Evolutionary Computation-based Kernel Optimal Component Analysis for Pattern Recognition

Jason C. Isaacs
isaacs@eng.fsu.edu

Simon Foo
Florida State University
College of Engineering
Tallahassee, FL
foo@eng.fsu.edu

Anke Meyer-Baese
amb@eng.fsu.edu

ABSTRACT

Kernel methods are mathematical tools that provide higher dimensional representation of given data set in feature space for pattern recognition and data analysis problems. Optimal Component Analysis (OCA) [4] poses the problem of finding an optimal linear representation. In this paper we present the results of six kernel functions and their respective performance for Evolutionary Computation-based kernel OCA on the Pima Indian Diabetes database. Empirical results show that we outperform existing techniques on this database.

Categories and Subject Descriptors

I.5 [**Pattern Recognition**]: *Feature evaluation and selection.*

General Terms

Algorithms

Keywords

Kernels, Evolutionary Computation, Component Analysis.

1. INTRODUCTION

The role of a pattern recognition system is to classify data (and the patterns within) based on either a priori knowledge or information extracted from the data. The patterns to be classified are usually groups of measurements or observations defining points in an appropriate multidimensional space.

2. OPTIMAL KERNEL PROJECTION ANALYSIS

2.1 Introduction to Kernel Methods

Kernel methods, first published in the 1964 [1], have only come into popularity within the last ten years [4]. Essentially, the 'kernel trick' is a method for easily converting a linear classification learning algorithm into non-linear one, by mapping the original observations into a higher-dimensional non-linear space. We used the Kernel functions listed in Table 1 from [2].

2.2 Selection, Reproduction and Fitness

The problem is to optimize a subspace projection matrix to improve the recognition rate over a benchmarking database. We utilized a rank based selection on a population of twenty chromosomes with the top five reproducing with bottom fifteen. In addition, we used immigration with randomly configured set of orthogonal matrices [U] with discriminant(U) > 0. For reproduction we employed a midpoint crossover and the mutation rate used was 0.02 and had not been optimized. Testing was done using the Pima Indian database. Fitness was measured using a nearest neighbor classifier for the sake of simplicity.

GECCO'07, July 7–11, 2007, London, England, United Kingdom
ACM 978-1-59593-697-4/07/0007.

2.3 Experimentation and Results

There are 768 total samples. 581 are used for training, and the remaining 187 are used for testing. Experimental results are shown in Table 1. Figure 1 shows a convergence plot.

Table 1. GA-OPA Experimental Results For The Pima Indian Diabetes Data Set (Percent Correct).

Kernel	Starting	Final
Rayleigh	79.68	85.56
Erlang	79.68	86.1
Bessel	78.61	85.56
Poly-3	78.07	83.96
Einstein1	80.21	87.17
Einstein2	79.14	86.1

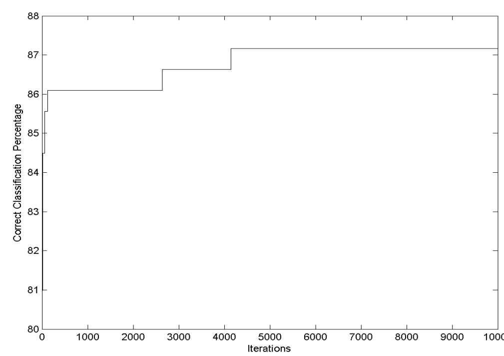

Figure 1. Convergence Plot for Einstine-1 Kernel.

3. CONCLUSION

We have demonstrated the novelty and effectiveness of this optimization technique as applied to OKCA for the suggested kernels. We also demonstrated a classification improvement of over ten percent on this database.

4. REFERENCES

[1] M. Aizerman, E. Braverman, and L. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition learning". Automation and Remote Control 25: 821-837 (1964).

[2] J. Isaacs, S. Foo, and A. Meyer-Baese, "Novel Kernels and Kernel PCA for Pattern Recognition", Proceedings of the 7th IEEE International Symposium on Computational Intelligence in Robotics and Automation.

[3] B. Schölkopf and A. J. Smola: Learning with Kernels. MIT Press, Cambridge, MA, 2002.

[4] Q. Zhang and X. Liu, "Kernel Optimal Component Analysis," proceedings of the 2004 IEEE Computer Vision and Pattern Recognition Workshops.