Draft of September 22, 2013, at 11:46 pm

Pomona College
Department of Computer Science

A Machine Learning Approach to Diagnosis of Parkinson's Disease

Sumaiya Hashmi

September 20, 2013

Submitted as part of the senior exercise for the degree of
Bachelor of Arts in Computer Science

Professor Sara Sood

Draft of September 22, 2013, at 11:46 pm

Abstract

I will investigate applications of machine learning algorithms to medical data, adaptations of differences in data collection, and the use of ensemble techniques.

Focusing on the binary classification problem of Parkinson's Disease (PD) diagnosis, I will apply machine learning algorithms to a dataset consisting of voice recordings from healthy and PD subjects. Specifically, I will use Artificial Neural Networks, Support Vector Machines, and an Ensemble Learning algorithm to reproduce results from [MS12] and [GM09].

Next, I will adapt a regression dataset of PD recordings and combine it with the binary classification dataset. I will determine the performance of the above algorithms on this consolidated dataset.

Performance of algorithms will be evaluated using k-fold cross validation and a confusion matrix. Specificity and sensitivity will be calculated, as these are of particular importance in medical diagnosis. I will also determine accuracy, precision, recall, and F-score.

Past related work has used either a regression dataset alone to predict a Unified Parkinson's Disease Rating Scale score for PD patients, or a classification dataset to determine healthy or PD diagnosis. The datasets have not been combined, and the regression set has not been used to contribute to evaluation of healthy subjects.

Contents

List of Figures

List of Tables

Chapter 1

Background

1.1 Machine Learning techniques

1.1.1 Artificial Neural Networks

Artificial neurons were first proposed in 1943. Motivated by biological neurons, the artificial neuron received several weighted inputs and produced an output, based on some threshold [MP43]. The perceptron model built on this early work, adding a learning rule to improve the performance of the neural network [Ros58]. However, the perceptron model was severely limited, unable to solve non-linearly separable functions such as XOR [MP69]. Eventually, multilayer perceptrons were developed to address the original perceptron model's shortcomings.

A checkers-playing program used neural networks to train a player [CF01]. The board was encoded as a vector of available board positions, with values assigned based on whether the square was empty, taken by a regular piece, or taken by a King. The neural network consisted of three hidden layers. The first hidden layer completed spatial preprocessing, representing each subsquare of the board as a node, for a total of 91 nodes. The second and third hidden layers had 40 and 10 nodes, respectively. The network outputted a value between -1 and 1, representing the goodness of the board from the current player's

perspective. The weights for the network were initially specified through a uniform sample, and several networks played against one another. The winners were declared 'parents', and they generated 'offspring networks' with weights varied by a parameter vector. The process was repeated for many generations to produce an ideal neural network.

1.1.2 Support Vector Machines

Support vector machines are binary classifiers that can be applied to linearly separable datasets. They separate data into classes using a hyperplane. SVMs can also be used non-linearly by mapping the data to a higher-dimensional space, thus making the data separable. This mapping is done by a kernel function. SVMs perform well with large feature spaces, as long as the data is separable with a wide margin. They also do well with sparse datasets, as in text classification [Joa98].

In the absence of large amounts of labeled data, pool-based active learning can be utilized with SVMs [TK01]. The learning algorithm has access to a pool of unlabeled data, and is able to choose a subset of that pool to use as training data. The learner chooses pool data to use such that the data minimizes the size of the learner's set of hypotheses, and brings it closer to a single hyperplane. This approach allows for using less labeled data.

1.1.3 Ensemble Learners

Ensemble learners combine different machine learning algorithms. There is no one algorithm that always performs well on all domains, and ensemble learners are a way of combining the advantages of different learners. A good ensemble algorithm will be made up of diverse base learners that have varied strengths. The different learners can be combined in a number of ways. They can work in parallel on all of the inputs, and their outputs can be combined in some way. Alternatively, a multistage combination will train the base learners

on different subsets of the input data. For example, AdaBoost first trains an initial learner, and then trains subsequent learners on data that the first learner misclassifies. This way, the weaknesses of each base-learner are made up for by the next learner [FS95].

1.2 Applications to Medical Data

Medical diagnosis presents an ideal domain for machine learning algorithms. A large part of diagnosis falls under pattern recognition, based on large amounts of data, and ML algorithms are well-suited to this task. For an algorithm to be effective in this domain, it needs to be able to handle noisy and missing data, rely on relatively few medical tests, and complement the role of physicians [Kon01]. Machine learning algorithms have been applied to a variety of medical data, some examples of which are outlined below.

