Predicting progression of Alzheimer's Disease with CSF biomarker, genotype, and MRI data

Josh Tycko JOSHTYCKO@GMAIL.COM
Spencer Penn SPENN321@GMAIL.COM
Juan Jose Lopez Delgado JUANLOP@SEAS.UPENN.EDU

Abstract

Machine learning algorithms have the potential to predict Alzheimer's disease (AD) progression by analyzing large clinical, genomic, biomarker, and MRI datasets. None of these data types has a standard predictor alone. Using data for 800 patients from the ADNI1 database, we implemented three strategies to generate clinically useful predictions. After attempting two strategies based on the Mini-Mental State Examination measure of cognition, we determined to implement a more clinically relevant strategy, quantifying changes in cognition with the Alzheimer's Disease Assessment Scale Cognitive Subscale (ADAS-cog). A decision tree learner was optimal because it can handle mixed data and missing values, while returning interpretable results that can be relatively transparent in the clinic. We used varying combinations of data types and feature selection methods to gain insight into which features may be most informative for AD physicians to analyze in the future.

1. Introduction

Alzheimer's disease (AD) is predicted to affect 1 in 85 people globally by 2050, causing dementia and eventual death. Care in the US costs $100 billion annually, and the available drugs can only help relieve some symptoms (Duthey, 2013).

1.1. Motivation

It is currently difficult to predict the progression of AD, and it often progresses undiagnosed for years. Machine learning algorithms have the potential to assist doctors and patients by accurately predicting disease progression based on clinical, genetic, MRI, and cerebrospinal fluid (CSF) biomarker data, which could enable accurate, early diagnoses. In addition, such an algorithm could potentially accelerate drug development by helping pharmaceutical companies better recruit patients with progressing dementia who stand to benefit from the drugs. Currently, patient recruitment for clinical trials is a major obstacle to drug development, in part because it is difficult for doctors to discern which of their cognitively normal (CN) or mildly cognitively impaired (MCI) patients are likely to develop more significant dementia in the future (Watson, 2014).

1.2. Related work

Since the causes of AD are currently unknown and there are no laboratory tests that can accurately perform a diagnosis, AD progression is quantified with psychological tests like the mini-mental state examination (MMSE) - a questionnaire used to measure cognitive impairment. This set of 30 questions was developed in 1975 and remains the most widely used test by doctors (Doerflinger, 2007). The Alzheimer's Disease Assessment Scale Cognitive Subscale (ADAS-cog) is another such test, developed in 1984, and is now the standard assessment used in clinical trials (Meade, 2005).

Machine learning algorithms have been used on ADNI data with varying success to predict the change in MMSE and ADAS-cog (Stonnington, 2010). Interestingly, in a large study comparing approaches, no single algorithm was superior across all AD datasets, particularly when progression was measured up to varying time points (Umer, 2011)

2. Materials & Methods

2.1. Data

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). First, we used data that was collected over 2 years in 767 patients, including mental examinations and genotype in order to predict the progression of AD over time. Progression is quantified by the change in MMSE score over a 24 month period

(ΔMMSE).

We also enlarged the data set to include the patients for whom the ADAS-cog score was measured, and then quantified progressing disease as a change in ADAS greater than 4 points after a time of 6 months or later. For these patients, we additionally considered MRI and CSF biomarker data. The MRI images were analyzed at UCSF with the Free Surfer software package to quantify thickness, volume, and surface area of brain regions. The CSF biomarker analysis was performed at UPenn, and includes key markers associated with AD: amyloid beta, tau, and p-tau 181.

2.2. Learning Approach

First, we split the data for 767 patients into training and testing sets. As we were interested in creating an algorithm that could predict disease progression quantitatively, we applied a linear regression to predict (ΔMMSE). Since several of our features were categorical data, we used dummy variables to break them into multiple features that could be weighted accordingly in the model.

Next, we developed a system to categorize patients by Normal Cognition, Mild Impairment, Moderate Impairment, and Severe Impairment; based on MMSE scores (Table 1). We used these classes to begin applying classification algorithms, including SVM, Decision Tree, and Naive Bayes.

Table 1. Interpretations of the MMSE Score, ranging from full normal cognition to sever impairment – which is closely correlated with the presence of dementia. We used this breakdown for classification style ML methods.

