Implementation of exemplar-based learning model for music cognition

Ichiro Fujinaga, Stephan Moore, David S. Sullivan, Jr.

Peabody Institute, Johns Hopkins University
1 E. Mt. Vernon Place
Baltimore, Maryland 21202 U. S. A.
{ich | stephan | sullivan}@peabody.jhu.edu
http://www.peabody.jhu.edu/~ich

Abstract

"We tend to think of what we 'really' know as what we can talk about, and disparage knowledge that we can' t verbalize." ((Dowling 1989), 252)

The exemplar-based learning model is proposed here as an alternative approach to modeling many aspects of music cognition. The implementation of this model is based on a k-NN (nearest neighbor) classifier, and on a genetic algorithm which is used for feature weighting.

Although humans are capable of consciously abstracting concepts and deriving rules, there are other cognitive tasks such as music knowledge acquisition that are largely non-verbal and defy generalizations, consequently making the application of traditional rule-based AI models problematic.

In exemplar-based learning, such as the k-NN rule, objects are categorized by their similarity to one or more sets of stored examples, which can be represented as weighted feature vectors. Similarity is often defined as the distance between the vectors. In the current implementation, the genetic algorithm is used to find a near-optimal set of weights.

This paradigm, also known as the lazy learning model, is attractive because training is not necessary, learning is extremely fast, algorithms are simple and intuitive, rules are not sought, and learning is incremental. The major drawback has been the high memory requirement, since all examples must be stored, but the recent decrease in memory cost makes this model quite feasible.

Exemplar-based recognition models have been successfully applied in many pattern recognition and classification tasks. Furthermore, cognitive psychologists have found this model evident in human and animal learning. In music, style recognition, harmonization, expressive performance, instrument recognition, and structural analysis are some of the obvious targets for the deployment of this model. Use of this model is illustrated with an optical music recognition system and a musical instrument identifier that uses only the steady-state (post-attack) portion of instrumental sounds.

Introduction

Most research in artificial intelligence and music has used rule-based models (e.g. (Balaban, Ebcioglu, and Laske 1992; Schwanauer and Levitt 1993)). Exemplar-based model, which is analogous to the idea of "learning by examples," is proposed here as an alternative approach to modeling many aspects of music cognition.

Although humans are capable of consciously abstracting concepts and deriving rules, there are other cognitive tasks such as music knowledge acquisition that are largely non-verbal and defy generalizations, consequently making the application of traditional rule-based AI models problematic. Laske (Laske 1992, 251) remarked that "in AI generally, and in AI and Music in particular, the acquisition of non-verbal knowledge is difficult, and no proven methodology exists." One of the ways to represent non-verbal, implicit knowledge is through examples. The implementation of this model is based on a combination of a k-nearest neighbor (k-NN) classifier and a genetic algorithm, which is used for feature selection and feature weighting.

This paradigm, also known as the lazy learning model (Aha 1997), is attractive because training is not necessary, learning is extremely fast, algorithms are simple and intuitive, rules are not sought, and learning is incremental. The major drawback has been the high memory requirement since all examples must be stored, but the recent decrease in memory cost makes this model quite feasible.

Exemplar-based recognition models have been successfully applied, for example, in optical music recognition (Fujinaga, Pennycook, and Alphonce

1989), vehicle identification (Lu, Hsu, and Maldague 1992), pronunciation (Cost and Salzberg 1993), cloud identification (Aha and Bankert 1994), respiratory sounds classification (Sankur et al. 1994), wine analysis and classification (Latorre et al. 1994; Moret, Scarponi, and Cescon 1994; Garciajares, Garciamartin, and Celatorrijos 1995), natural language translation (Sato 1995), and credit risk assessment (Henley and Hand 1996). Furthermore, cognitive psychologists have found this model evident in human and animal learning. In music, style recognition, harmonization, expressive performance, instrument recognition, and structural analysis are some of the obvious targets for the deployment of this model.

Exemplar-based model

The exemplar-based model is based on the idea that objects are categorized by their similarity to one or more stored examples. There is much evidence from psychological studies to support exemplar-based categorization by humans (Brooks 1978; Hintzman 1986; Medin and Schaffer 1978). This model differs both from rule-based or prototype-based models of concept formation in that it assumes no abstraction or generalizations of concepts (Nosofsky 1984; 1986).

