Journal of Convergence for Information Technology
Vol. 9. No. 8, pp. 248-253, 2019
e-ISSN 2586-4440
DOI : https://doi.org/10.22156/CS4SMB.2019.9.8.248

제한된 볼츠만 기계학습 알고리즘을 이용한 우리나라 지역사회 노인의 경도인지장애 예측모형

변해원
호남대학교 보건과학대학 언어치료학과 교수

Mild Cognitive Impairment Prediction Model of Elderly in Korea Using Restricted Boltzmann Machine

Haewon Byeon
Professor, Department of Speech Language Pathology, Honam University

요 약 노인성 치매의 전 임상단계인 경도인지장애(MCI)를 조기 진단하고, 조기 개입한다면, 치매의 발병률을 줄일 수 있다. 본 연구는 우리나라 지역사회 노인의 MCI 예측 모형을 개발하고 노년기 인지장애의 예방을 위한 기초자료를 제공하였다. 연구대상은 2012년 Korean Longitudinal Survey of Aging(KLoSA)에 참여한 65세 이상 지역사회 노인 3,240명(남성 1,502명, 여성 1,738명)이다. 결과변수는 MCI유병으로 정의하였고, 설명변수는 성, 연령, 혼인상태, 교육수준, 소득수준, 흡연, 음주, 주1회 이상의 정기적인 운동, 월평균 사회활동 참여시간, 주관적 건강, 고혈압, 당뇨병을 포함하였다. 예측모형의 개발은 Restricted Boltzmann Machine(RBM) 인공신경망을 이용하였다. RMB 인공신경망을 이용하여 우리나라 지역사회 노인의 MCI 예측 모형을 구축한 결과, 유의미한 요인은 연령, 성별, 최종학력, 주관적 건강, 혼인상태, 소득수준, 흡연, 규칙적 운동이었다. 이 결과를 기초로 MCI 고위험군의 특성을 고려한 맞춤형 치매 예방 프로그램의 개발이 요구된다.

주제어 : 인공신경망, 제한된 볼츠만 기계, 예측모형, 경도인지장애, 위험요인

Abstract Early diagnosis of mild cognitive impairment (MCI) can reduce the incidence of dementia. This study developed the MCI prediction model for the elderly in Korea. The subjects of this study were 3,240 elderly (1,502 men, 1,738 women) aged 65 and over who participated in the Korean Longitudinal Survey of Aging (KLoSA) in 2012. Outcome variables were defined as MCI prevalence. Explanatory variables were age, marital status, education level, income level, smoking, drinking, regular exercise more than once a week, average participation time of social activities, subjective health, hypertension, diabetes Respectively. The prediction model was developed using Restricted Boltzmann Machine (RBM) neural network. As a result, age, sex, final education, subjective health, marital status, income level, smoking, drinking, regular exercise were significant predictors of MCI prediction model of rural elderly people in Korea using RBM neural network. Based on these results, it is required to develop a customized dementia prevention program considering the characteristics of high risk group of MCI.

Key Words : Neural network, Restricted boltzmann machine, Predictive model, Mild cognitive impairment, Risk factors

*This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2018R1D1A1B07041091).
*Corresponding Author : Haewon Byeon(bhwpuma@naver.com)
Received July 8, 2019 Revised August 1, 2019
Accepted August 20, 2019 Published August 28, 2019

1. 서론

우리나라는 급속한 고령화로 인해 치매 환자가 급증하는 추세이다. 보건복지부가 수행한 치매유병률조사에 따르면[1], 2012년을 기준으로 치매환자는 54만명으로 조사되었는데 이 같은 경향이 지속된다면 2020년에는 84만명, 2030년에는 127만명, 2050년에는 271만명으로 40년 사이에 치매 인구가 5배 이상 증가할 것으로 전망된다. 특히, 우리나라는 전 세계적으로 볼 때, 치매 인구 증가세가 가장 빠르기 때문에 노년기 인지장애의 대책 마련이 시급하다[2].

지난 20년 동안 세계적으로 치매 치료법의 개발을 시도하였지만, 현재까지 치매를 완치할 수 있는 치료법은 개발되지 않았다. 다만, 도네피질(donepezil) 등의 약물을 포함하여 언어-인지 치료를 통해서 초기 단계부터 인지-언어 기능을 체계적으로 관리한다면 치매환자의 인지 감퇴를 지연시킬 수 있다[3]. 이 같은 약물치료는 일찍 시작할수록 효과가 크기 때문에 치매는 조기 검진과 조기 개입이 무엇보다 중요하다.

