Applying Machine Learning to an Alzheimer's Database [1]

Piew Datta **W.R. Shankle** **Michael Pazzani**
Neurology Department

Information and Computer Science Department
University of California, Irvine
(pdatta@ics.uci.edu, rshankle@uci.edu, pazzani@ics.uci.edu)

Abstract. This paper explores the application of Machine Learning (ML) methods for classifying dementia status to improve accuracy over current dementia screening tools: the Blessed Orientation, Memory, and Concentration Exam (BOMC), and the Functional Activities Questionnaire (FAQ). We apply six ML methods to a database of 578 patients and controls. The ML methods are applied in conjunction with these two tests and experimental results show a 15-20% increase in classification accuracy over applying either the BOMC, FAQ, or their combined cutoff criteria. These ML methods also provide simple comprendable criteria for diagnosing dementia status.

Introduction

Correctly diagnosing dementia is a complex problem that requires historical patient data, physical exams, cognitive testing, and laboratory tests. Unfortunately, often times patients' initial visits are to community physicians who frequently do not realize a problem exists (Hoffman, 1982) or misdiagnose the disease (O'Connor et al., 1993). This decreases the time to treat the disease which leads to difficulty when attempting to slow the progression of the disease or minimize behavioral effects. Dementia is defined as multiple cognitive impairments with loss of related functional skills without delirium. Our goal is to analyze Machine Learning (ML) methods to determine if they can improve the accuracy of dementia screening tools recommended by the Agency for Health Care Policy Research (AHCPR) (Williams et al., 1995). ML methods attempt to induce criteria to separate dementia patients when given a set of patient descriptions and their dementia state.

Sample Description

The Alzheimer's Disease Research Center at the University of California, Irvine contains a database of the initial visits of 578 possible patients and controls (either community volunteers or caregivers). Subjects with delirium were excluded from this analysis. The neurologist

[1] This paper will be published as a AAAI Technical Report and be presented at the 1996 AAAI Symposium: Artificial Intellingence in Medicine.

and neuropsychologist applied the DSM-IV criteria to classify the subject's dementia status as either normal, cognitively impaired but not demented, or demented. To test the potential predictive power of ML methods for predicting dementia state, we extracted from the Alzheimer's database subject age, sex, job, and education, plus responses of patients to questions contained in two recommended tests from the AHCPR: the six-item Blessed Orientation, Memory and Concentration Exam (BOMC) and the Functional Activities Questionnaire (FAQ). The BOMC and FAQ tests provide a method for accessing the functional and cognitive abilities of the patient. These data constitute the attributes of the examples used by the ML methods to predict a patient's dementia state.

Machine Learning Methods

Machine Learning methods attempt to learn a description that best separates the different classes of dementia state. As input, each of the initial visits are represented with the attributes described above. The output of each of the learning methods is a representation of the dementia state that can be applied to classify patients with an unknown dementia status. Each learning method applies a different search technique and concept representation to describe the possible outcomes of the diagnosis. We describe briefly how each of the six ML methods learn a representation of the dementia state.

We used the Machine Learning Library in C++ (Kohavi et al., 1994) to test ML methods (C4.5, C4.5 Rules, Naive Bayes, and IB1), plus FOCL (Pazzani & Kibler, 1992) and a new algorithm called PL (Datta & Kibler, 1995) on the Alzheimer's database. C4.5 (Quinlan, 1993) learns a decision tree to classify dementia status based on the input data's attribute values. C4.5 recursively examines these attribute values to create groups of examples all belonging to the same dementia class. The attributes chosen are then used to classify unseen examples. C4.5 Rules (Quinlan, 1993) simplifies C4.5's decision tree and converts it into a set of if-then rules. Naive Bayes (Duda & Hart, 1973) uses Bayes theorem to calculate, for an unseen example, the probability of each dementia class given the unseen example's attribute values and the prior probabilities obtained from the training sample. IB1 (Aha et al, 1991) is an instance-based learner which classifies an unseen example according to the closest example FOCL is a system that learns rules about classes. It learns a rule to describe a class and continues specializing the rule until it excludes examples from other classes. It repeats this process until all objects of a class are described by at least one rule. PL (Prototype Learner) learns at least one prototype for each of the classes. It represents disjunctions within classes with more than one prototype for the class. PL applies a top down hill climbing search to find these disjunctions. Each learned prototype describes the most typical attribute values for members of the class. PL classifies the unseen examples by finding the example's closest prototype and predicting that prototype's class.