K-nearest-neighbor classifier

The exemplar-based model can be implemented by k-NN classifier (Cover and Hart 1967), which is a classification scheme to determine the class of a given sample by its feature vector. Distances between feature vectors of an unclassified sample and previously classified samples are calculated. The distance can be defined in a variety of ways, for example, the Euclidean distance between two feature vectors **X** and **Y** in an N-dimensional feature space is defined as:

$$d = \left[\sum_{i=1}^{N} \left(x_i - y_i \right)^2 \right]^{1/2}.$$

The class represented by the majority of k-closest neighbors is then assigned to the unclassified sample. Besides its simplicity and intuitive appeal, the classifier can be easily modified, by continually adding new samples that it "encounters" into the database, to become an incremental learning system. In fact, "the nearest neighbor algorithm is one of the simplest learning methods known, and yet no other algorithm has been shown to outperform it consistently" ((Cost and Salzberg 1993), 76).

Since the calculation time of the distance is proportional to the number of features in the feature vector, one of the ways to improve both performance and accuracy is to determine relevant features.

Feature selection

Feature selection involves deciding which subset of features that best distinguish among the various object types. The procedure of selecting "good" features is not formalized; as Castleman states, "frequently intuition guides the listing of potentially useful features" (Castleman 1979), 321). Cover and Van Campenhout (Cover and Van Campenhout 1977) rigorously showed that in determining the best feature subset of size *m* out of *n* features, one needs to examine all possible subsets of size *m*. For practical consideration, some non-exhaustive feature selection methods must be employed. Many methods exist for finding near-optimal solutions to this problem in a finite time (Jain and Zongker 1997), such as sequential backward elimination (Kittler 1978), sequential forward floating selection (Pudil, Novovicova, and Kittler 1994), and branch and bound (Hamamoto et al. 1990; Narendra and Fukunaga 1977; Yu and Yuan 1993). The latter method guarantees the optimal features subset without explicitly evaluating all possible feature subsets under the assumption that the criterion function used satisfies the "monotonicity" property. Unfortunately, in many situations there is no guarantee that this constraint or even the more relaxed "approximate monotonicity" (Foroutan and Sklansky 1987) can be met. Furthermore, although branch and bound can reduce the search space drastically, the calculation may become impractical in cases where there are many features.

Feature selection using genetic algorithms, introduced by (Siedlecki 1989), is near optimal and efficient. It makes no a priori assumptions about the features and the subset can be incrementally modified as more samples are added to the database, thus making it adaptive to changing environment. This technique of finding a good feature subset is used by various researchers in neural networks (Brill, Brown, and Martin 1992), parallel classification (Kuncheva 1993), biology (Punch et al. 1993), music (Fujinaga 1995), mammography (Sahiner et al. 1996), and agriculture (Chtioui, Bertrand, and Barba 1998).

Genetic algorithms

Genetic algorithms (GA) (Davis 1987; Goldberg 1989; Holland 1975) are often used whenever exhaustive search of the solution space is impossible or prohibitive, and is based on computational models of the evolution of individual structures via processes of selection and reproduction.

The algorithm maintains a population of individuals that evolve according to specific rules of selection and other operators such as crossover and mutation. Each individual in the population receives a measure of its fitness in the environment. Selection focuses attention on high-fitness individuals, thus exploiting the available fitness information. Since the individual’s genetic information (chromosomes) is represented as arrays of binary data, simple bit manipulations allow the implementation of mutation and crossover operations.

GA have been successfully applied to solve many optimization and other computationally intensive problems (Davis 1991). In music, genetic algorithms have been used for sound synthesis (Cheung and Horner 1996; Horner, Beauchamp, and Haken 1993; Horner, Beauchamp, and Haken 1992; Horner, Beauchamp, and Packard 1993; Takala et al. 1993; Vuori and Välimäki 1993), for optimal placement of microphones (Wang 1996), and as a compositional aid (Fujinaga and Vantomme 1994; Horner and Goldberg 1991).

For the feature selection, the set of features is converted to “genes” (chromosomes), where each feature is represented by a bit in the binary array. Therefore, each gene, having different sequence of bits represents a subset of features to be used for classification and those that have high recognition rates are made to survive in this pseudo-biological environment.