특히, 치매는 사회경제적으로 막대한 비용의 부담을 필요로 하므로 조기 개입을 통한 치매 노인의 체계적 관리는 보건학적으로 중요한 이슈이다. 2014년 국회보고서에 따르면 2050년에 치매환자에 따른 사회적 비용은 373억 달러에 육박할 것으로 추정이 되는데, 이러한 비용은 국내총생산(GDP)의 약 1.5%에 해당한다[2]. 따라서 치매를 조기에 발견하여 유병률을 감소시킬 수만 있다면, 불필요한 사회경제적 비용의 지출도 줄일 수 있다[4].

치매의 조기 검진의 중요성이 대두되면서, 치매의 전 임상단계로 알려진 경도인지장애(Mild Cognitive Impairment; MCI)에 대한 연구가 주목받고 있다. MCI는 인지기능이 정상의 범위를 벗어날 정도로 감퇴되었지만, 임상적으로 치매의 진단 기준에는 해당되지 않은 상태로서 정상 노화과정에서의 인지감퇴와 치매의 중간단계로 정의된다. 이들은 향후 치매의 발병 가능성이 정상 노인에 비해서 매우 높은 것으로 알려져 있다[5]. 즉, MCI는 임상적으로 치매를 가장 빠른 시기에 발견할 수 있는 단계이다. MCI를 조기에 감별하고 치료를 하면, 치매로의 진행을 지연시킬 가능성이 높기 때문에 MCI의 예측은 치매 중재의 일차적인 목표로서 중요하다.

지난 20년간 다수의 연구를 통해서 MCI의 위험요인으로 성별, 연령, 흡연, 음주, 식습관, 운동, 당뇨병, 고혈압 등이 보고되었다[4,6]. 그러나 이 처럼 개별적인 차원의 위험요인을 탐색하는 수준만으로 MCI의 발병을 설명하는데 한계가 있다. 게다가 MCI의 다양한 위험 요인들 중에서 최우선적으로 영향을 미치는 요인에 대해서도 연구마다 상이한 결과가 보고된다는 점은 임상가로 하여금 치매의 위험요인과 특성을 이해하는 데 장애로 작용한다[7]. 또한, 최근에는 MCI의 관련 요인을 예측하는 데 있어서 우울증 등의 정신건강에 대한 고려의 필요성이 제기되고 있다[8]. 인지장애의 발생양상과 위험요인은 인종에 따라 큰 차이를 보이므로 우리나라 노인의 특성을 반영한 한국형 경도인지장애 예측모형의 개발이 필요하다.

한편, 컴퓨터의 발달로 인해 빅데이터 분석이 확산되면서, 데이터마이닝을 이용한 예측모형의 개발이 주목받고 있다. 데이터마이닝은 훈련 데이터를 이용하여 이미 알려진 속성을 기초로 빅데이터 자료에 대해 예측하는 분석방법이다[9]. 최근 컴퓨터의 발달과 함께 비선형 분석 방법 중 하나인 인공신경망에 대한 연구도 활발히 진행되었는데, 그 중에서도 그 중에서도 제한된 볼츠만 기계학습 (Restricted Boltzmann Machine; RBM) 인공신경망은 구조가 단순함에도 불구하고, 심층 신뢰 신경망(Deep Belief Network; DBN)을 구성하므로 예측력이 높다[10].

본 빅데이터 기반 연구는 우리나라의 지역사회에 거주하는 노인을 대상으로 수행된 빅데이터 자료를 이용하여 RBM 기반 MCI예측 모형을 개발하고 성능을 평가하였다.

2. 연구 방법

2.1 분석 자료

본 연구의 자료원은 2012년도 Korean Longitudinal Survey of Aging(KLoSA)에 참여한 65세 이상 총 3,240명(남성 1,502명, 여성 1,738)의 데이터이다. KLoSA는 한국노동연구원의 주관 하에 조사 수행기관인 TNS Korea가 위탁을 받아서 2012년 7월 7일부터 2012년 12월 31일까지 조사를 수행하였다[11]. 조사표본에 대한 모집단은 전수조사인 2005년 인구주택총조사를 기반으로 하였고, 261,237의 구역이 표본단위(sampling units)로 설정되었다. 설문 조사 방법은 노트북을 사용한 대면조사 방법인 computer-assisted personal interviews(CAPI)를 이용하였다.