Table 1: Accuracy of ML algorithms

Training size	C4.5	C4.5 Rules	Naive Bayes	IB1	PL	FOCL
112	81.60%	81.95%	80.23%	78.99%	83.10%	82.14%
223	82.51%	82.43%	82.61%	80.86%	82.48%	82.38%
335	82.75%	83.47%	83.43%	81.36%	82.18%	82.78%
446	83.31%	83.74%	84.44%	81.54%	83.43%	82.85%

Methodology and Results

We ran these algorithms with a training and a testing sample. The algorithms learn a description of the three dementia classes from the training sample, which is randomly selected from the 578 examples. The algorithms were then run on the remaining examples (testing sample) to rate how well they can correctly predict the classes for unseen examples. The classification accuracy of an algorithm is the number of correct predictions divided by the total number of predictions. To view the effect of increasing the training sample, we experimented with training sizes of 112, 223, 335, and 446 examples [2].

Table 1 shows the results of this experiment. The first column represents the number of patient examples in the training sample and the remaining columns show each of the algorithms' classification accuracies on the disjoint test sample. Each data point in the table is the mean of 30 runs of randomly sampled examples. The maximum standard deviation was 0.64% for any point. For this sample, using the dementia decision criteria for the FAQ (FAQ $>$ 8) and BOMC (BOMC $>$ 10) tests recommended by the AHCPR, we obtained classification accuracies of 69% and 63% respectively, which is 14% to 20% worse than the best results obtained with these ML methods. Combining the two tests (FAQ & BOMC) results in a 60% classification accuracy.

All of the algorithms have similar classification accuracies, with Naive Bayes performing the best. Table 2 shows the specific class accuracies for C4.5, FOCL, and PL with a training size of 446 and class accuracies for the BOMC, FAQ, and FAQ&BOMC tests. Although the AHCPR cutoffs for the BOMC and FAQ tests results in 100% accuracy for the normal class, they have very poor accuracies for the high risk cognitive impairment class. C4.5, PL and FOCL have higher classification accuracies for both the demented and cognitive impairment groups.

Only three algorithms, C4.5 Rules, FOCL, and PL generate easily understandable representations of dementia status described with the attributes. An example of the more directly interpretable if-then rules generated by C4.5 Rules is shown in Figure 1. The numbers shown in brackets after each rule indicates its accuracy on a particular test sample. The FAQ and BOMC totals have been normalized between 0 to 1. Although these if-then rules can aid an

[2] These training sizes were chosen by MLC++. We merely stated the number of intervals and the minimum testing sample size.

Table 2: Class Accuracies

Class	C4.5	PL	FOCL	BOMC	FAQ	FAQ&BOMC
sensitivity for demented	91%	88%	95%	83%	74%	64%
sensitivity for cog. imp.	60%	70%	42%	19%	7%	24%
specificity for normal	68%	81%	52%	100%	100%	100%

informed observer in diagnosing possible dementia patients, these rules do not provide a general description of each class. The prototypes learned by PL can aid in describing the typical dementia patient attributes on the BOMC and FAQ tests. Figure 2 shows an example of one of the dementia prototypes learned with training size 446. Combining the rules from C4.5 rules and the prototypes from PL can give an informed observer the necessary information to make a more knowledgable choice for diagnosis.

