Predicting MMSE Score from Finger-Tapping Measurement

Ma Jian*

Hitachi (China) Research & Development Corporation, Beijing, China

Abstract

Dementia is a leading cause of diseases for the elderly. Early diagnosis is very important for the elderly living with dementias. In this paper, we propose a method for dementia diagnosis by predicting MMSE score from finger-tapping measurement with machine learning pipeline. Based on measurement of finger tapping movement, the pipeline is first to select finger-tapping attributes with copula entropy and then to predict MMSE score from the selected attributes with predictive models. Experiments on real world data show that the predictive models such developed present good prediction performance. As a byproduct, the associations between certain finger-tapping attributes ('Number of taps' and 'SD of inter-tapping interval') and MMSE score are discovered with copula entropy, which may be interpreted as the biological relationship between cognitive ability and motor ability and therefore makes the predictive models explainable. The selected finger-tapping attributes can be considered as dementia biomarkers.

Keywords: MMSE; finger tapping; copula entropy; linear regression; support vector regression; dementia biomarkers

1. Introduction

The world population is aging [1] and the diseases spectrum of world changes accordingly. In 2013, the WHO study [2] estimated the top 20 leading causes of death in 2030 showing that noncommunicable diseases, instead of communicable disease, will become the major threats to human lives. Among them are dementias, which are common in the elderly people. In the World Alzheimer Report 2009 [3], Alzheimer's Disease International estimated that 36 million people worldwide were living with dementia, with numbers doubling every 20 years to 66 million by 2030, and 115 million by 2050.

Early diagnosis is crucial for dementia patients to make a timely plan of treatment,

* Email: majian@hitachi.cn

intervention and long-term disease management, which will definitely benefit not only patients themselves but their families, caregivers and health service providers. However, research [4] showed that most patients currently living with dementia have not received a timely formal diagnosis. In high income countries, only 20-50% of dementia cases are diagnosed and documented in primary care, and this gap between early diagnosis and unidentified patients is even much greater in low and middle income countries.

Diagnosing dementia involves cognitive assessment which measures brain functions, such as memory, thinking, language skills, attention, problem-solving, and many other mental abilities [5]. Additionally, diagnosis based on blood tests and brain imaging tests may also be helpful for checking for biological evidences to find or rule out the cause of certain symptoms in clinical practice.

Since the first criteria for clinical diagnosis of Alzheimer Disease (also called the NINCDS-ADRDA criteria) was proposed at 1984 [6], clinical guideline for AD diagnosis has been updated several times (at 1994 [7], 2001 [8], 2007 [9], 2011 [10]). In those updates criteria for blood tests and brain imaging tests were evolving as new pathophysiological evidences were accumulated, while criteria for cognitive tests has experienced enduring clinical successes and hardly been revised. Among many candidate instruments for cognitive assessment, the MMSE (Mini-Mental State Examination) has always been recommended as the preference for diagnosis of dementia due to its reliable performance in practice [6-10].

The MMSE, proposed by Folstein et al. [11], is a 30-points instrument for cognitive assessment, consists of 7 groups of questions measuring different aspects of mental state, and takes about 4-20 minutes to conduct. It has been reported to enjoy a very popularity among clinicians since its birth [12].

Though there are many reliable instruments for cognitive assessment besides MMSE (such as Diagnostic and Statistical Manual, Kokmen Short Test, among many others), those tests are still considered to be too complicated and time-consuming for certain clinical settings. In this sense, simple instrument with less time cost and comparable performance is dearly needed for early diagnosis of dementia. Previously, researchers have developed a type of finger-tapping device [13], tried to utilize it as diagnosis tool to estimate MMSE score, and achieved promising results [14]. In this research, we will propose a method for predicting MMSE score from finger-tapping attributes developed with Machine Learning (ML) pipeline.