COGNITIVE STATES	MMSE SCORE (OUT OF 30)
NORMAL COGNITION	$>$ 27 POINTS
MILD IMPAIRMENT	27 - 23 POINTS
MODERATE IMPAIRMENT	23 - 20 POINTS
SEVERE IMPAIRMENT	$<$ 20 POINTS

Then, we presented our results to ADNI scientists at UPenn and gained further insight into their database. We determined ADAS-cog was a more clinically useful and accurate quantifier of AD disease progression than MMSE. We incorporated additional features for the patients at baseline, such as MRI and CSF biomarkers, when possible. However, not every patient had every possible test done (Table 2).

The datasets were pre-screened to remove irrelevant features. Researchers at UCSF included quality control (QC) results for each instance of the MRI data - we removed any instance that did not pass all QC tests (977 instances). Many of the physician diagnostic features were scarcely reported (for $<$25% of patients), so we limited this data set to only include the physician's overall diagnosis and their confidence in that diagnosis.

Table 2. The updated dataset quantifies progression with the ADAS-cog test, which is more thorough than MMSE. Varying fractions of these patients had other features tested at baseline. A physician diagnosis takes into account a holistic approach, including self-reported memory loss, and includes the physician's confidence in the diagnosis. The ApoE genotype, and specifically the ApoE4 allele, is a strong risk factor for AD. MRI data quantifies brain morphology. The CSF biomarkers of amyloid beta and tau are associated with plaques and tangles in the AD brain.

FEATURE TYPE	NUMBER OF PATIENTS
ADAS-COG	819
PHYSICIAN DIAGNOSIS	819
APOE GENOTYPE	747
BRAIN MRI	556
CSF BIOMARKERS	416

We determined to use a decision tree learner on this data set, given its advantages in this specific domain (see Discussion). We used the `SciKitLearn` decision tree, which is an optimized version of the CART algorithm. This is an updated version of the C4.5 decision tree, that constructs binary trees using the feature and threshold with the largest information gain. Trees are considered as sets of if-then rules, and the rules are ordered depending on their accuracy. We trained the learner on combinations of the data groups {Diagnosis, Genotype, CSF Biomarkers, MRI}. Their ADAS-cog scores and physician diagnosis at baseline were included in every data group. The decision tree classified patients' disease as "progressing" if their ΔADAS ≤ -4 six months after baseline or later, and "not progressing" otherwise (Figure 1). This same threshold has been used in clinical trials to determine which patient's neurological health is declining (Rockwood, 2007). The tree's predictions were compared against actual test data, in order to create confusion matrices from which we could compute accuracy, precision, and recall.

3. Results

3.1. Linear Regression with MMSE as output

With an 80:20 training/testing split, the linear regression predictor explained 60% of the variance in our (ΔMMSE).

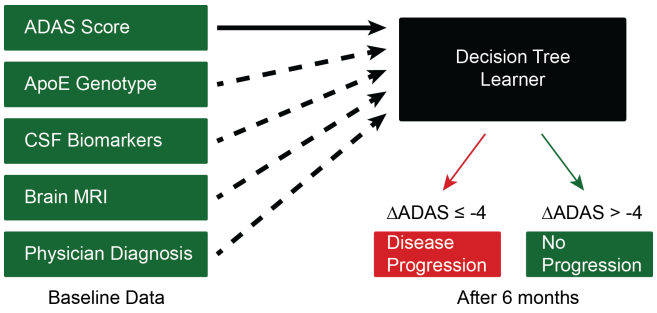

Figure 1. An illustration of the classification approach using ADAS-cog to quantify disease progression. For any given patient, data may be available for their ApoE Genotype, CSF Biomarkers, Brain MRI, and Physician Diagnosis, but the ADAS score at baseline is always included.

The residual sum of squares was 10.42.

3.2. Classification Algorithms with MMSE as output

After switching to a classification problem, we were able to obtain somewhat better results, while still using MMSE to quantify disease progression (Table 3). We tried five different classification algorithms, SVM, K-Nearest Neighbors, Naive Bayes, Decision Trees and ADABoost, on different parameters. The SVM model with a penalty of 0.1 and a linear kernel gave us the best results, with a training accuracy of 64.4% and testing accuracy of 67.5%.

