Journal of the Korea Convergence Society
Vol. 10. No. 8, pp. 21-26, 2019
ISSN 2233-4890
https://doi.org/10.15207/JKCS.2019.10.8.021

C4.5 알고리즘을 이용한 피로도 식별 시스템 구현

김우진[1], 이덕진[2*]
[1]극동대학교 정보통신학과 박사과정, [2]극동대학교 항공IT융합학과 교수

Implementation of Fatigue Identification System using C4.5 Algorithm

You Zhen Jin[1], Deok-Jin Lee[2*]
[1]Doctoral Course, Dept. of Information and Communication, Far East University
[2]Professor, Dept. of Aviation and IT Convergence, Far East University

요 약 본 논문은 C4.5 알고리즘을 이용한 피로 인식 방법을 제안한다. 피로 평가에 대한 국내외 연구를 바탕으로 중국인의 생활습관 및 문화적 특성과 결합하여 피로 자기평가 척도를 완성하였다. 본문에서 사용한 척도는 58개 하위 항목에 적용되어 있으며 피로의 유형과 정도를 평가하는 데 사용되었다. 이 항목들은 육체적 피로, 정신적 피로, 개인의 습관 및 피로의 결과 등을 측정하는 4가지 분류 항목에 포함된다. 본 연구의 목적은 피로 형성의 주요 원인을 분석하고 그에 따른 피로 정도를 인식함으로써 피로에 대한 주관적 관심을 증가시키고 과도한 피로로 인한 심뇌혈관계 질환의 위험을 감소시키는 데 있다. C4.5 알고리즘을 활용한 피로 인식 시스템의 인식률은 평균 85%로 나타나 본 제안의 유용성을 확인하였다.

주제어 : 의사결정 트리, C4.5알고리즘, 인식시스템, 피로도, 패턴 인식

Abstract This paper proposes a fatigue recognition method using the C4.5 algorithm. Based on domestic and international studies on fatigue evaluation, we have completed the fatigue self - assessment scale in combination with lifestyle and cultural characteristics of Chinese people. The scales used in the text were applied to 58 sub items and were used to assess the type and extent of fatigue. These items fall into four categories that measure physical fatigue, mental fatigue, personal habits, and fatigue outcomes. The purpose of this study is to analyze the leading causes of fatigue formation and to recognize the degree of fatigue, thereby increasing the personal interest in fatigue and reducing the risk of cerebrovascular disease due to excessive fatigue. The recognition rate of the fatigue recognition system using the C4.5 algorithm was 85% on average, confirming the usefulness of this proposal.

Key Words : Decision tree, C4.5 algorithm, Recognition system, Degree of fatigue, Pattern recognition

1. 서론

현재 급속하게 발전하는 시장경제 체제는 현대인에게 물질적 풍요를 가져다주었지만, 그와 동시에 현대인을 자기 분야에서 최고가 되어야 생존한다는 무한경쟁 사회로 내몰고 있다. 시장경제의 일반적인 특징은 국가, 기업, 개인들이 경쟁을 통해 생존 및 발전의 공간을 확보하려고 노력한다는 것이다. 경쟁력을 높이기 위해 사람들이 더 많은 노동시간과 노력을 지불하는 것이 일반화되었다. 이에 따라 과로가 현대사회에서 건강에 유해한 중요한 요인 중 하나가 되었다.

일본 세이코쿠 대학의 전임 교수 우에하타 테츠노조는

*Corresponding Author : Deok-Jin Lee(djlee@kdu.ac.kr)
Received May 31, 2019 Revised August 2, 2019
Accepted August 20, 2019 Published August 28, 2019

