoda, H. (1998), *Feature Selection for Knowledge Discovery and Data Mining*, Kluwer Academic Publishers, Norwell, MA
13. Narendra, P. and Fukunaga, K. (1977), "A Branch and Bound Algorithm for Feature Subset Selection", IEEE Transactions on Computer, Vol.C-26, No.9, pp917-922
14. Yang, J. and Honavar, V. (1998), "Feature Subset Selection Using a Genetic Algorithm", IEEE Intelligent Systems, Vol. 13, No.2, pp44-49

15. Raymer, M.L., Punch, W.F., Goodman, E.D., Kuhn, L.A., and Jain, A.K. (2000), Dimensionality reduction using genetic algorithms", IEEE Transactions on Evolutionary Computation, Vol. 4, No. 2, pp164-171

16. Doak, J. (1992), "An Evaluation of Feature Selection Methods and their Application to Computer Security", Technical report, Davis, CA: University of California, Department of Computer Science

17. Kitter, J. (1978), "Feature Set Search Algorithm", Pattern Recognition and Signal Processing, Sithoff and Noordhoff, Alphen ann Den Riju, The Netherlands, pp41-60

18. Pudil, P., Novovicova, J. and Kittler, J. (1994), " Floating Search Methods in Feature Selection", Pattern Recognition Letters 15 (1), pp1119-1125

19. Setiono, R. and Liu, H. (1997), "Neural-Network Feature Selector", IEEE Transactions on Neural Networks, Vol. 8, No. 3, pp 654-662

20. Rezaee, M. R., Goedhart, B., and Lelieveldt, B. (1999), "Fuzzy Feature Selection", Pattern Recognition, Vol. 32, No.12, pp2011-2019

21. Duda, R. O., Hart, P. E., and Stork, D. G. (2000), *Pattern Classification*, John Wiley & Sons, Inc., New York, NY

22. Piramuthu, S. (1999), "The Hausdorff Distance Measure for Feature Selection in Learning Applications", Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences

23. Ben-Bassat, M. (1982), "Pattern Recognition and Reduction of dimensionality", Handbook of Statistics (P. Krishnaiah and L. Kanal, eds.) North Holland, pp773-791

24. Hall, M.A. (2000), "Correlation-based feature selection for discrete and numeric class machine learning", Proceedings of the 17th International Conference on Machine Learning, Stanford University, CA. Morgan Kaufmann Publishers

25. John, G., Kohavi, R. and Pfleger, K. (1994), "Irrelevant Features and the Subset Selection Problems", Proceedings of 11th Int'l Conference on Machine Learning, San Mateo, CA, pp121-129

26. Wang, H., Bell, D. and Murtagh, F. (1999), "Axiomatic Approach to Feature Subset Selection Based on Relevance", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, No.3, pp271-277

27. Houck, C.R., Joines, J.A., and Kay, M.G. (1995), "A genetic algorithm for function optimization: a Matlab implementation", Technical Report NCSU-IE TR95-09, North Carolina State University, Raleigh, NC