Table 3. Classification accuracies for different variations the machine learning models SVM, Decision Trees, and Naive Bayes on dataset categorized by MMSE labels from Table 1.

MODEL		PARAMS	TRAIN	TEST ACC.
SVM	C = 0.1	LINEAR	64.4%	67.5%
SVM	C = 3	LINEAR	62.8%	66.2%
SVM	C = 0.5	GAUSS	70.0%	65.6%
SVM	C = 1	POLY	70.0%	61.0%
K-NN	K = 5		65.9%	65.6%
N. BAYES	MULTI		62.5%	65.6%
D. TREES	DEPTH	2	61.5%	65.6%
ADABOOST	D.TREE		57.1%	59.7%

3.3. Decision tree classification with ADAS-cog as output

4. Next Steps

Enlarging the dataset

This approach could be applied to the ADNI GO and ADNI 2 datasets, which are extensions of the data investigated here in some ways. They have many features in common, but also include additional features which may also be informative. In the future, we could include data from a study done in 2012, the AddNeuron trial. This includes additional clinical evidence and better genomic features. Furthermore, we are also considering complementing our database with the Penn AD Core Center database. Namely, we want to look specifically at some of the genome sequencing data to attempt to derive meaning from imputed genomes, and determine if genes besides ApoE would inform our predictions.

5. Discussion

Our first few attempts to create a useful algorithm resulted in little success but much insight into the importance of domain specific knowledge in machine learning. Initially, we quantified progression using a regressor to predict an MMSE numerical value after 24 months. This was misguided because clinicians are more interested in using MMSE to categorize patients by the severity of their dementia, as opposed to pinning down an exact MMSE number. So, we classified patients into four cognitive state classes by MMSE and achieved test accuracies $>$ 65%. Upon discussing this work with specialist physician-scientists, we determined that AD drug developers and doctors would be more interested in a classifier that quantified AD progression with the ADAS-cog test. We also included new features that are thought to be informative, including CSF biomarkers and MRI quantification of brain morphology, as well as physician's holistic diagnoses. After pre-screening this new dataset for the highest standard of QC, we determined to use a decision tree to classify patients disease progression, using their baseline data.

The decision tree was an appropriate learning algorithm for several reasons. They are strong performers for mixed data types, handle missing values well, are robust to outliers, scale easily, handle irrelevant inputs, and are mostly interpretable. It was crucial that our system have a high degree of interpretability, given that we aspire to apply it in a clinical setting to aid physicians in their decision-making. They would not be likely to accept a "black box" algorithm. Also, ADNI data and other AD patient databases often include irrelevant features. Here, we pre-screened the data to ensure all features' relevance, but others may not follow the same procedures, so it is useful to use an algorithm that can appropriately ignore such features. Since the data comes from several institutions there are sometimes discrepancies in which features are reported for a given patient. Decision trees can handle such missing values at a node by assigning the most common value of all examples sorted to that node, for example. Other methods can also handle this issue effectively.

Acknowledgments

We would like to thank Dr. Leslie Shaw and Dr. John Trojanowski for their advice on the best usage of the database and the relationship between the data features and the disease, especially with regards to CSF biomarkers and MRI data. Dr. Jon Toledo's help in selecting critical datasets of high quality from the large ADNI database was invaluable.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

Doerflinger, D. Carolan. How to try this: The mini-cog. *Elektronika IR Elektrotechnika*, 107(12):62–71, 2007.

Duthey, B. Background paper 6.11 alzheimer disease and other dementias. Technical report, World Health Organization, Paris, France, 2013.

Meade, C. Diagnosing dementia: mental status testing and beyond. *Australian Prescriber*, 28(1):11–13, 2005.

Rockwood, K. The clinical meaningfulness of adas-cog changes in alzheimer's disease patients treated with donepezil in an open-label trial. *BMC Neurology*, 7(1): 26–34, 2007.

Stonnington, C. Predicting clinical scores from magnetic resonance scans in alzheimer's disease. *NeuroImage*, 51 (4):1405–1413, 2010.

Umer, R. *Machine learning approaches for the computer aided diagnosis and prediction of Alzheimer's disease based on clinical data*. PhD thesis, Department of Computer Science, University of Georgia, 2011.

Watson, J. Obstacles and opportunities in alzheimer's clinical trial recruitment. *Health Affairs*, 33(4):574–579, 2014.