처음으로 '과로사'라는 개념을 제시하였다[1]. 그는 현재까지 과로사 연구를 진행하고 있으며, "과로사, 자살협상센터"를 운영하고 있다. 1989년에 발간된 『의학의 발전』이라는 저서에서 과로사를 "뇌혈관 질환 및 심장병과 같은 순환계 질환의 원인과 관련된 사회의학 용어로 사용되며, 과도한 노동은 고혈압 및 동맥경화를 악화시켜 뇌출혈과 지주막하 출혈, 심뇌경색을 비롯한 기타 뇌혈관 질환, 허혈성 심장 질환, 급성 심부전 등 급성 발병을 유발하여 영구적인 노동력 손실, 궁극적으로 사망에 이르는 상태를 조성한다."라고 정의를 내렸다[2-4]. 과로사를 직접적으로 촉진하는 5종의 질병을 순서대로 나열해보면 관상동맥 질환, 대동맥 동맥류, 판막 심장 질환, 심근 병증, 뇌출혈 등이다. 임상 연구에서는 표준근무시간을 초과한 노동시간이 길어질수록 심혈관 질환의 발병률이 높아진다는 점에서 발병 위험이 노동시간에 직접적인 영향을 받는다는 것이 밝혀졌다[5-7]. 고용 노동부 자료에 따르면 일주일에 평균 60시간 이상 근무하면 뇌혈관 질환의 위험이 증가하는 것으로 나타났다[8-11].

이에 본 연구의 목적은 피로 형성의 주요 원인 분석과 함께 개인이 피로 정도를 인식하게 함으로써 피로에 대한 관심을 높이고 과로로 인한 심뇌혈관 질환의 위험을 줄이는 데 있다.

2. 피로 설문지 구조

본 연구에서는 피로 평가를 위한 4가지 분류 항목을 58개 항목으로 상세화한'피로 설문지'를 작성하여 실험 대상자가 자기평가를 할 수 있도록 하였다. 설문의 결과는 5가지의 척도로 나누어 C4.5 알고리즘에 적용하였다. 자기평가 피로 설문지의 구조는 제시된 Fig. 1과 같다.

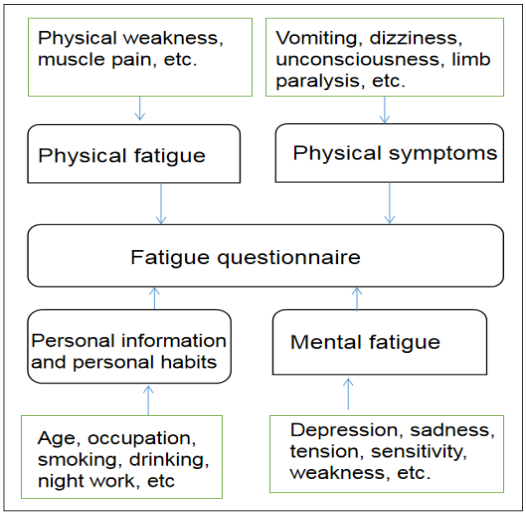

Fig. 1. Diagram of fatigue questionnaire

피로도 평가는 대체로 (1)개인의 사업과 가정 요인 및 자신의 습관적 요인 (2)피로가 신체에 주는 느낌 (3)정신적 피로 (4)주요 증상의 발현과 증상에 대한 환경적 영향 등에 대한 평가로 나뉜다.

Table 1.Fatigue questionnaire item structure

Fatigue questionnaire	
Measurement project	Number of questions
Physical fatigue	10
Mental fatigue	10
Personal habits and personal information	10
Disease symptoms	28

본 논문에서 피로 설문지는 심리적 측정의 개념을 적용한다. 척도는 심리학적 특성과 성향, 인상 등을 정량화하는 도구로서, 정신분석을 위한 정량적 도구이다. 척도에는 정량화, 표준화, 정제 및 객관화의 특성이 있다. 척도의 형식은 이상적인 연구 표본, 색인 및 변수 형식이다. 이러한 척도를 분석한 결과는 과학적 요구 사항과 조건, 규모, 구조의 다차원적 특성에 잘 부합한다. 특히 척도 항목은 명확한 정의와 채점 기준을 가지고 있어야 평가의 중복을 피할 수 있다. 이러한 특성은 임상 실험에서 다차원 주관 증상에 대한 좋은 평가 도구로 활용될 수 있다. 피로 설문지 척도 항목 구조는 Table 1과 같다.