Figure 1. An example of the generated C4.5 Rules

```
Rule 1: If Months_backwards > 0 and
        Recall_item1_score > 0 and
        Total_BOMC_Error_Score > 0.111111 and
        FAQtotal <= 0
        Then class normal  [80.5%]

Rule 2: If Total_BOMC_Error_Score > 0.357143 and
        FAQtotal > 0
        Then class demented  [98.5%]

Rule 3: If FAQtotal > 0.259259 and age > 67
        Then  class demented  [98.3%]
...
```

Figure 2. An example of one of the dementia prototypes

```
Prototype: Class Demented
   YEAR_SCORE = 0, MONTH_SCORE = not relevant, TIME_OF_DAY_SCORE = 1,
   MONTHS_BACKWARDS = 0, COUNT_20_BACKWARDS_TO_1_SCORE = 2,
   RECALL_ITEM1_SCORE = 0, RECALL_ITEM2_SCORE = 0, RECALL_ITEM3_SCORE= 0,
   RECALL_ITEM4_SCORE = 0, RECALL_ITEM5_SCORE = 0,
```

```
TOTAL_BOMC_ERROR_SCORE = 0.5236 to 0.6344
FAQ1 = 3, FAQ2 = 3, FAQ3 = 3, FAQ4 = 0, FAQ5 = 2,FAQ6 = 3, FAQ7 =0,
FAQ8 = 0, FAQ9 = 2, FAQ10 = 0, FAQTOTAL = 0.5176 to .6131,
SEX = not relevant, AGE =  65.82 to 86.08, EDUCATION = 10.5 to 15.5,
OCCUPATION = 5.
```

Discussion

These experimental results show that ML methods can detect dementia with a 15-20% increase in classification accuracy over applying either the FAQ, BOMC or their combined cutoff criteria. The AHCPR criteria do not separate the cognitively impaired status from those considered normal. ML methods can separate these two groups, resulting in the high risk cognitively impaired group receiving medical attention at an earlier time. More importantly, ML methods in conjunction with the results of the FAQ and BOMC results can be applied to create simple statements to help classify the dementia status of patients. In addition to these statements, prototypes learned by PL can aid in better understanding the characteristics of each class. Future work entails developing a guideline from ML methods that can be applied by clinicians for diagnosing dementia status.

Bibliography

Aha, D., Kibler, D., & Albert, M (1991). Instance-based Learning Algorithms. Machine Learning v. 6 pp 37-66 Boston: Kluwer Academic Publishers.

Datta, P. & Kibler, D. (1995). Learning prototypical concept descriptions, in Proceedings of the Twelfth International Conference on Machine Learning. Morgan Kaufmann, Los Altos, California.

Duda, R. & Hart, P. (1973). Pattern classification and scene analysis. New York: John Wiley & Sons.

Hoffman, R. S. (1982). Diagnostic errors in the evaluation of behavioral disorders. *JAMA*, 248:225-8.

Kohavi, R., John, G. & et al. (1994). MLC++: a machine learning library in C++ in Tools for Artificial Intelligence Conference, IEEE Computer Society Press.

O'Conner D., Fertig A., Grande M., Hyde J., Perry J., Roland M., Silverman J., Wright S., (1993). Dementia in general practice: the practical consequences of a more positive approach to diagnosis. *Br J Gen Pract*, 43:185-8.

Pazzani, M. & Kibler, D. (1992). The utility of knowledge in inductive learning. Machine Learning v. 9 no. 1. Boston, MA: Kluwer Academic Publishers.

Quinlan, J. R. (1993). C4.5: Programs for Machine Learning, Morgan Kaufmann, Los Altos, California.

Williams, TF. & Costa, PT. (1995) Recognition and Initial Assessment of Alzheimer's Disease and Related Dementias: Clinical Practice Guidelines. Department of Health and Human Services, Agency for Health Care Policy and Research: Office of the Forum for Quality and Effectiveness in Health Care Department.

Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). 4th edition. Publisher American Psychiatric Association, Washington, D. C., 1994.