1.2.1 Self-reported input

Some diagnoses rely largely on patient-reported information, rather than biological tests. A prime example of this is diagnosis of mental disorders, which is based on how a patient's symptoms compare to criteria outlined in the Diagnostic and Statistical Manual of Mental Disorders. Symptoms are determined through consultation with a mental health professional, and are largely reported by the patient. Automated systems have been proposed which will produce a diagnosis based on user-reported information [YC96].

1.2.2 Clinical Decision Support Systems

Clinical decision support systems help healthcare professionals make diagnosis decisions based on patient data. These systems can be rule-based, in which case they are created with a knowledge base and a set of rules. Alternatively, they can utilize machine learning to learn from past data and recognize patterns. Several such systems have been proposed,

including a statistical approach to diagnosing digestive disorders based on a computer interview of the patient [SK84]. However, the use is not yet widespread, in part due to lack of data availability and to limited adoption of uniform computer systems [Gre07].

1.2.3 EEG and EKG data

Recordings of electrical activity in the body can be used to diagnose a variety of disorders. Electroencephalograms (EEGs) are recordings from the brain and contain a wealth of features that can be used by machine learning algorithms. A classification algorithm using EEGs was able to diagnose Alzheimer's Disease with 86.05% accuracy [Pod12].

Electrocardiograms (EKGs) are often used to detect arrhythmia, which is any abnormality of the heartbeat. They can be indicative of heart disease and other conditions. In 1989, a model was derived from the Cleveland Clinic heart disease data set and compared it to CADENZA, a Bayesian algorithm. Both models were found to overpredict heart disease, though this occurred more with CADENZA [Det89].

After that, a novel machine learning approach to diagnosing and classifying cardiac arrhythmia was presented, called the VF15 algorithm. It used a genetic algorithm to learn feature weights. Then, each feature 'voted on a class prediction. The algorithm had a 62% accuracy on this task and was found to outperform Naive Bayes [GAD98].

Another study collected data on ischemic heart disease, including signs and symptoms, EKG, and scintigraphy. Several algorithms were applied, including Naive Bayes, neural networks, k-nearest neighbors, and two decision tree algorithms. These were compared to clinicians diagnoses. Naive Bayes had the best sensitivity/recall, whereas clinicians, followed by neural nets, had the highest specificity [KKG+99].

More recently, a group compared various machine learning algorithms for arrhythmia diagnosis based on EKG data, with an emphasis on minimizing false positives and dealing

with noisy data. They used the UCI Machine Learning Repository Arrhythmia dataset, and highlighted the need to improve on VF15s 62% accuracy. They evaluated a Bayesian artificial neural network classifier as compared to Naive Bayes, decision trees, logistic regression, and neural networks [GMCL05].

1.3 Diagnosing Parkinson's Disease

Parkinson's Disease (PD) is a degenerative neurological disorder marked by decreased dopamine levels in the brain. It manifests itself through a deterioration of movement, including the presence of tremors and stiffness. There is commonly a marked effect on speech, including dysarthria (difficulty articulating sounds), hypophonia (lowered volume), and monotone (reduced pitch range). Additionally, cognitive impairments and changes in mood can occur. Traditional diagnosis of Parkinson's Disease involves taking a neurological history of the patient and observing motor skills in various situations. Monitoring progression of the disease over time requires repeated clinic visits by the patient. There is no cure, but pharmacological treatment to manage the condition includes dopaminergic drugs.

Speech tests can be used for monitoring Parkinsons disease, due to vocal impairment being a common symptom and early indicator. Using an at-home recording device, such as one developed by Intel for PD telemonitoring, can conveniently allow PD patients' health to be monitored remotely. Specified voice recordings can be passed through signal processing algorithms and a classification and regression tree to predict a rating on the unified PD rating scale [TLMR10].

Another study described a weakly supervised multiple instance learning approach to detecting symptoms of Parkinsons Disease. This approach addressed the issue of self-reporting resulting in inaccurate or incomplete labels [DAlTH12].