3. "피로도" 식별시스템

본 논문의 척도는 4가지 방향의 특성을 고려한 58개 항목의 주제로 되어있으며 크게 (1)개인정보(예: 성별, 혼인여부, 주소, 연령 등) (2)개인습관(예: 휴식시간, 야간활동 여부, 취미, 음주여부, 흡연 등) (3)신체적 피로(몸에 힘이 부족하다, 사지에 힘이 없다, 허리와 무릎이 시큰거린다, 근육통이 있다 등) (4)정신적 피로(우울, 비관, 초조, 예민함, 기력 감퇴, 절망감, 지연성, 불만, 망설임 등) (5)질병을 통해 나타나는 증상(구토, 현기증, 갑작스런 언어장애 및 기침, 의식 장애, 가슴이 답답하고 두근거림, 팔다리 마비 등) (6)환경 요인(날씨, 온도, 작업 환경 등)

으로 분류할 수 있다. 이를 기존의 의료전문가 시스템과 결합하여 다차원적 주관 증상으로 판단하여 관련된 질병을 추측할 수 있을 것으로 본다.

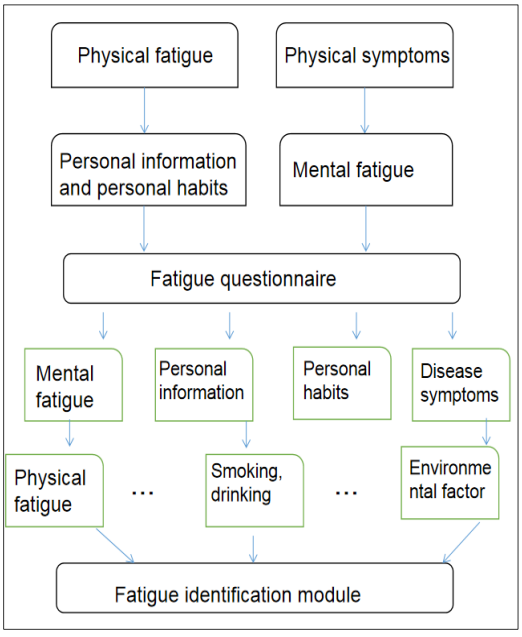

Fig. 2. Fatigue identification module structure

이 피로도 식별구조는 Fig. 2와 같다. 이를 바탕으로 중국인의 특성에 적합하도록 척도를 개발하면 피로와 관련된 질병 및 다른 증상을 가진 사람들의 다양한 피로도와 성격을 평가할 수 있다. 이는 피로 완화의 효과를 효율적으로 평가할 있다는 매우 실용적 의미를 가지게 된다.

3.1 의사 결정 트리

C4.5 알고리즘은 ID3 알고리즘의 장점을 취하고 다음과 같은 측면에서 ID3 알고리즘을 향상시킨다. 1)정보이득비율을 이용하여 속성을 선택하면, 더 많은 속성을 선택하는 경우 정보이득을 사용할 때의 단점을 극복할 수 있다. 2)트리를 구성 때 가지치기 할 수 있다 3)연속적 이산화 수행 능력이 우수하다. 4)불완전한 데이터처리가 가능하다. 한편 C4.5 알고리즘에서는 ID3가 가진 알고리즘의 문제점을 해결하기 위해 의사결정의 깊이를 어느 정도까지 진행할 것인가에 대하여 제안하였으며, 이는 “과적합 문제(over fitting)”와 “가지치기(pruning)”와 관련이 있다[12-16].

C4.5 알고리즘에서 의사결정 트리를 형성하기 위하여 처음 수행하는 작업은 분할정복이다. 입력되는 훈련 집합(Training Set)이 성공적으로 분할되도록 모든 하부집합에 하나의 클래스가 속하는 경우들로 구성될 때까지 트리를 형성한다. m개 클래스의 집합인 D에서 무작위로 선택한 한 클래스를 $C_j(j=1,2,3\cdots,m)$로 하고, $|D|$는 D의 모든 클래스의 개수이며, $|C_j,D|$ 는 집합 D에서 C_j에 속하는 클래스의 개수이다. 집합 D에서 기대정보를 구하기 위해서는 각 클래스가 전달하는 정보를 가중평균하면 된다.

$$Info(D) = -\sum_{j=1}^{m} \frac{|C_j,D|}{|D|} \times \log_2\left(\frac{|C_j,D|}{|D|}\right) \quad (1)$$

위의 info(D)와 비슷한 개념으로, T가 X에 의해 n개로 분할된 후에 기대 정보를 구하려면 식 (2)를 이용한다.