By using GA from the beginning of the learning process, a set of good genes (set of weights) are saved so that they can be used as the starting points for the future selection processes.

It should be noted that the genetic algorithm can also be used to make k-NN classifier more efficient by reducing the number of exemplars in the database (Brill, Brown, and Martin 1992; Kuncheva 1997; Kuncheva 1995; Zhao and Higuchi 1996).

Timbre recognition

A timbre recognition experiment was conducted using k-NN classifier and using GA for the feature selection. The system was used to classify 39 different orchestral instrument timbres. The data comprised of the steady-state (post-attack) spectrum of each of the instruments played at different pitches (total of 1338 spectrums) (Sandell 1994). Some timbre contained over 400 features since each value of the harmonics was considered as a feature. Additional features calculated from the spectral data included centroid and other higher order moments, such as skewness and kurtosis.

The initial analysis showed that, as expected, the centroid alone was the best single feature with the recognition rate of 13%, which is much better than chance (2.5%). When the genetic algorithm was used for feature selection, i.e. finding the subset of 400 features that results in the best recognition score, the best results were obtained using five features: the fundamental, the integral of the spectrum, the centroid, the standard deviation, and the skewness. What was surprising was the recognition varied greatly between instruments. While the French horn and the muted trumpet were recognized near 100%, other instruments did very poorly: such as the oboe (19%), the viola with martele (14%), and the violin with martele (6%). The average overall was 46%, which is far superior to using a single feature such as the centroid while maintaining calculation time to a minimum by using five features, instead of hundreds of features.

This study illustrates the benefit of feature selection using genetic algorithms. The exhaustive search would have meant prohibitive calculation of 2^{400} combinations of features.

Feature weights

The k-NN classifiers can be further enhanced by modifying the feature space, or equivalently, changing the weights in the distance measure (Kelly and Davis 1991). A commonly used weighted-Euclidean metric between two vectors **X** and **Y** in an N-dimensional feature space is defined as:

$$d = \left(\sum_{i=1}^{N} \omega_i \left(x_i - y_i \right)^2 \right)^{1/2}$$

By changing the weights, ω_i , the shape of the feature space can be changed. (See Figure 1.) The feature selection is a trivial case of feature weighting where ω_i is binary.

In determining the weights for each feature, the problem becomes more complex as Cash and Hatamian (Cash and Hatamian 1987) have shown. On the other hand the weighting of each feature used in a similarity measure can markedly improve the recognition rate and also provide relative importance of each feature. In other words, the optimal use of features involves not only choosing the correct subset of the features but also how much of each feature should contribute to the final decision. In feature selection, the goal was to find a set of binary weights for the features (0 or 1), but the problem now is to determine the weights that can be any real numbers. Since no known deterministic method for finding the optimal solution exits, GA is again a useful tool for finding the near-optimal set of weights from this infinite possibility (Wettschereck, Aha, and Mohri 1997).

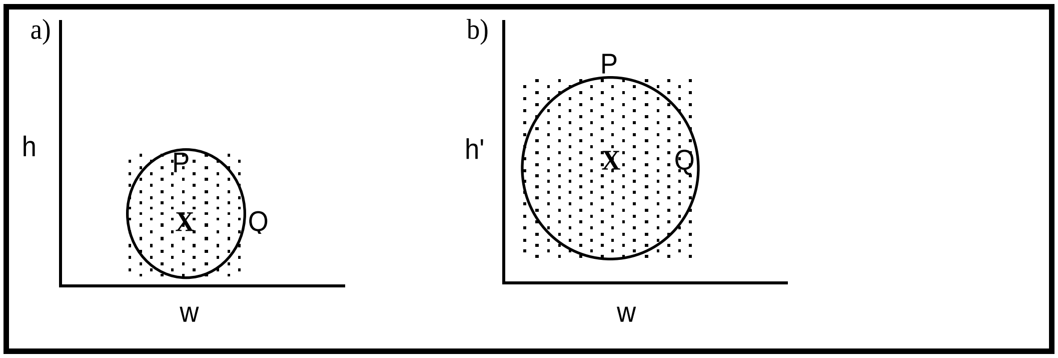

Figure 1. Unknown object X and its neighbors in two feature spaces. By changing the weights in the distance metric, the shape of the feature can be modified. In a), the nearest neighbor of X is P. In b), where the vertical axis has been scaled, the nearest neighbor is Q,

Although, this hybrid learning system, combining k-NN classifier and GA, does not guarantee the optimal solution, near-optimal results can be obtained relatively quickly and preliminary experiments with the system have shown dramatic improvements in the recognition.