The contributions of this paper are as follows:

1. A ML pipeline comprising of Copula Entropy (CE) based variable selection and predictive models is proposed, which is generally applicable to other similar ML problems and can leads to explainable ML models;
2. The associations between certain attributes of finger-tapping movement and MMSE score are discovered by CE, which may be interpreted as the relationship between cognitive and motor ability of the elderly people;
3. A method for predicting MMSE score from the selected finger-tapping attributes, is proposed and its effectiveness is validated on real world data. Due to the above associations, the predictive models are explainable for clinical use.

2. Related Works

There are several academic collaboration platforms on Alzheimer and Dementia research, such as Alzheimer's Disease Neuroimaging Initiative (ADNI) [15] and its Japanese and European counterparts (J-ADNI [16] and E-ADNI [17]), European Prevention of Alzheimer's Dementia (EPAD) Consortium [18], Beijing Aging Brain Rejuvenation Initiative [19], which have been contributing to dementia research [20].

ML methods have been applied to build models for diagnosing dementia, especially at early stage, on demographic, biological, neuropsychological, medical imaging data, or their combinations [20-24]. The outcome of the diagnostic models can be degrees of dementia severity or clinical scores. Though promising, those researches are far from practical clinical use, mainly due to limited data, unreliable performance, the 'Black-box' ML models, and fundamentally biomedical in-plausibility.

Smart home, a broader topic for in-home elderly health monitoring technologies, has been reviewed by Liu et al. [25], Demirir and Hensel [26]. Cognitive monitoring is an important topic in these frameworks of technologies. Due to space limits, please refer to the papers and references therein for more details.

3. Methodology

3.1 Copula Entropy

3.1.1 Theory

Copula theory is about representation of multivariate statistical dependence with copula function [27, 28]. At the core of copula theory is Sklar theorem [29], which states that multivariate joint density function can be represented as a product of its marginal functions and copula density which represents dependence structure among random variables. Please refer to [30] for notations.

With copula density, one can define a new mathematical concept, called *Copula Entropy* [30], as follows:

Definition 1 (Copula Entropy). Let $\mathbf{X}$ be random variables with marginals $\mathbf{u}$ and copula density $c(\mathbf{u})$. CE of $\mathbf{X}$ is defined as

$$H_c(\mathbf{X}) = -\int_{\mathbf{u}} c(\mathbf{u}) \log c(\mathbf{u}) d\mathbf{u}. \quad (1)$$

In information theory, Mutual Information (MI) and entropy are two different concepts [31]. Ma and Sun proved that MI is actually a type of entropy -- negative CE [30], as follows:

Theorem 1. MI of random variables is equivalent to negative CE:

$$I(\mathbf{X}) = -H_c(\mathbf{X}). \quad (2)$$

Theorem 1 has simple proof [30] and an instant corollary on the relationship between information containing in joint density function, marginals and copula density.

Corollary 1.

$$H(\mathbf{X}) = \sum_i H(X_i) + H_c(\mathbf{X}). \quad (3)$$

The above worthy-a-thousand-words results cast insight into the relation between MI and copula and therefore build a bridge between information theory and copula theory.

3.1.2 Estimation

It is widely considered that estimating MI is notoriously difficult. Under the blessing of Theorem 1, Ma and Sun proposed a non-parametric method for estimating CE (MI) from data [30], which comprises of only two steps:

- Step 1. estimating Empirical Copula Density (ECD);
- Step 2. estimating CE.

For Step 1, if given data samples $\{\mathbf{x}_1, \dots, \mathbf{x}_T\}$ i.i.d. generated from random variables

$\mathbf{X} = [x_1, \dots, x_N]^T$, one can easily derive ECD using empirical functions

$$F_i(x_i) = \frac{1}{T} \sum_{t=1}^T \chi(X_t^i \leq x_i^i), \quad (4)$$

where $i = 1, \dots, N$ and χ represents for indicator function. Let $\mathbf{u} = [F_1, \dots, F_N]$, and then one derives a new samples set $\{\mathbf{u}_1, \dots, \mathbf{u}_T\}$ as data from ECD $c(\mathbf{u})$.