$$Info_X(T) = \sum_{i=1}^{n} \frac{|T_i|}{|T|} \times Info(T_i) \quad (2)$$

X에 의한 분할로 얻어진 정보는 다음 식 (3)에서 얻을 수 있다.

$$Gain(X) = Info(T) - Info_X(T) \quad (3)$$

따라서 T에 있는 한 개의 클래스에 속하는 부분집합을 식별 정의하는 데 필요한 평균 정보의 양으로 정규화한다. 이는 분열 정보 T가 n개의 부분집합으로 분할됨에 따라 발생하는 정보의 양이다. 분열 정보는 다음 식(4)에서 얻을 수 있다.

$$SplitInfo_X(T) = -\sum_{i}^{n} \frac{|T_i|}{|T|} \times \log_2 \frac{|T_i|}{|T|} \quad (4)$$

Gain을 분열 정보로 나누어 줄 때, Split에 의해 생성된 유용한 정보의 비율변수별로 Gain Ratio를 최대화시켜 주는 Split Point를 찾아 각 변수별로 실시하여 변수끼리 최대 Gain Ratio를 비교하고, 그중에서 가장 큰 변수를 선택한다. 정보 이득 비율 정의는 다음 식 (5)로 구성된다.

$$GainRatio(X) = \frac{Gain(X)}{SplitInfo(X)} \quad (5)$$

이로써 트리가 구성된다[17-18].

4. 실험 방법과 결과

4.1 데이터 수집

데이터 수집 대상은 두 팀으로 구성하였다. 2018년 2월 15일부터 2018년 9월 2일까지 연변자치주 6개 주요 도시의 주요병원과 외래 진료소에서 고혈압 및 기타 병으로 앓고 있는 입원 응답자 250명(고혈압병력 200건, 기타 질병 50건)을 A팀으로 구성하였다. 그리고 중소기업에 근무하는 일반인 250명에 대해 설문 결과를 얻었으며 이를 B팀으로 하였다. 총 500명의 사례 중 남성 수는 277명(55.4%), 남성 평균연령은 38(37.6±18.7)세, 여성 수는 223명(44.6%), 여성 평균연령은 37.29(37.29±17.50)세였다.

4.2 피로 자아측정 테이블의 평가 기준

등급 매기기로 인해 항목의 민감도에 영향을 줄 수 있으며 이에 따라 너무 많은 선택 항목을 설치하면 선택의 어려움이 생기는 등의 상황을 고려하여 이 표의 선택 옵션은 매 항목마다 모두 5개 등급으로 설정하였다. 등급은 1-5의 숫자로 표시되며 1에서 5까지의 수는 각각 아래와 같은 의미를 나타낸다. 어떤 항목에서 번호를 선택하면 해당 번호의 역순의 값으로 채점된다. 예: 선택 1, 2, 3, 4, 5등급은 각각 5, 4, 3, 2, 1점으로 채점된다.

4.3 채점기준

1. NFS(No fatigue symptoms) 58개 옵션 중 조건을 만족하는 10% 미만의 상황은 "피로"가 없는 것으로 판단된다. 2. MFS(Mild fatigue symptoms) 58개 옵션 중 조건을 만족하는 20% 미만의 상황은 경미한 "피로" 증상으로 판단된다. 3. ModerateFS(Moderate fatigue symptoms) 58개 옵션 중 조건을 만족하는 30% 미만의 상황은 중등 "피로" 증상으로 판단된다. 4. SFS(Severe fatigue symptoms) 58개 옵션 중 조건을 만족하는 43.1% 미만의 상황은 심한 "피로" 증상으로 판단된다. 5. VSFS(Very severe fatigue symptoms) 58개 옵션 중 조건을 만족하는 44% 이상의 상황은 극심한 피곤 상태로 판단된다.

4.4 분류결과

MATLAB (2014a) 를 사용하여 수집된 데이터 분석 결과는 Fig. 3, 4, 5, 6과 같다.

정신적 피로특성(우울, 비관, 초조, 예민함, 능력감퇴, 절망감, 지연성, 불만, 망설임 등)으로 구성된 정신적 피로 착도 설문을 적용하여 분류한 결과는 Fig. 3과 같다.