Raymer et al. (Raymer et al. 1997a) comparing a simple k-NN and a feature-weighted k-NN showed recognition accuracy increase from 61% to 77% in one case and 69% to 80% in another case, both using four floating-point features. In a medical application using 21 features, they (Raymer et al. 1997b) showed an improvement from 70% to 94%. In addition, the GA was modified (using a 987-bit chromosome) for feature selection as well as feature weighting that resulted in reduction of the number of features to five and increase of accuracy to 98%.

Other successful implementations include speaker identification (Charlet and Jouvet 1997) and optical music recognition.

Optical music recognition

In the optical music recognition task (Fujinaga 1996b; 1996a), segmented music symbols are represented by 15 features, such as width, height, area. Prior to introducing the GA into the classifier, the recognition rates were in the 85% to 90% range depending on the complexity of music. The feature selection up to this point was done manually over several months. After a few days of calculation, the application of GA resulted in a significant improvement achieving accuracy of 95%-97%. By using floating-point weights with 4 bits per feature, the system recognized up to 99% of the symbols

Conclusions

The implementation of the exemplar-based learning model using k-NN classifiers and genetic algorithms was demonstrated using examples from music applications. There are many other possible areas in music where this model can be applied. Some of these are harmonization, counterpoint, orchestration, piano reduction, expressive performance, automatic accompaniment, composition / improvisation, score-based analysis, transcription, beat-induction, tempo / key tracking, key finder, phrase detection, style imitation, style identification, intelligent instrument lessons.

The exemplar-based model, which places rote memory as the fundamental mechanism for learning, offers a promising and alternative approach for music cognition and to understand the implicit knowledge that musicians possess and may be applied to other types of categorization and learning tasks.

Bibliography

Aha, D. W. 1997. Lazy learning. *Artificial Intelligence Review* 11 (1-5): 7-10.

Aha, D. W., and R. L. Bankert. 1994. Feature selection for case-based classification of cloud types: An empirical comparison. *Case-Based Reasoning: Papers from the 1994 Workshop*.

Balaban, M., K. Ebcioglu, and O. Laske, eds. 1992. *Understanding music with AI: Perspectives on music cognition*. Menlo Park: AAAI Press.

Brill, F. Z., D. E. Brown, and W. N. Martin. 1992. Fast genetic selection of features for neural network classifiers. *IEEE Transactions On Neural Networks* 3 (2): 324-8.

Brooks, L. R. 1978. Non-analytic concept formation and memory for instances. In *Cognition and categorization*, edited by E. Rosch and B. Lloyd. Hillside, N. J: Erlbaum.

Cash, G. L., and M. Hatamian. 1987. Optical character recognition by the method of moments. *Computer Vision, Graphics, and Image Processing* 39 (3): 291-310.

Castleman, K. R. 1979. *Digital image processing*. Englewood Cliffs: Prentice-Hall.

Charlet, D., and D. Jouvet. 1997. Optimizing feature set for speaker verification. In *Audio- and Video-Based Biometric Person Authentication*. Berlin: Springer-Verlag Berlin.

Cheung, N. M., and A. B. Horner. 1996. Group synthesis with genetic algorithms. *Journal of the Audio Engineering Society* 44 (3): 130-47.

Chtioui, Y., D. Bertrand, and D. Barba. 1998. Feature selection by a genetic algorithm: Application to seed discrimination by artificial vision. *Journal of the Science of Food and Agriculture* 76 (1): 77-86.

Cost, S., and S. Salzberg. 1993. A weighted nearest neighbor algorithm for learning with symbolic features. *Machine Learning* 10: 57-78.

Cover, T., and P. Hart. 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* 13 (1): 21-7.

Cover, T. M., and J. M. Van Campenhout. 1977. On the possible orderings in the measurement selection problem. *Transactions on Systems, Man, and Cybernetics* 7 (9): 657-61.