Once ECD is estimated, Step 2 is essentially a problem of entropy estimation which can be tackled by many existing methods. Among them, k-Nearest Neighbor method [32] was suggested in [30], which leads to a non-parametric way of estimating CE.

3.2 Predictive Models

Linear Regression (LR) models linear relationship between dependent and some independent random variables. Suppose there are dependent random variable Y and an independent random vector X , the LR model is as:

$$y = \mathbf{A}x + \beta + \varepsilon, \quad (5)$$

where $\mathbf{A}, \beta$ are parameters to be estimated, and ε is noise.

Support Vector Regression (SVR) is a popular ML method that learns complex relationship from data [33]. Theoretically, SVR can learn the model with simple model complexity and meanwhile do not compromise on predictive ability, due to the max-margin principle. The learning of SVR model is formulated as an optimization problem [33], which can be solved by quadratic programming techniques after transformed to its dual form. SVR has its nonlinear version with kernel tricks. The final SVR model is represented as

$$f(x) = \sum_i v_i k(x, x_i) + b, \quad (6)$$

where x_i represents support vectors, and $k(\cdot, \cdot)$ represents kernel function.

3.3 The Machine Learning Pipeline

We propose a ML pipeline with the above concept and methods for predicting MMSE score, as illustrated in Figure 1. The pipeline is based on the finger-tapping attributes generated from the collected raw data (magnetic data of finger-tapping movement). With CE as association measure, the pipeline first selects the attributes mostly associated with MMSE score. Such selected attributes are then fed into the trained predictive models (LR and SVR) to predict MMSE scores. The predictive models in the pipeline is not limited to the above two models, and open to others models.

Figure 1. The ML pipeline of prediction tasks.

4. Experiments and Results

4.1 Data

The data in the experiments were collected at Tianjin and Beijing with finger-tapping device [13]. To collect data from finger-tapping test, two type of movements were measured for each test: bimanual in-phase, bimanual anti-phase, which derives 168 attributes totally from both hands on aspects of distance, velocity, acceleration, time, etc. (For more details on the attributes, please refer to [14]). In the experiments, each movement lasts for 15 seconds. All the participants signed informed consent. At Tianjin, 40 people were recruited as subjects, whose age range at 45-84. From 20 out of 40 subjects, finger-tapping data were collected, the rest subjects only presented corrupted or noisy data. A total 142 finger-tapping tests were performed by these 20 subjects. At Beijing, 117 people were recruited as subjects, whose age range at 48-96. Each subject presented one sample data. Each subject at both locations was tested to derive a MMSE score after each finger-tapping task was performed. Note that subjects at Tianjin are mostly healthy people and present high MMSE scores while those at Beijing are mostly with dementias and present relatively low MMSE scores (Figure 2).

Figure 2. Distribution of MMSE scores in the data from Beijing and Tianjin.

4.2 Experiments

With the above data, we did experiments on predicting MMSE score. Experiments used the ML pipeline mentioned in Section 3.3. Association between the finger-tapping attributes and MMSE was measured with CE and the most associated attributes were selected as inputs to the predictive models. The predictive models

were trained and then test on a testing dataset. To evaluate the selected associated attributes on prediction performance, several additional attributes with less association strength were included as inputs of the predictive models in another experiment. In the experiments, CE was estimated with the non-parametric method in Section 3.1.2, and the hyper-parameters of SVR were tuned to get optimal prediction results.

For prediction, the whole dataset was randomly separated into two parts, 80% for training the predictive models and 20% for testing the accuracy of the trained models. The Mean-Absolute-Error (MAE) was used to measure the performance of the predictive model on testing data:

$$MAE = \frac{1}{T} \sum_{t=1}^T |MMSE_t^{pred} - MMSE_t^{true}|, \quad (8)$$

where $MMSE^{pred}$ and $MMSE^{true}$ represent for predicted and true MMSE scores respectively. The final MAE is an average on the results of 100 run of experiments with the above setting. To further measure the power of the proposed method, we divided the samples with clinical cutoff value ($MMSE=24$) into healthy people and patient, and then calculated the diagnosis accuracy of the method.