신체적 피로 특성(몸에 기운이 빠진다, 사지에 힘이 없다, 허리와 무릎이 시큰거린다, 근육통이 있다 등)과 운동이 피로에 주는 영향, 피로 지속시간 등의 특성으로 구성된 신체적 피로 설문을 적용하여 분류한 결과는 Fig. 4와 같다.

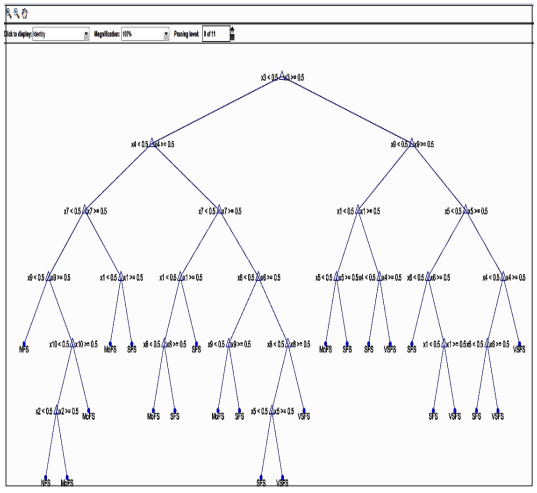

Fig. 3. Mental fatigue classification scheme

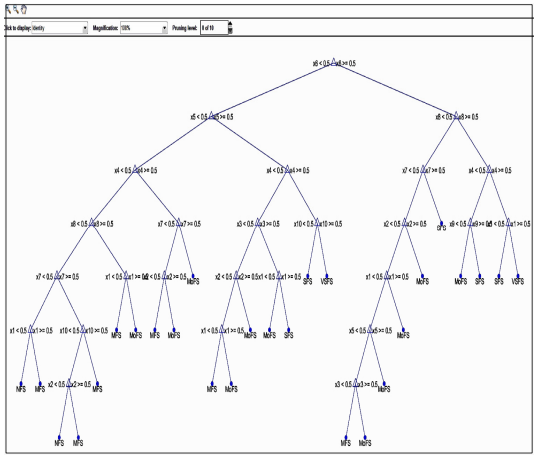

Fig. 4. Physical fatigue classification diagram

개인정보 특성(예: 성별, 혼인여부, 주거지, 연령, 체중 등)와 개인습관 특성(예: 휴무시간, 야간활동 여부, 취미, 음주여부, 흡연 등)으로 구성된 개인습관 착도 설문을 적용하여 분류한 결과는 Fig. 5와 같다.

관련된 질병의 증상(구토, 어지럼증, 언어장애, 기침, 의식 장애, 흉부 압박감, 심계항진, 팔다리가 저린 감각)과 피로가 체력에 미치는 결과, 피로가 신체 기능에 미치는 결과, 피로가 감정에 미치는 결과, 피로가 직장생활 정보에 미치는 결과, 피로가 인지기능에 주는 결과, 피로로 인한 신체의 부적합한 반응 등의 특성으로 구성된 질병의 증상 착도 설문을 적용하여 분류한 결과는 Fig. 6과 같다.