2.2 측정

예측을 위한 결과변수는 KLoSa의 원시데이터에 제시되어 있으며 건강검진에 의해 조사된 이분형의 MCI 유병(네, 아니오)으로 설정하였다. 설명변수는 성, 연령(65세 부터

75세, 75세 이상), 교육수준(고등학교 이하, 고등학교 이상), 소득수준(가구 총 소득), 혼인상태(배우자 있음, 이혼/별거, 사별) 흡연(비흡연, 과거흡연, 현재흡연), 음주습관(비음주, 과거음주, 현재음주), 주1회 이상의 정기적인 운동(한다, 안한다), 월평균 사회활동 참여시간(1시간 미만, 1시간 이상), 주관적 건강(좋음, 보통, 나쁨), 당뇨병(네, 아니오), 고혈압(네, 아니오)을 포함하였다.

2.3 분석 방법

2.3.1 모형개발 및 평가

경도인지장애 발병 예측모형 개발을 위해 KLoSA자료를 학습용자료(training data) 70%, 평가용자료(test data) 30%로 분할하였다. 예측모형의 개발은 RBM 인공신경망 알고리즘을 활용하였고, 개발된 예측 모형의 결과는 정분류율과 민감도 및 특이도를 이용하여 성능을 평가하였다.

2.3.2 RBM 인공신경망 알고리즘

RBM 인공신경망은 차원의 감소 및 분류, 선형 기반 회귀 분석, 협업 필터링(collaborative filtering), 특징값 학습(feature learning), 주제 모델링(topic modelling)에 사용할 수 있는 알고리즘이다[9]. RBM 인공신경망은 Fig. 1과 같이 입력층(visible layer) 1개와 은닉층 1개의 두 개의 층으로 구성되는데, 이러한 구조는 심층 신뢰 신경망(DBN:Deep Belief Network)을 구성하는 요소로 이용될 수 있다[10].

각 입력 노드는 데이터에 대한 저수준의 특징값이나 속성을 출력한다. 예를들면 Fig. 2처럼 입력값으로 x를 투입하면 그 은닉층의 노드에 있는 가중치 w와 곱해지고 바이어스(b)를 더한 값(w*x+b)을 활성함수 f()에 넣으면서 노드가 출력된다. 알고리즘에 대한 수식은 식(1)에 제시하였다.

activation f((weight w * input x) + bias b) = output a (1)

위에서 제시한 식(1)은 단일 은닉 노드와 연결되어진 다수의 입력 노드 중 하나만 고려한 식으로서, 실제로는 RBM 인공신경망의 연산은 Fig. 3처럼 여러 개의 입력 노드의 데이터(값)를 전달받아서 서로 다른 가중치 w와 곱해지는 과정을 거친다. 이후 그 총합에 바이어스를 더한 뒤 활성함수를 산출한다. 즉, 실제 은닉 노드의 연산과정은 식(2)와 같다.

f(x_1*w_1 + x_2*w_2 + x_3*w_3 + x_4*w_4 + b) (2)

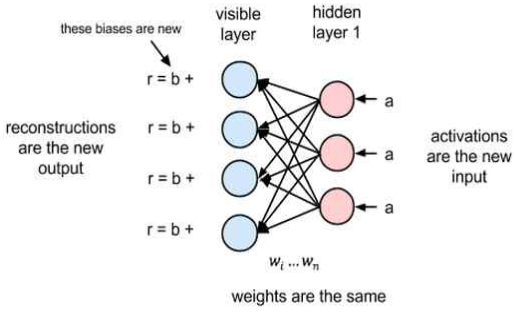


Fig. 1. Concept of Restricted Boltzmann Machines [9]