4.3 Results

Experimental results are shown in Figure 3-6. The associations between 168 attributes and MMSE score measured with CE are shown in Figure 3, from which it can be learned that only some attributes of bimanual in-phase are associated with MMSE while those of bimanual anti-phase do not, and that the attributes ‘Average of interval’, ‘Frequency of taps’, and ‘SD of inter-tapping interval’ of both hands of bimanual in-phase task are mostly associated with MMSE score, as marked with red in Figure 3. The joint distribution of these attributes and MMSE score is plotted in Figure 4.

In the first experiment, LR and SVR with the above 6 attributes as inputs present high accuracy, as illustrated in Figure 5. Comparison on performance of 100 independent experiments between the two models in terms of MAE and diagnosis accuracy are listed in Table 1 and 2 respectively, from which it can be learned that the proposed method achieves good prediction performance (less than 3 points deviation and above 90% accuracy).

In the second experiment, we also try to select more attributes as inputs of the ML models based on association strength. This time, the attributes, ‘Average of local maximum acceleration in flexing movement’ and ‘SD of contact duration’ of both hands of bimanual in-phase task are additionally considered (marked with blue in Figure 3) because they are associated with MMSE scores stronger than the rest attributes. With these 10 attributes, the ML models present a comparable results with the case with 6 attributes (see Figure 6). It can be learned that a few additional attributes can hardly improve the accuracy of the models. Comparison on the performances of between the two models were done based on the results of 100 independent experiments in terms of MAE and diagnosis accuracy, as listed in Table 3 and 4 respectively.

Figure 3. Associations between MMSE and finger-tapping attributes.

Figure 4. Scattering plot of MMSE and the associated attributes.

Figure 5. Prediction Results with 6 attributes.

Figure 6. Predicted Results with 10 attributes.

Table 1. Average Accuracy of LR and SVR in Experiment 1.

Model	LR	SVR
MAE	2.693	2.548

Table 2. Average Diagnosis Accuracy of LR and SVR in Experiment 1.

Model	LR	SVR
Accuracy (%)	91.7	92.6

Table 3. Average Accuracy of LR and SVR in Experiment 2.

Model	LR	SVR
MAE	2.731	2.566

Table 4. Average Diagnosis Accuracy of LR and SVR in Experiment 2.

Model	LR	SVR
Accuracy (%)	91.1	91.9

5. Discussion

In the experiments, the most associated attributes are actually related with each other by definition. Since all the tasks in the experiments last for the same 15 seconds, it is obvious that 'Average of interval' equals to 15 divided by 'Number of taps' and that 'Frequency of taps' equals to 'Number of taps' divided by 15. So they are essentially same things, despite of possible numerical errors of device software. So the two most associated attributes with MMSE are 'Number of taps' (or 'Frequency of taps' or 'Average of interval') and 'SD of inter-tapping interval', which can be considered as dementias biomarkers. The meaning of the former association can be easily understood while how to explain the association between the latter attribute and MMSE is unclear yet.

Among many available predictive models, only LR and SVR are adopted in our experiments because the sample sets in our cases are small. Here simple model, like LR with linearity assumption, is preferred rather than complex models (such as ensemble models or neural networks) in case of over-fitting. SVR is also a good choice for such small sample problems due to its theoretical advantage and practical successes. Experimental results show that the performances of LR and SVR are close in every experiment despite SVR are in nonlinear mode after hyper-parameters are tuned for optimal results. This might suggest MMSE prediction problem could be

approximately linear.

Regarding the accuracy of the two models in the experiments, one can easily learn that LR and SVR show high prediction accuracy with less than 3-points deviation in terms of MAE. Following the common clinical practice that a cutoff score for MMSE is used in dementia diagnosis, most dementia patients in our experiment are labeled correctly by our method in spite of less than 3-points deviation. It is reasonable to expect further improvement on prediction accuracy if larger data sets are collected for better models.