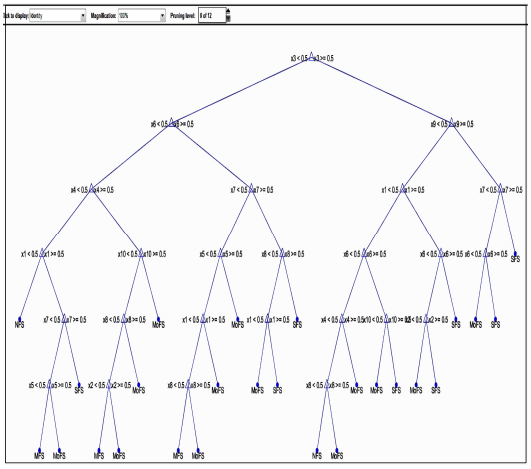

Fig. 5. Personal information and personal habits

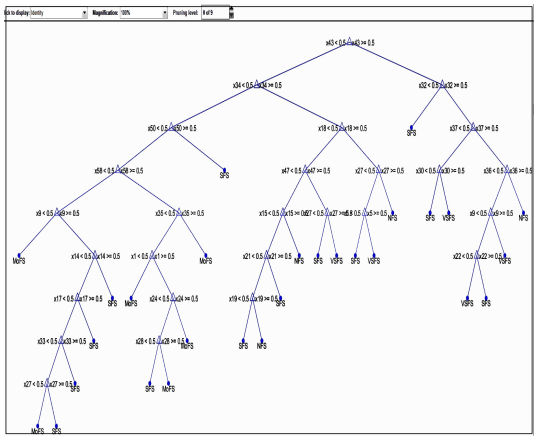

Fig. 6. Results of fatigue classification

이처럼 신체적 피로, 정신적 피로, 개인 습관 및 질병의 증상을 측정하는 4가지 평가기준을 이용하여 분류한 결과는 85% 인식률을 나타냈다.

5. 결론 및 향후 연구과제

국내외에서 피로 평가에 대한 연구를 바탕으로, 중국인의 습관 및 문화적 특성과 결합하여, 피로 자기평가 척도가 작성되었다. 본문에서 서술한 척도는 58개의 항목으로 구성되었으며 피로의 유형과 심각도를 평가하는 데 사용되었다. 신체적 피로, 정신적 피로, 개인 습관 및 피로의 결과를 측정하는 4가지 분류를 C4.5 알고리즘을 적용하여 실험한 결과는 Fig. 7과 같다.

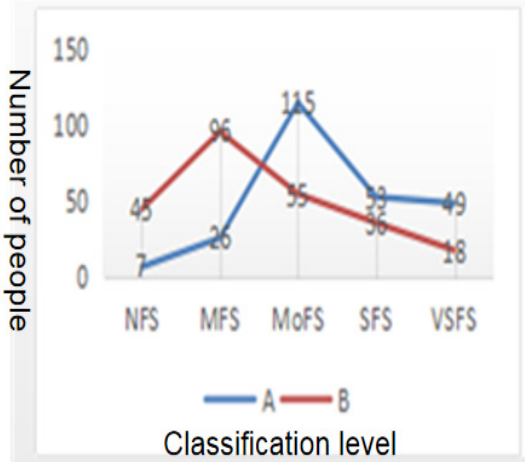

Fig. 7. Recognition rate result figure

전체적으로 A팀의 인식률이 B팀보다 높게 나타났다. 결과적으로 A팀의 40%는 피로의 영향을 받는 것으로 판단되었고, 그중 20%는 향후 급사 위험이 있는 것으로 판단되었다. B팀의 경우 22%가 피로의 영향을 받는 것으로 판단되었으며 그중 7%는 급사 위험이 있었다. 종합하면 500명 응답자 중 31% 이상은 일상생활 중 피로의 영향을 받는 것으로 나타났고 그중 13% 이상이 갑작스런 사망 위험에 처해있는 것으로 나타났다. 결론적으로 피로감은 현대인이 삶의 질과 행복도 저하에 심각한 영향을 미친다.

C4.5 알고리즘을 적용하여 실험한 결과 평균 85%의 인식률을 나타내어 본 제안의 유용성이 확인된다.

본 논문에서 사용한 척도는 피로의 일반적 인 증상을 평가한 질병 인구 및 건강 인구의 특성을 통합하여 평가한 것이라는 점에서 의미가 있다. 따라서 피로로 인한 발병의 억제 효과를 위한 임상적 활용 가능성이 크다고 할 수 있다.

향후 다양한 질병 및 증상이 피로와 어떤 관계를 갖는

지 정확하게 규명하기 위한연구가 필요한 것으로 될 사료된다, 또한 증상의 이름과 피로로 인한 질병의 이름을 질병의 심각성과 연결하고 정리함으로써 데이터 샘플 문제를 해결하고 이를 위해 수많은 데이터 샘플을 수집하여 다른 임상 테스트 데이터에 나타난 특성과 연관 인식률을 향상시키고자 한다.

REFERENCES

[1] F. B. Liu & J. J. Li. (2011). Research on Suicide and Related Legislation in Japan. *Journal of Yunnan University Law Edition, 24(3)*, 124-129.