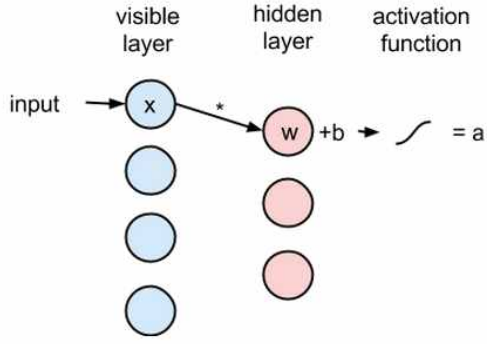


Fig. 2. Input path of Restricted Boltzmann Machines [9]

3. 연구 결과

3.1 대상자의 일반적 특성

총대상자 4,134명의 평균 연령은 72.3세(범위: 65-99, 표준편차: 5.9)로 성별은 남성이 46.4%, 여성이 53.6%이었다. 최종학력은 고등학교 졸업이상이 24.8%이었고, 배우자와 함께 동거하는 노인은 71%이었다. 현재 음주자는 28.8%, 현재흡연자는 13.1%, 주1회 이상 정기적으로 운동을 하는 노인은 37.3%이었다. 20.9%는 당뇨병이 있었고, 49.8%는 고혈압 유병자 이었다.

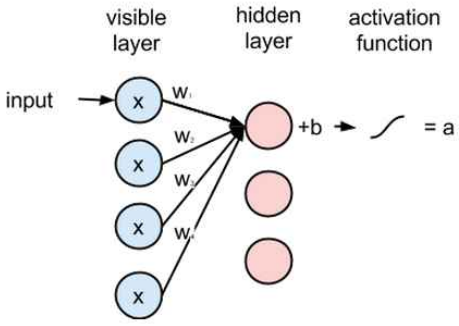


Fig. 3. Weighted input path of Restricted Boltzmann Machines [9]

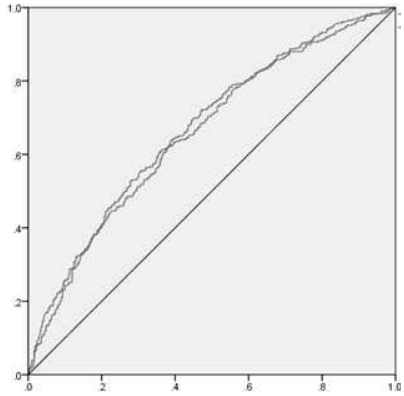


Fig. 4. ROC curve of Restricted Boltzmann Machines

3.2 MCI의 잠재적 위험요인 탐색

MCI의 잠재적 요인은 Table 1에 제시하였다. 전체 대상자 4,134명 중에서 MCI 유병률은 31.3%이었다. 카이제곱검정 결과, 건강한 노인과 MCI는 성, 연령, 혼인상태, 소득수준, 교육수준, 흡연, 음주, 주1회 이상의 정기적인 운동, 주관적 건강, 고혈압, 월평균 사회활동 참여시간에서 유의미한 차이가 있었다($p<0.05$). 75세이상(43.7%), 여성(38.6%), 사별(42.3%), 중학교졸업 이하(36.4%), 주관적 건강이 나쁜 노인(43.9%), 고혈압 유병 노인(34.3%), 비흡연자(34.6%), 비음주자(35.6%), 정기적인 운동을 하지 않는 노인(35.9%), 월 평균 사회활동 참여시간이 1시간 미만인 노인(31.8%)에서 MCI의 유병률이 각각 높았다.

3.3 RBM 인공신경망을 이용한 MCI 예측 요인 탐색

훈련 표본 59.2%, 검정 표본 30.4%, 검증 표본 10.4%에 대한 RBM 인공신경망을 분석한 결과, 데이터의 오차를 최소한으로 감소시키는 은닉층의 수는 총 8개가 도출되었고, 제곱합 오차는 230.5, 분류정확도는 훈련 표본 85.3%, 검정 표본 82.8%, 검증 표본 83.9%로 도출되었다. Fig. 4와 같이 ROC 곡선 면적(AUROC)은 0.71로 모형의 적합도와 설명력은 보통 수준으로 평가되었다.

입력 변수중요도는 Table 2에 제시하였다. 본 RBM 인공신경망에서 연령, 성별, 최종학력, 주관적 건강, 혼인상태, 소득수준, 흡연, 규칙적 운동은 MCI의 가중치가 높은 주요 변수로 가정되었다.

