IJSRM
INTERNATIONAL JOURNAL OF SCIENCE AND RESEARCH METHODOLOGY
An Official Publication of Human Journals

Human Journals

Research Article

December 2017 Vol.:8, Issue:2

A Study of Using Decision Trees and Artificial Neural Network Technology on Factors Influencing Alzheimer's Disease

IJSRM
INTERNATIONAL JOURNAL OF SCIENCE AND RESEARCH METHODOLOGY
An Official Publication of Human Journals

Chun-Lang, Chang*[1], Yung-Shun, Liao[1], Chun-Jung, Chen[2]

[1]Institute of Industrial Engineering and Management
National Formosa University, Huwei, Yunlin, Taiwan

[2]Neurology Department
Yunlin Christian Hospital, Siluo, Yunlin, Taiwan

Submission: 27 November 2017
Accepted: 5 December 2017
Published: 30 December 2017

HUMAN JOURNALS

www.ijsrm.humanjournals.com

Keywords: Alzheimer's disease, Back-Propagation Network, Decision Tree

ABSTRACT

Alzheimer's Disease (AD) is a very common recent illness for the elders over the age of 65 suffering from dementia. Its prevalence increases by 1% every year, with 4.6% of the total population, and with exceeding 140 thousands of people were diagnosed with dementia now. The AD is a form of dementia resulted from the continuing gradual loss of brain functions over time. The Mini-Mental State Examination (MMSE) test scores are used to assess the extent of degradation in patients. In this study, we conducted three experiments and data was designated for each experiment according to the MMSE scores and data mining classification results. The classification models were constructed using decision tree in accordance with the back-propagation network (BPN) of the artificial neural network (ANN). The results indicate that the decision tree model has the highest accuracy at 92.26% in classification, whereas the model employed both the decision tree and ANN has the lowest accuracy at 81.45%. The objective is in two steps: first, to find factors influencing intelligence degradation. Factors such as smoking habit, gender, carotid intima-media thickness, depression index, etc. all appeared in the experiments. Then, the main factor that affects the scores will be located to provide the reference for physicians when conducting the diagnosis.

INTRODUCTION

Most people have the impression that dementia only strikes the elder population. However, it can be divided into three major types of dementia based on the different causes: (1) degenerative dementia; (2) vascular dementia; (3) dementia caused by illnesses or brain damages. Degenerative dementia accounts for 60% of all dementia cases; vascular dementia takes up about 25%, and approximately 15% of dementia is caused by brain damages or other diseases. Regardless of the cause, they share similar basic characteristics of symptoms. Among all, Alzheimer's disease (AD) in the degenerative dementia category is the most common form. (Jung-Lung Hsu, 2009)

The age of onset in AD patients is usually between the ages of 40 to 90, though AD most commonly occurs to people over 65 years old. There are approximately 2.1 millions of people over the age of 65 in Taiwan areas, of which 140 thousand suffer from dementia. The incidence rate of the AD is around 4.4% to 4.6%; Pai MC (2009); degenerative dementia accounts for 60% of the dementia population. With MMSE and depression scale, this study aims to analyze patients with different degrees of intelligence degradations to find the major factors influencing intelligence degradation, to slow down the rate of degradation of intelligence through the correlation between each factor.

The brain neurons in AD patients show stepwise degradation, making it very difficult to track back the correct onset time of the disease since the memory function continues aging and worsening. The process may take about 8 to 10 years or as long as 15 years; Zeelandia Dementia Association (2009). In this study, we found that the MMSE scores show a stepwise decrease as dementia advances to different stages. Thus, we used the individual health data and pathology reports provided by the case hospital to discuss which factors

impact the MMSE scores most significantly and to increase awareness, screening, and detection, and eventually prevention of dementia.

With recent technological advancement, many data-mining technologies and classification methods have been applied to medical fields to provide clinical assistance; however, patients receive confirmed diagnoses after symptom manifestation. Not to mention that only very little part of the state-of-the-art technologies was used in preventative medicine. In this study, we analyzed some factors to find certain hidden risk factors that are related to the disease when symptoms are not obvious so that prevention can be achieved. Through artificial intelligence, the ultimate goal is locating helpful knowledge in the huge database-using computer's operation and analytical ability to achieve the goals of preventative medicine.

MATERIALS AND METHODS

The modeling analysis is mainly in three stages. The first stage involves data collection and pre-processing, focusing on normalization of the database and coding; the second stage is decision tree analysis, mainly targeting four types of classification methods of C5.0 decision tree, C&R tree (Classification and Regression Tree), QUEST, and CHAID to compare which type has the highest accurate analysis result; the third stage focuses on comparison and explanation of the results from each model.

1. Collect MMSE provided by the case hospital and the physiological data of patients.

2. Combine relevant literature and suggestions from hospital doctors and find possible factors that may affect AD and MMSE scores.