Davis, L. 1991. *Handbook of genetic algorithms*. New York: Van Norstrand Reinhold.

Davis, L. 1987. *Genetic algorithms and simulated annealing*. London: Pitman.

Dowling, W. J. 1989. Simplicity and complexity in music and cognition. *Contemporary Music Review* 4: 247-53.

Foroutan, I., and J. Sklansky. 1987. Feature selection for automatic classification of non-Gaussian data. *IEEE Transactions on Systems, Man, and Cybernetics* 17 (2): 187-98.

Fujinaga, I. 1996b. Adaptive optical music recognition. Ph.D. Dissertation, Music, McGIll University.

Fujinaga, I. 1996a. Exemplar-based learning in adaptive optical music recognition system. *Proceedings of the International Computer Music Conference*. 55-6.

Fujinaga, I. 1995. Exemplar-based music structure recognition. *Workshop Notes for IJCAI-95 Workshop on Artificial Intelligence and Music*. 75-6.

Fujinaga, I., B. Pennycook, and B. Alphonce. 1989. Computer recognition of musical notation. *Proceedings to the First International Conference on Music Perception and Cognition*. 87–90.

Fujinaga, I., and J. Vantomme. 1994. Genetic algorithms as a method for granular synthesis regulation. *Proceedings of the International Computer Music Conference.* 138-41.

Garciajares, C., S. Garciamartin, and R. Celatorrijos. 1995. Analysis of some highly volatile compounds of wine by means of purge and cold trapping injector capillary gas-chromatography: *Journal of Agricultural and Food Chemistry* 43 (3): 764-8.

Goldberg, D. E. 1989. *Genetic algorithms in search, optimization, and machine leaning*. Reading: Addison-Wesley.

Hamamoto, Y., S. Uchimura, Y. Matsuura, T. Kanaoka, and S. Tomita. 1990. Evaluation of the branch and bound algorithm for feature selection. *Pattern Recognition Letters.* 11: 453-6.

Henley, W. E., and D. J. Hand. 1996. A k-nearest-neighbour classifier for assessing consumer credit risk. *Statistician* 45 (1): 77-95.

Hintzman, D. L. 1986. 'Schema abstraction' in a multiple-trace memory model. *Psychological Review* 93: 411–28.

Holland, J. H. 1975. *Adaptation in natural and artificial systems*. Ann Arbor: U. of Michigan Press.

Horner, A., J. Beauchamp, and L. Haken. 1993. Methods for multiple wavetable synthesis of musical-instrument tones. *Journal of the Audio Engineering Society* 41 (5): 336-56.

Horner, A., J. Beauchamp, and L. Haken. 1992. Wavetable and FM matching synthesis of musical instrument tones. *Proceedings of the International Computer Music Conference*. 18-21.

Horner, A., J. Beauchamp, and N. Packard. 1993. Timbre breeding. *Proceedings of the International Computer Music Conference*. 396-8.

Horner, A., and D. E. Goldberg. 1991. Genetic algorithms and computer-assisted music composition. *Proceedings of the International Computer Music Conference*. 479-82.

Jain, A., and D. Zongker. 1997. Feature selection: Evaluation, application, and small sample performance. *IEEE Transactions On Pattern Analysis and Machine Intelligence* 19 (2): 153-58.

Kelly, J. D., and L. Davis. 1991. Hybridizing the genetic algorithm and the k nearest neighbors classification algorithm. *Fourth International Conference on Genetic Algorithms and their Applications*. 377-83.

Kittler, J. 1978. Feature set search algorithms. In *Pattern Recognition and Signal Processing*, edited by C. H. Chen. The Netherlands: Sijthoff and Norddhoff.

Kuncheva, L. I. 1997. Fitness functions in editing k-NN reference set by genetic algorithms. *Pattern Recognition* 30 (6): 1041-9.

Kuncheva, L. I. 1995. Editing for the k-nearest neighbors rule by a genetic algorithm. *Pattern Recognition Letters* 16 (8): 809-14.

Kuncheva, L. I. 1993. Genetic algorithm for feature selection for parallel classifiers. *Information Processing Letters* 46 (4): 163-8.

Laske, O. E. 1992. The OBSERVER tradition of knowledge acquisition. In *Understanding Music with AI*, edited by M. Balaban, K. Ebcioglu and O. E. Laske. Cambridge: The AAAI Press.