4. 논의

MCI는 치매를 조기에 발견할 수 있는 전임상 단계이자 조기 발견을 통해서 인지 감퇴를 지연시킬 수 있는 질병 전단계라는 점에서 임상적으로 매우 중요하다. 본 연구에서는 RBM 인공신경망 알고리즘에 기반한 예측모형을 개발한 결과, 연령, 성별, 최종학력, 주관적 건강, 혼인상태, 소득수준, 흡연, 규칙적 운동은 MCI의 주요 위험요인이었다. 고령, 교육수준 등의 인구사회학적 요인과 흡연, 음주 등의 건강위험행위는 이미 잘 알려진 MCI의 위험요인이다[2,4].

특히, 본 연구모형에서 흡연은 연령, 성별 등 변경이 불가능한 인구사회학적 요인을 제외한 변수들 중에서 중요도가 가장 높은 변수로 도출되었다. 우리나라에 거주하는 만 65세 이상 남성의 흡연율은 23.3%로서 경제 협력 개발 기구에 가입한 국가들과 비교하면 여전히 높은 수준이다[2]. 상대적으로 노인은 청장년에 비해서 흡연에 노출된 누적 기간이 길뿐만 아니라 금연에 대한 의지도 크지 않기 때문에[12], MCI를 예방하고 건강한 인지기능을 유지하기 위해서는 변경가능한 생활습관인 금연이 필요하다.

본 연구에서 RBM 인공신경망의 예측력은 훈련 표본 85.3%, 검정 표본 82.8%, 검증 표본 83.9%로 83%에서 85%의 높은 정확도를 보였다. 이는 RBM 인공신경망의 알고리즘이 각 층마다 최대한 오차를 줄이는 방향으로 학습이 이루어지는 층별 탐욕 비지도 선행학습(layerwise, greedy unsupervised pre-training)에 기인했기 때문으로 추측된다[10]. 향후 continuous RBM 등의 방법을 이용하여 인공신경망의 예측모형의 성능을 보다 높일 수 있는 연구가 요구된다.

Table 1. Characteristics of the subjects based on MCI, n(%)

Characteristics	MCI		p
	No (n=2,225)	Yes (n=1,015)	
Age			<0.001
Aged 65-75	1,528 (76.3)	475 (23.7)	
75 Years or Older	697 (56.3)	540 (43.7)	
Gender			<0.001
Male	1,157 (77.0)	345 (23.0)	
Female	1,068 (61.4)	670 (38.6)	
Marital Status			<0.001
Marriage	1674 (72.8)	625 (27.2)	
Divorce or Separation	47 (69.1)	21 (30.9)	
Separation By Death	504 (57.7)	369 (42.3)	
Final Education			<0.001
Below Junior High School Graduation	1,549 (63.6)	887 (36.4)	
Above High School Graduation	676 (84.1)	128 (15.9)	
Income Level (Quartile)			<0.001
First Quartile	703 (59.5)	478 (40.5)	
Second Quartile	659 (70.9)	270 (29.1)	
Third Quartile	579 (76.9)	174 (23.1)	
Fourth Quartile	284 (75.3)	93 (24.7)	
Subjective Health			<0.001
Good	502 (82.6)	106 (17.4)	
Usually	1,061 (73.0)	392 (27.0)	
Poor	662 (56.1)	517 (43.9)	
High Blood Pressure			<0.001
Yes	1,061 (65.7)	554 (34.3)	
No	1,164 (71.6)	461 (28.4)	
Diabetes			0.185
Yes	450 (66.6)	226 (33.4)	
No	1,775 (69.2)	789 (30.8)	
Smoking			<0.001
Nonsmoker	1,449 (65.4)	767 (34.6)	
Past Smoker	459 (76.5)	141 (23.5)	
Current Smoker	317 (74.8)	107 (25.2)	
Drinking			<0.001
Non-Drinker	1,138 (64.4)	628 (35.6)	
Past Drinker	382 (70.6)	159 (29.4)	
Current Drinker	705 (75.6)	228 (24.4)	
Regular Exercise			<0.001
Yes	923 (76.3)	287 (23.7)	
No	1,302 (64.1)	728 (35.9)	
Monthly Average Time of Participation In Social Activities			0.001
Less Than 1 Hour	2,134 (68.2)	996 (31.8)	
More Than 1 Hour	91 (82.7)	19 (17.3)	