Gil and Johnson used a multilayer network with one hidden layer and an output layer

that output healthy or PD. The inputs were passed through a sigmoidal activation function, and gradient descent backpropagation was used to modify the weights. They achieved a classification accuracy of 92.31%. They also trained an SVM using the sequential minimal optimization (SMO) algorithm. SMO speeds up training of SVMs, particularly those with non-linear kernel functions (Platt 1998), using a divide and conquer approach. Gil and Johnson used a linear kernel with 91.79% accuracy, and a Pearson VII function kernel, with accuracy of 93.33% [GM09].

Mandal and Sairam also used a neural network with a sigmoidal activation function. They modified weights using backpropagation with dynamic learning rate and momentum, and achieved an accuracy of 97.6471%. They also used SVM with a linear kernel and obtained an accuracy of 97.6471% [MS12].

Bibliography

[CF01] Kumar Chellapila and David B. Fogel. Evolving an expert checkers playing program without using human expertise. *IEEE Transactions On Evolutionary Computation*, 5:422–428, 2001.

[DAlTH12] Samarjit Das, Breogan Amoedo, F. De la Torre, and Jessica Hodgins. Detecting parkinsons' symptoms in uncontrolled home environments: A multiple instance learning approach. In *Annual International Conference of the IEEE*, pages 3688–3691, 2012.

[Det89] R Detrano. International application of a new probability algorithm for the diagnosis of coronary artery disease. *American Journal of Cardiology*, 64:304–310, 1989.

[FS95] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Proceedings of the Second European Conference on Computational Learning Theory*, EuroCOLT '95, pages 23–37, London, UK, UK, 1995. Springer-Verlag.

[GAD98] H. Altay Guvenir, Burak Acar, Glsen Demirz, and Ayhan ekin. A supervised machine learning algorithm for arrhythmia analysis. In *In: Proc. Computers in Cardiology Conference*, pages 433–436, 1998.

[GM09] David Gil and Devadoss Johnson Manuel. Diagnosing parkinson by using artificial neural networks and support vector machines. *Global Journal of Computer Science and Technology*, 9(4), 2009.

[GMCL05] Dayong Gao, Michael Madden, Des Chambers, and Gerard Lyons. Bayesian ann classifier for ecg arrhythmia diagnostic system: A comparison study. In *Proc. of Int. Joint Conf. on Neural Networks*, 2005.

[Gre07] Robert A. Greens. *Clinical Decision Support: The Road Ahead*. Academic Press, 2007.

[Joa98] Thorsten Joachims. Text categorization with suport vector machines: Learning with many relevant features. In *Proceedings of the 10th European Conference on Machine Learning*, ECML '98, pages 137–142, London, UK, UK, 1998. Springer-Verlag.

[KKG+99] Matjaz Kukar, Igor Kononenko, Ciril Groselj, Katarina Kralj, and Jure Fettich. Analysing and improving the diagnosis of ischaemic heart disease with machine learning. *Artificial Intelligence in Medicine*, pages 25–50, 1999.

[Kon01] Igor Kononenko. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, 23:89–109, 2001.

[MP43] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.

[MP69] M. Minsky and S. Papert. *Perceptrons*. Cambridge, MA: MIT Press, 1969.

[MS12] Indrajit Mandal and N. Sairam. Accurate telemonitoring of parkinson's disease diagnosis using robust inference system. *International Journal of Medical Informatics*, 2012.

[Pod12] V. Podgorolec. Analyzing eeg signals with machine learning for diagnosing alzheimer's disease. *Electronics and Electrical Engineering*, 18:61–64, 2012.

[Ros58] Frank Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.

[SK84] Dj Spiegelhalter and Rp Knilljones. Statistical and knowledge-based approaches to clinical decision-support. systems, with an application in gastroenterology. *Journal of the Royal Statistical Society, Series A: Statistics in Society*, 147:35–77, 1984.

[TK01] S. Tong and D. Koller. Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research*, 2:45–66, November 2001.

[TLMR10] Athanasios Tsanas, Max A. Little, Patrick E. McSharry, and Lorraine O. Ramig. Accurate telemonitoring of parkinsons disease progression by noninvasive speech tests. *IEEE Transactions on Biomedical Engineering*, 57(4):884–893, 2010.

[YC96] R.H. Yap and D.M. Clarke. An expert system for psychiatric diagnosis using the dsm-iii-r, dsm-iv and icd-10 classifications. In *AMIA Annual Symposium Proceedings Archive*, 1996.