In our ML pipeline, CE plays a vital role on preparing the inputs of the predictive models. As a rigorously defined mathematical concept, CE has its own advantage. Unlike the correlation which is only applicable to linear cases with Gaussian assumption, CE is defined as a model-free measure of association which makes no assumption of the underlying mechanisms. CE, universal measure of statistical dependence, can be explained as a quantity of energy exchange or information transmission within real physical or biological systems. Since CE is a ideal measure of association with good properties, such as multivariate, symmetric, non-negative, etc., the predictive model based on the associations discovered by CE is explainable, which is of vital importance for clinical use.

In this research, strong associations between certain finger-tapping attributes and MMSE score measured by CE could also has biological meanings, which indicate that hands motor ability and cognitive ability are biologically related with each other and are both syndromes of dementia, more broadly aging. Note that, intuitively, finger-tapping task is so easy to perform that it might relate to only some but not all aspects of mental state measured by MMSE. We suggest further physiological and pathophysiological research on the relation between motor ability and cognitive ability, which could lead to refinement of our technology and better understandings mechanism of dementia.

6. Conclusion

In this paper, we propose a method for predicting MMSE score from finger-tapping measurement with machine learning models. The method is developed with a ML pipeline which comprises of CE based variable selection and predictive models. In the variable selection step, based on the data collected from healthy subjects and patients, the associations between certain finger-tapping attributes and MMSE score are measured with CE, and the mostly associated attributes ('Number of taps' and

‘SD of inter-tapping interval’) are selected. Then the selected attributes are fed into the predictive models (LR and SVR) to predict MMSE score. Experimental results showed that the models such built present good prediction performance. Due to the associations discovered by CE, the predictive model is explainable for clinical use.

The experimental results that certain finger-tapping attributes are associated with MMSE score indicates that motor ability and cognitive ability are associated since the finger-tapping attributes measure motor ability and MMSE measures cognitive ability. Such association may means that both abilities are intrinsically related with each other and that declines of both abilities are syndromes of dementias, or broadly aging.

Acknowledgment

The author thanks Lin Xiaolie and Yin Ying for providing finger-tapping data.

References

- [1] United Nations. World Population Prospects. 2017.
- [2] World Health Organization. Global Health Estimates Summary Tables: Deaths by Cause, Age and Sex by various regional grouping. July 2013.
- [3] Alzheimer's Disease International. World Alzheimer Report 2009. 2009.
- [4] Prince M, Bryce R, Ferri C. World Alzheimer Report 2011: The benefits of early diagnosis and intervention. Alzheimer's Disease International, 2011.
- [5] Petersen R C, Stevens J C, Ganguli M, et al. Practice parameter: early detection of dementia: mild cognitive impairment (an evidence-based review). *Neurology*, 2001, 56:1133-1142.
- [6] McKhann G, Drachman D, Folstein M, et al. Clinical diagnosis of Alzheimer's disease Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 1984, 34(7): 939-939.
- [7] Alter M, Byrne T N, Daube J R, et al. Practice parameter for diagnosis and evaluation of dementia (summary statement). *Neurology*, 1994, 44(11): 2203-2206.
- [8] Knopman D S, DeKosky S T, Cummings J L, et al. Practice parameter: diagnosis of dementia (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*, 2001, 56(9): 1143-1153.
- [9] Dubois B, Feldman H H, Jacova C, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *The Lancet Neurology*, 2007, 6(8): 734-746.
- [10] Albert M S, DeKosky S T, Dickson D, et al. The diagnosis of mild cognitive impairment

due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 2011, 7(3): 270-279.