Table 2. Relative importance of inputs

Inputs	RI	NI
Age	0.186	68.7%
Gender	0.095	45.8%
Final Education	0.088	40.5%
Subjective Health	0.075	35.3%
Marital Status	0.072	33.5%
Income Level	0.068	30.1%
Smoking	0.048	22.6%
Regular Exercise	0.045	20.2%

RI=Relative importance; NI=Normalized importance

5. 결론

본 연구는 인구집단을 대표할 수 있는 검진자료를 이용하여 MCI예측모형을 개발하였다는 강점이 있다. 본 RBM 인공신경망 기반 예측모형에 의하면, 연령, 성별, 최종학력, 주관적 건강, 혼인상태, 소득수준, 흡연, 규칙적 운동은 MCI의 주요 위험요인이었다. 본 MCI예측모형을 바탕으로 노년기 인지장애를 조기 진단이 가능한 맞춤형 모니터링 시스템에 대한 구축이 요구된다.

REFERENCES

[1] Ministry of Health & Welfare. (2013). *Nationwide Study on the Prevalence of Dementia in Korean Elders 2012.* Sejong : Ministry of Health & Welfare.

[2] S. Kim. (2014). *Analysis on Management Policies for the Dementia.* Seoul : National Assembly Budget Office.

[3] P. Anand & B. Singh. (2013). A review on cholinesterase inhibitors for Alzheimer's disease. *Archives of pharmacal research, 36(4),* 375-399.
DOI : 10.1007/s12272-013-0036-3.

[4] H. Byeon et al. (2015). Association of alcohol drinking with verbal and visuospatial memory impairment in older adults: Clinical Research Center for Dementia of South Korea (CREDOS) study. *International Psychogeriatrics, 27(3),* 455-461.
DOI : 10.1017/S104161021400146X.

[5] H. A. Tuokko & D. F. Hultsch. (2013). *Mild cognitive impairment: International perspectives.* New York : Psychology Press.

[6] G. Cheng, C. Huang, H. Deng & H. Wang. (2012). Diabetes as a risk factor for dementia and mild cognitive impairment: a meta-analysis of longitudinal studies. *Internal medicine journal, 42(5),* 484-491.
DOI : 10.1111/j.1445-5994.2012.02758.x.

[7] T. Etgen, D. Sander, H. Bickel & H. Förstl. (2011). Mild cognitive impairment and dementia: the importance of modifiable risk factors. *Deutsches Ärzteblatt International, 108(44),* 743-750
DOI : 10.3238/arztebl.2011.0743.

[8] R. C. Petersen, B. Caracciolo, C. Brayne, S. Gauthier, V. Jelic & L. Fratiglioni. (2014). Mild cognitive impairment: a concept in evolution. *Journal of internal medicine, 275(3),* 214-228.
DOI : 10.1111/joim.12190.

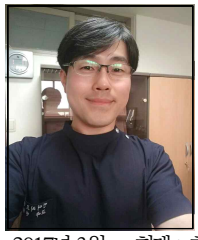
[9] Sky mind. (2019). *Restricted Boltzmann Machines.* Enterprise ML Platform.
https://skymind.ai/kr/wiki/restrictedboltzmannmachine

[10] H. Larochelle, M. Mandel, R. Pascanu & Y. Bengio (2012). Learning algorithms for the classification restricted boltzmann machine. *Journal of Machine Learning Research, 13,* 643-669.

[11] Korea Labor Institute. (2014). *Korean Longitudinal Survey of Ageing 2011.* Sejong : Korea Labor Institute.

[12] D. W. Appel & T. K. Aldrich. (2003). Smoking cessation in the elderly. *Clinics in geriatric medicine, 19(1),* 77-100.

변 해 원 (Haewon Byeon) [정회원]



· 2013년 2월 : 아주대학교 대학원 의학과 (의학박사)
· 2011년 3월 ~ 2013년 2월 : 대림대학교 언어재활과 교수
· 2013년 3월 ~ 2017년 2월 : 남부대학교 언어치료청각학과 교수
· 2017년 3월 ~ 현재 : 호남대학교 보건과학대학 언어치료학과 교수
· 관심분야 : 노년기의사소통장애, 예측 모형 개발
· E-Mai : bhwpuma@naver.com