Geometric Decision Rules for High Dimensions

Binay Bhattacharya [*]
School of Computing Science
Simon Fraser University

Kaustav Mukherjee
School of Computing Science
Simon Fraser University

Godfried Toussaint [†]
School of Computer Science
McGill University

Abstract

In this paper we report on a new approach to the instance-based learning problem. The new approach combines five tools: first, editing the data using Wilson-Gabriel-editing to smooth the decision boundary, second, applying Gabriel-thinning to the edited set, third, filtering this output with the ICF algorithm of Brighton and Mellish, fourth, using the Gabriel-neighbor decision rule to clasify new incoming queries, and fifth, using a new data structure that allows the efficient computation of *approximate* Gabriel graphs in high dimensional spaces. Extensive experiments suggest that our approach is the best on the market.

1 Introduction

In the typical non-parametric classification problem (see Devroye, Gyorfy and Lugosi [3]) we have available a set of d measurements or observations (also called a feature vector) taken from each member of a data set of n objects (patterns) denoted by $\{X, Y\} = \{(X_1, Y_1), ..., (X_n, Y_n)\}$, where X_i and Y_i denote, respectively, the feature vector on the ith object and the class label of that object. One of the most attractive decision procedures, conceived by Fix and Hodges in 1951, is the nearest-neighbor rule (1-*NN*-rule). Let Z be a new pattern (feature vector) to be classified and let X_j be the feature vector in $\{X, Y\} = \{(X_1, Y_1), ..., (X_n, Y_n)\}$ closest to Z. The nearest neighbor decision rule classifies the unknown pattern Z into class Y_j.

A key feature of this decision rule is that it performs remarkably well considering that no explicit knowledge of the underlying distributions of the data is used. Furthermore, a simple generalization of this rule called the *k*-*NN*-rule, in which a new pattern Z is classified into the class with the most members present among the k nearest neighbors of Z in $\{X, Y\}$, can be

[*] This research was supported by NSERC.

[†] This research was supported by NSERC and FCAR.

used to obtain good estimates of the Bayes error and its probability of error asymptotically approaches the Bayes error (Devroye et al. [3]).

In practice the size of the training set $\{X, Y\}$ is not infinite. This raises several practical questions. How can the storage of the training set be reduced without degrading the performance of the decision rule? How large should k be? How can the rule be made robust to overlapping classes or noise present in the training data? How can new decisions be made in a practical and computationally effcient manner? Geometric proximity graphs such as Voronoi diagrams and their many relatives provide elegant approaches to these problems.

2 Editing the Training Data to Improve Performance

Methods that have as their goal the improvement of recognition acuracy and generalization, rather than the reduction of the size of the stored training set, are called *editing* rules in the pattern recognition literature. In 1972 Wilson [9] first conceived the idea of editing with this goal in mind, and proposed the following elegant and simple algorithm. Delete all points (in parallel) misclassified by the *k-NN*-rule. Classify a new unknown pattern Z using the 1-*NN* rule with the *edited* subset of $\{X, Y\}$.

This simple editing scheme is so powerful that the error rate of the 1-*NN* rule that uses the edited subset converges to the Bayes error as n approaches infinity.

3 Thinning the Training Data to Reduce Storage

In 1979 Toussaint and Poulsen [8] used d-dimensional Voronoi diagrams to delete "redundant" members of $\{X, Y\}$ in order to obtain a subset of $\{X, Y\}$ that implements *exactly* the same decision boundary as would be obtained using all of $\{X, Y\}$. For this reason they called their method *Voronoi condensing*. The algorithm in [8] is very simple. Two points in $\{X, Y\}$ are called *Voronoi neighbors* if their corresponding Voronoi polyhedra share a face. First mark each point X_i if all its Voronoi neighbors belong to the same class as X_i. Then discard all marked points. The main problem with Voronoi-condensing is that the complexity of computing all the Voronoi neighbors of a point is prohibitive in high dimensions.

Bhattacharya [1] and Toussaint, Bhattacharya and Poulsen [7] generalized Voronoi condensing so that it would discard more points in a judicious and organized manner so as not to degrade performance unnecessarily. The Delaunay triangulation is the dual of the Voronoi diagram. Therefore an equivalent description of Voronoi-condensing is to discard all points (in parallel) if all their Delaunay neighbors belong to the same class. The idea is then to use subgraphs of the Delaunay triangulation in exactly the same manner. Experimental results obtained in [1] and [7] suggest that the Gabriel graph is the best in this respect. This procedure will be referred to as Gabriel-thinning in the following. Two points are Gabriel neighbors if the smallest hypersphere that contains them, contains no other points. The Gabriel graph is obtained by connecting two points with an edge if they are Gabriel neighbors.