Latorre, M. J., C. Garciajares, B. Medina, and C. Herrero. 1994. Pattern-recognition analysis applied to classification of wines From Galicia (Northwestern Spain) with certified brand of origin. *Journal of Agricultural and Food Chemistry* 42 (7): 1451-5.

Lu, Y. J., Y. H. Hsu, and X. Maldague. 1992. Vehicle classification using infrared image-analysis. *Journal of Transportation Engineering-Asce* 118 (2): 223-40.

Medin, D. L., and M. M. Schaffer. 1978. Context theory of classification learning. *Psychological Review* 85: 207-38.

Moret, I., G. Scarponi, and P. Cescon. 1994. Chemometric characterization and classification of 5 Venetian white wines. *Journal of Agricultural and Food Chemistry* 42 (5): 1143-53.

Narendra, P. M., and K. Fukunaga. 1977. A branch and bound algorithm for feature subset selection. *IEEE Transactions of Computers* 26: 917-22.

Nosofsky, R. M. 1986. Attention, similarity, and the identification categorization relationship. *Journal of Experimental Psychology: General* 115 (1): 39-57.

Nosofsky, R. M. 1984. Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10 (1): 104-14.

Pudil, P., J. Novovicova, and J. Kittler. 1994. Floating search methods in feature selection. *Pattern Recognition Letters* 15 (11): 1119-25.

Punch, W. F., E. D. Goodman, M. Pei, L. Chia-Shun, P. Hovland, and R. Enbody. 1993. Further research on feature selection and classification using genetic algorithms. *Proceedings of the Fifth International Conference on Genetic Algorithms*. 557-64.

Raymer, M. L., W. F. Punch, E. D. Goodman, P. C. Sanschagrin, and L. A. Kuhn. 1997b. Simultaneous feature scaling and selection using a genetic algorithm. *Proceedings of the Seventh International Conference on Genetic Algorithms*. 561-7.

Raymer, M. L., P. C. Sanschagrin, W. F. Punch, S. Venkataraman, E. D. Goodman, and L. A. Kuhn. 1997a. Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *Journal of Molecular Biology* 265 (4): 445-64.

Sahiner, B., H. P. Chan, D. T. Wei, N. Petrick, M. A. Helvie, D. D. Adler, and M. M. Goodsitt. 1996. Image feature selection by a genetic algorithm: Application to classification of mass and normal breast tissue. *Medical Physics* 23 (10): 1671-84.

Sandell, G. J. 1994. SHARC timbre database.

Sankur, B., Y. P. Kahya, E. Cagatay Guler, and T. Engin. 1994. Comparison of AR-based algorithms for respiratory sounds classification. *Computers in Biology and Medicine* 24 (1): 67-76.

Sato, S. 1995. MBT2: A method for combining fragments of examples in example-based translation. *Artificial Intelligence* 75 (1): 33.

Schwanauer, S. M., and D. A. Levitt, eds. 1993. *Machine models of music*. Cambridge, Mass: MIT Press.

Siedlecki, W. S., J. 1989. A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters* 10 (5): 335-47.

Takala, T., J. Hahn, L. Gritz, J. Geigel, and J. W. Lee. 1993. Using physically-based models and genetic algorithms for functional composition of sound signals, synchronized to animated motion. *Proceedings of the International Computer Music Conference*. 180-5.

Vuori, J., and V. Välimäki. 1993. Parameter estimation of non-linear physical models by simulated evolution-application to the flute model. *Proceedings of the International Computer Music Conference*. 402-4.

Wang, B. T. 1996. Optimal placement of microphones and piezoelectric transducer actuators for far-field sound radiation control. *Journal of the Acoustical Society of America* 99 (5): 2975-84.

Wettschereck, D., D. W. Aha, and T. Mohri. 1997. A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artificial Intelligence Review* 11 (1-5): 272-314.

Yu, B., and B. Z. Yuan. 1993. A more efficient branch-and-bound algorithm for feature selection. *Pattern Recognition* 26 (6): 883-9.

Zhao, Q., and T. Higuchi. 1996. Minimization of nearest neighbor classifiers based on individual evolutionary algorithm. *Pattern Recognition Letters* 17 (2): 125-31.