- [11] Folstein M F, Folstein S E, McHugh P R. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 1975, 12(3): 189-198.
- [12] Tombaugh T N, McIntyre N J. The mini-mental state examination: a comprehensive review. *Journal of the American Geriatrics Society*, 1992, 40(9): 922-935.
- [13] Kandori A, Yokoe M, Sakoda S, et al. Quantitative magnetic detection of finger movements in patients with Parkinson's disease. *Neuroscience Research*, 2004, 49(2): 253-260.
- [14] Suzumura S, Osawa A, Nagahama T, et al. Assessment of finger motor skills in individuals with mild cognitive impairment and patients with Alzheimer's disease: Relationship between finger-to-thumb tapping and cognitive function. *Japanese Journal of Comprehensive Rehabilitation Science*, 2016, 7: 19-28.
- [15] Mueller S G, Weiner M W, Thal L J, et al. The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics*, 2005, 15(4): 869-877.
- [16] Iwatsubo T. Japanese Alzheimer's Disease Neuroimaging Initiative: present status and future. *Alzheimer's & Dementia*, 2010, 6(3): 297-299.
- [17] Frisoni G B, Henneman W J P, Weiner M W, et al. The pilot European Alzheimer's disease neuroimaging initiative of the European Alzheimer's disease consortium. *Alzheimer's & Dementia*, 2008, 4(4): 255-264.
- [18] Ritchie C W, Molinuevo J L, Truyen L, et al. European Prevention of Alzheimer's Dementia (EPAD) Consortium. Development of interventions for the secondary prevention of Alzheimer's dementia: the European Prevention of Alzheimer's Dementia (EPAD) project. *Lancet Psychiatry*, 2016, 3(2): 179-186.
- [19] Chen Y J, Xu K, Yang C S, et al. Beijing Aging Brain Rejuvenation Initiative: aging with grace (in Chinese). *SCIENTIA SINICA Vitae*, 2018, 48: 721-734.
- [20] Weiner M W, Veitch D P, Aisen P S, et al. 2014 Update of the Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimer's & Dementia*, 2015, 11(6): e1-e120.
- [21] Williams J A, Weakley A, Cook D J, et al. Machine learning techniques for diagnostic differentiation of mild cognitive impairment and dementia. *Workshops at the twenty-seventh AAAI conference on artificial intelligence*. 2013: 71-76.
- [22] Chen R, Herskovits E H. Machine-learning techniques for building a diagnostic model for very mild dementia. *Neuroimage*, 2010, 52(1): 234-244.
- [23] Korolev I O, Symonds L L, Bozoki A C, et al. Predicting progression from mild cognitive impairment to Alzheimer's dementia using clinical, MRI, and plasma biomarkers via probabilistic pattern classification. *PloS one*, 2016, 11(2): e0138866.

- [24] Doyle O M, Westman E, Marquand A F, et al. Predicting progression of Alzheimer's disease using ordinal regression. PloS one, 2014, 9(8): e105542.
- [25] Liu L, Stroulia E, Nikolaidis I, et al. Smart homes and home health monitoring technologies for older adults: A systematic review. International Journal of Medical Informatics, 2016, 91: 44-59.
- [26] Demiris G, Hensel B K. Technologies for an aging society: a systematic review of "smart home" applications. Yearbook of Medical Informatics, 2008, 17(01): 33-40.
- [27] Nelsen R B. An introduction to copulas. Springer, 2007.
- [28] Joe H. Dependence modeling with copulas. Chapman and Hall/CRC, 2014.
- [29] Sklar M. Fonctions de repartition an dimensions et leurs marges. Publications de l'Institut de statistique de l'Université de Paris, 1959, 8: 229-231.
- [30] Ma Jian, Sun Zengqi. Mutual information is copula entropy. Tsinghua Science & Technology, 2011, 16(1): 51-54. See also arXiv preprint, arXiv:0808.0845, 2008.
- [31] Cover T M, Thomas J A. Elements of information theory. John Wiley & Sons, 2012.
- [32] Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. Physical Review E, 2004, 69(6): 066138.
- [33] Smola A J, Schölkopf B. A tutorial on support vector regression. Statistics and Computing, 2004, 14(3): 199-222.