[2] C. X. Luo. (2005). On the Legal Nature of the Overworked Death of Intellectuals and the Perfection of Labor Law. *Journal of Agents, (1),* 70-73.

[3] Y. Meng & L. Luo. (1992). Japan's overworked death. *World Knowledge*, *16*, 29.

[4] T. Z. Shangyu, X. C. Liu & X. L. Ma. (1992). Background of the so-called overwork death. *Introduction to Japanese Medicine*, *13(1),* 5-6.

[5] M. Kivimaki, G. Batty & M. Hamer. (2011). Using Additional information on working hours to predict coronary heart disease: a cohort study. *Ann Intern Med*, *154(7),* 457-63.

[6] T. W. Jang, H. R. Kim, H. E. Lee & J. P. Myong & J. W. Koo. (2013). Long Work hours and obesity in Korean adult workers. *Journal of Occupational Health*, *55,* 359-366.

[7] T. Amagasa & T. Nakayama. (2013). Relationship between long working hours and depression: a 3-year longitudinal study of clerical workers. *Journal of OCCUP Environ Med, 55(8),* 863-872.

[8] S. Viviers et al. (2008). Burnout, psychological distress, and overwork: the case of Quebec's ophthalmologists. *Can J Ophthalmol*, *43(5),* 535-546.
DOI : 10.3129/i08-132

[9] A. Wirtz, D. A. Lombardi, J. L. Willetts, S. Folkard & D. C. Christiani. (2012). Gender differences in the effect of weekly working hours on occupational injury risk in the United States working population. *J Work Environ Health*, *38(4),* 349-357.
DOI : 10.5271/sjweh.3295

[10] M. Y. Kang et al. (2012). Long working hours and cardiovascular disease. *Journal of OCCUP Environ Med*, *54(5),* 532-537.

[11] T. W. Jang et al. (2014). Overwork and cerebrocardiovascular disease in Korean adult workers. *Journal of occupational health, 57(1),* 51-57.
DOI : 10.1539/joh.14-0086-OA.

[12] T. Devashsish, M. Nisarga & R. D. Sharan. (2010). Re optimization of ID3 and C4.5 decision tree. *2010 International Conference on Computer and Communication Technology(ICCCT).*
DOI : 10.1109/iccct.2010.5640492

[13] Y. Zhang & L. Gong. (2004). *Principle and technology of data mining*, Beijing: Publishing House of Electronics Industry Press.

[14] H. Byeon. (2015). The Factors of Participating in a Smoking Cessation Program using Integrated Method of Decision Tree and Neural Network Algorithm. *Journal of the Korea Convergence Society, 6(2),* 25-30.

[15] S. Y. Oh. (2013). Decision Tree Learning Algorithms for Learning Model Classification in the Vocabulary Recognition System. *Journal of Digital Convergence, 11(9)*, 153-158.

[16] J. K. Lee & H. W. Lee. (2018). Meltdown Threat Dynamic Detection Mechanism using Decision-Tree based Machine Learning Method. *Journal of Convergence for Information Technology, 8(6)*, 209-215.

[17] D. M Li, Li Yan, C. Yuan, C. R. Li & H. Liu. (2016). The application of decision tree C4.5 algorithm to soil quality grade forecasting model. *2016 First IEEE International Conference on Computer Communication and the Internet(ICCCI),*
DOI : 10.1109/cci.2016.7778985

[18] D. John, J. Cai & Z. Cai. (2005). Decision tree technique and its current research, *Control Engineering of China*, *12(1),* 15-18.

김 우 진(You Zhen Jin) [정회원]

· 2013년9월 : 광운대학교 컴퓨터공학과 (공학석사)
· 2015년 9월 ~ 현재 : 극동대학교 정보통신학과 박사과정
· 관심분야 : 패턴 인식 ,알고리즘, 인공지능.
· E-Mail : jinritian23@naver.com

이 덕 진(Deok-Jin Lee) [정회원]

· 1995년 8월 : 광운대학교 전기공학과 (공학석사)
· 2000년 8월 : 광운대학교 전기공학과 (공학박사)
· 2002년 3월 ~ 현재 : 극동대학교 항공IT융합학과 교수
· 관심분야 :센서정보시스템, 패턴인식
· E-Mail : djlee@kdu.ac.kr