4 Filtering the Training Data for Fine Tuning

Brighton and Mellish [2] proposed a new hybrid method and compared it to several other hybrid methods on 30 different classification data sets. Their elegant and simple algorithm, which appears to be the previous best in practice, is called *iterative case filtering* (ICF), and may be described as follows. The first part of the algorithm consists of preprocessing with the original Wilson editing scheme. The second part of their algorithm, their main contribution, is an adaptive condensing procedure. The rule for discarding an element (X_k, Y_k) of $\{X, Y\}$ depends on the relative magnitute of two functions of (X_k, Y_k) called the *reachable* set of (X_k, Y_k) and the *coverage* set of (X_k, Y_k). The *reachable* set of (X_k, Y_k) consists of all the data points contained in a hypersphere centered at X_k with radius equal to the distance from X_k to the nearest data point belonging to a class different from that of X_k. More precisely, let $S(X_k, Y_k)$ denote the hypersphere with center X_k and radius $r_k = min\{d(X_k, X_j)|Y_j \neq Y_k\}$ minimized over all j. Then all the data points of $\{X, Y\}$ that are contained in $S(X_k, Y_k)$ constitute the reachable set of (X_k, Y_k) denoted by $R(X_k, Y_k)$. The *coverage* set of (X_k, Y_k), denoted by $C(X_k, Y_k)$, consists of all the data points in $\{X, Y\}$ that have (X_k, Y_k) in their own reachable set. More precisely, $C(X_k, Y_k)$ consists of all data points $(X_i, Y_i), i = 1, 2, ..., n$ such that (X_k, Y_k) is a member of $R(X_i, Y_i)$. The condensing (thinning) step of the ICF algorithm of Brighton and Mellish [2] can now be made precise. First, for all i mark (X_i, Y_i) if $|R(X_i, Y_i)| > |C(X_i, Y_i)|$. Then discard all marked points. This condensing step is repeated until no marked points are discarded. We will refer to this second iterative step of their overall procedure as the *filtering* step of ICF.

5 The New Hybrid Gabriel Graph Algorithm

Our new approach to the problem of instance-based learning depends heavily on the use of the Gabriel graph. First we describe the approach using the exact Gabriel graph, and after than we turn to the practical version for high dimensional problems.

Step 1: The original training set $\{X, Y\}$ is subjected to editing with a modification of Wilson editing. Instead of editing with the k nearest neighbors of a point, we use the Gabriel neighbors, thus dispensing with the problem of choosing a value of k. Let $\{X, Y\}'$ denote the edited training set.

Step 2: The set $\{X, Y\}'$ is subjected to thinning (condensing) using the Gabriel graph rule: points are discarded (in parallel) if all their Gabriel neighbors belong to the same class. Let $\{X, Y\}''$ denote the resulting edited-thinned training set.

Step 3: Subject the set $\{X, Y\}''$ to the *filtering* step of the ICF algorithm of Brighton and Mellish [2]. Let $\{X, Y\}'''$ denote the final training set obtained.

Decision rule: A new query point Z is classified according to the majority vote among the Gabriel neighbors of Z in $\{X, Y\}'''$.

In high dimensional spaces computing the exact Gabriel graph may be costly for large training sets. Therefore we use a new data structure that computes *approximate* Gabriel neighbors in a practical and efficient manner. The practical version of our algorithm uses the *approximate* Gabriel graph instead of the Gabriel graph at every step. We call this the Hybrid Approximate Gabriel-Graph Algorithm. The data structure called GSASH [6] is a modification of SASH [4] to handle Gabriel neighbors rather than the originally intended k nearest neighbors. The query time of GSASH is $O(dk \log k + \log m)$, where m is the size of $\{X, Y\}'''$ and k is the number of approximate Gabriel neighbors computed. Extensive experiments comparing our algorithm with the best available using the standard data sets from the UCI Machine Learning repository [5] suggest that our algorithm is better than any other on the market.

References

[1] Binay K. Bhattacharya. Application of computational geometry to pattern recognition problems. Ph.d. thesis, School of Computer Science, McGill University, 1982.

[2] Henry Brighton and Chris Mellish. Advances in instance selection for instance-based learning algorithms. *Data Mining and Knowledge Discovery*, 6:153–172, 2002.

[3] Luc Devroye, László Györfi, and Gábor Lugosi. *A Probabilistic Theory of Pattern Recognition.* Springer-Verlag New York, Inc., 1996.

[4] Michael Houle. SASH: A spatial approximation sample hierarchy for similarity search. Tech. Report RT-0517, IBM Tokyo Research Laboratory, 2003.

[5] C. J. Merz and P. M. Murphy. UCI repository of machine learning database. Internet http://www.ics.uci.edu/mlearn/MLRepository.html, Department of Information and Computer Science, University of California.

[6] Kaustav Mukherjee. Application of the gabriel graph to instance-based learning. M.sc. project, School of Computing Science, Simon Fraser University, 2004.

[7] G. T. Toussaint, B. K. Bhattacharya, and R. S. Poulsen. The application of Voronoi diagrams to nonparametric decision rules. In *Computer Science and Statistics: The Interface*, pages 97–108, Atlanta, 1985.

[8] G. T. Toussaint and R. S. Poulsen. Some new algorithms and software implementation methods for pattern recognition research. In *Proc. IEEE Int. Computer Software Applications Conf.*, pages 55–63, Chicago, 1979.

[9] D. L. Wilson. Asymptotic properties of nearest neighbor rules using edited data. *IEEE Transactions on Systems, Man and Cybernetics*, 2:408–420, 1972.