3. Select, screen, and delete unsuitable data.

4. Conduct normalization and coding on the data and produce objectives of related analysis.

5. Construct models using above-mentioned impact factors and analyzed objectives.

a. Model One: Enter variables that influence and analyze causes of illness into the decision tree for the analysis results.

b. Model Two: Enter variables that influence and analyze causes of illness into the artificial neural network for the analysis results.

c. Model Three: Use the nodes produced from decision tree model and enter them into the artificial neural network for the analysis results.

6. Based on the established rules, conduct classification on testing data.

7. Focus on the accuracy of each experiment, analyze, compare, and explain the results.

Data used in this study consists of the MMSE and physiological database of patients at a total of 1156 pieces of information, including patient personal information, MMSE data, medical examination information, and depression scale scores. The analysis is mainly conducted according to the standard as shown in Table 1, separating 881 data as normal patients, 224 as patients with mild cognitive impairment, 76 as patients with mild Alzheimer's disease, and 5 as moderate Alzheimer's patients. 60 factors are used in this study as variables influencing causes of the illness, including two major attributes below:

1. Clinical Attributes: Clinical attributes consist of personal information, medical history, family medical history, etc. There is data of 43 types of attributes.

2. Histopathological Attributes: Histopathological attributes 17 types of attributes data including MMSE scores, depression scale scores, carotid intima-media thickness, and etc.

RESULTS

Model One:

Enter 1156 pieces of data into the clustering algorithm of the decision tree and classify through 60 related factors to obtain a tree diagram and the error rate report of the decision tree. Based on the result, the classification accuracy for C5.0 algorithm, CART algorithm, QUEST classification, and CHAID classification respectively is 92.26%, 80.62%, 79.5%, and 80.19%. Decision rules of C5.0 are selected since it has the highest accuracy at 92.26%.

Model Two:

Conduct normalization of the 60 related factors of the 1150 data using the artificial neural network to obtain the importance of each factor. The parameters used in the artificial neural network of this study are as shown in Table 1. After testing from the experiment group, the classification accuracy obtained is 85.36%.

Table 1. Parameters of Artificial Neural Network

Parameter Setting	Value
Mining Function	BPN classification
Input Layer	59 Neurons
Output Layer	4 Neurons
Hidden Layer	One Hidden Layer, 30 Neurons
Number of Input Data	771 Training Data

Model Three:

Integrate 1156 medical records with the decision tree model in Experiment 1 produces 15 nodes, which are used as the important possible factors that may affect MMSE scores. Use 771 pieces of data as training data and another 385 pieces as test data into ANN models to verify the accuracy of the clinical diagnosis model constructed. The ANN parameters

adopted in this study are as shown in Table 2. The minding result of Experiment Three combining decision tree and artificial neural network shows the final accuracy at 81.45% and an error rate at 18.55%.

Table 2. Parameters of Hybrid Decision Tree and Artificial Neural Network

Parameter Setting	**Value**
Mining Function	Decision Tree & BPN classification
Input Layer	15 Neurons (as Decision Tree Nodes)
Output Layer	4 Neurons
Hidden Layer	One Hidden Layer, 8 Neurons
Number of Input Data	771 Training Data

Among the three sets of different data obtained from the three models above, C5.0 decision model has the highest accuracy followed by BPN model while the combination model of decision tree nodes and the artificial neural network has the lowest accuracy, as illustrated in Table 3.

Table 3. Comparisons of Three Models

	Model One	Model Two	Model Three
Mining Mode	C5.0 Decision Tree	Artificial Neural Network Model	Hybrid Decision Tree and Artificial Neural Network Model
Accuracy	92.26%	85.36%	81.45%

Upon analyzing sorting of the related factors between each experiment, as shown in Table 4, education level is the most important factor based on the weight. After discussions with the physicians, it is found that smoking habit and depression index will also influence MMSE scores of patients given that subjective factors are excluded.

Table 4. Sorting of the Top 15 Factors

Model One	Model Two	Model Three
Education Level	Education Level	Education Level
Depression Score	Type of Career	Age
Sleep Habit	Age	Hip-to-Waist Ratio
Age	Marriage Status	Smoking Habit
Smoking Habit	Exercise Habit	Drinking Habit
Hip-to-Waist Ratio	Vegan Habit	Cholesterol
Stroke or not?	Gender	Heart Beat
Heart Beat	Family History of Blood Pressure Diseases	Depression Score
Having eye problem or not?	Blood Type	Glycated hemoglobin (HbA1C)
Gender	Smoking Habit	Having eye problem or not?
Myocardial infarction	Carotid Intima-Media Thickness	Sleep Habit
Carotid Intima-Media Thickness	Drinking Habit	Carotid Intima-Media Thickness
Glycated hemoglobin (HbA1C)	Vitamin Assumption	Gender
Drinking Habit	Perivascular Disease	Stroke or not?
Cholesterol	Depression Score	Myocardial infarction

DISCUSSION

The objective of this study is to focus on the four stages of intelligence degradation as the research topic while using the genuine data provided by the case hospital as the research subject. After categorizing using the decision tree and artificial neural network of data mining, the results show over 80% accuracy on all the experiments designed specifically for this study. Amongst all, experiment one, the decision tree model has the highest accuracy at 92.26%, followed by the model that uses BPN at 85.36% accuracy and the least accurate model has 81.45% accuracy for the model that combines decision tree and BPN. Among the same factors that appear in both experiment one and experiment two, it is found that education is

the major factor that impacts the degradation of intelligence. Other factors include age, smoking habit, gender, carotid intima-media thickness, and depression scores. Based on currently known studies, only age and education level have shown significant impact on MMSE scores. Through the experiments in this study, we found that smoking habit, gender, depression scores, and carotid intima-media thickness would also affect MMSE scores. With the use of artificial intelligence, the factors that are discovered can provide physicians important reference basis in clinical diagnoses so that dementia can be detected early and preventative measures can be provided to decrease the intelligence degradation speed.

CONCLUSION

1. In this study, 1136 data is collected, including 60 impact factors and 4 degradation stages. Within statistical samples, four types of sample classifications are unevenly distributed, with 851 pieces of the first type of data accounting for 73.6% of the whole. Since each type of data is not collected evenly, more validated data in each different classification is needed in the future to support the results of this study. With enough data from different types, the proposed reference for clinical diagnoses will be inferential when used to assist in early detection of dementia and to launch preventative measures to help decrease the degradation speed of the intelligence.

2. In this study, BPN of artificial neural networks and C5.0 decision tree are used as a predictive model. For future alternatives, data mining technologies such as K-means, genetic algorithms along with association rules should be considered and used to find more representative rules to provide more effective preventative measures.

3. Subjects in this study are from data of the case hospital which restraints to the local regions and characteristics of the samples thereby not suitable for the dementia prevention of the whole Taiwan areas. Future studies are recommended to use the information from the

database of the National Health Bureau so that effective preventative measures for senile dementia in Taiwan can be developed to lower expenditures on health resources and costs as well as for patients to detect symptoms early and receive better quality care.

4. Factors such as smoking habits, gender, depression scores, and carotid intima-media thickness can be used to discuss reasons affecting MMSE scores. Possible reasons for consideration include but not limited to hormonal influence, nicotine influence, the level of depression and the level of vascular obstruction. Once the factors affecting the degradation of intelligence are found, they can be quantified.

REFERENCES

1. Pai MC. Taiwan Alzheimer's Disease Association. [Online] Available http://www.tada2002.org.tw/index.php, August 20, 2009.

2. Tainan City Zeelandia Dementia Association. [Online] Available http://zda.org.tw/, August 19, 2009.

3. Wang, Y.L. (2007). Episodic Memory Feeling-of-Knowing (FOK) in Early Demented Patients with Alzheimer's Disease. Chinese Journal of Psychology. Vol. 49. No. 4: pp. 365-382.

4. Wu, Y.Q., Han, L.C., Lin, X.Z. (2008). Longevity and Successful Aging. Journal of Internal Medicine of Taiwan. No. 19: pp. 394-400.

5. Lee, H.B., Hsu, J.R. (2002). Use Decision Trees to Analyze Factors Influencing the Perceptions of the Public on Causes of Pollutions from Construction projects. Papers from the Fourth Conference on Sustainable Development of Environment and Resources in Taiwan. Disc No: 4-2.

6. Huang, M.H. (2002). A Study of Decision Trees for Search Terms Selections and Mental Models—Using University Student Index in National Taiwan University. Journal of Library and Information Science. No. 17.

7. Hsu, R.L. "Get to Know Dementia" Department of Neurology Shin Kong Wu Ho-Su Memorial Hospital Web Pages. [Online] Available http://www.skh.org.tw/Neuro/know.htm, August 20, 2009.

8. Liu, L.P., Geng, X.Y. (2008). Analysis on Mild Cognitive Impairment for the Seniors-Using 42 cases of Blood Tests. Shaanxi Medical Journal. Vol. 37. No. 9: pp. 1224-25.

9. Petersen, R.C. (2003). Conceptual overview. IN: Petersen RD, ed. Mild Cognitive Impairment: Again to Alzheimer's Disease. New York: Oxford University Press, 1-14.

10. Ronald, C. (2005).Petersen, Ph.D., MD, Mild Cognitive Impairment: Where Are We, Alzheimer Dis Assoc Disord, 19, pp. 166-169.