

<https://doi.org/10.7236/JIIBC.2019.19.5.151>
JIIBC 2019-5-21

뉴럴네트워크(NEWFM)를 이용한 심근경색의 특징추출과 분류

Feature selection and Classification of Heart attack Using NEWFM of Neural Network

윤희진*

Heejin Yoon*

요약 최근 심근경색은 중장년층의 돌연사의 80%로 밝혀졌다. 심근경색의 발병 원인은 복잡적이고 갑자기 발생하게 되어 예방이나 건강검진을 하더라도 발병을 예측하기 어렵다. 따라서 빠른 진단과 적절한 치료가 가장 중요하다. 이 논문에서는 심근경색에 대한 정확하고 빠른 진단을 위해 가중퍼지소속함수를 이용한 신경망으로 정상과 비정상 분류에 대한 정확도를 나타내었다. 실험에 사용된 데이터는 14개의 특징과 303개의 샘플 데이터로 이루어진 UCI Machine Learning Repository에서 제공하는 데이터 사용하였다. 2개의 특징을 선택하여 제거하였다. 특징선택을 위한 알고리즘은 average of weight method를 사용하였다. 가중퍼지소속함수를 이용하여 심근경색을 정상과 비정상으로 분류(1-normal, 2-abnormal)하였다. 실험 결과 정확도가 87.66%가 나왔다.

Abstract Recently heart attack is 80% of the sudden death of elderly. The causes of a heart attack are complex and sudden, and it is difficult to predict the onset even if prevention or medical examination is performed. Therefore, early diagnosis and proper treatment are the most important. In this paper, we show the accuracy of normal and abnormal classification with neural network using weighted fuzzy function for accurate and rapid diagnosis of myocardial infarction. The data used in the experiment was data from the UCI Machine Learning Repository, which consists of 14 features and 303 sample data. The algorithm for feature selection uses the average of weight method. Two features were selected and removed. Heart attack was classified into normal and abnormal(1-normal, 2-abnormal) using the average of weight method. The test result for the diagnosis of heart attack using a weighted fuzzy neural network showed 87.66% accuracy.

Key Words : classification, heart attack, feature selection, Neural Network

1. 서론

돌연사의 80%를 차지하고 발병 후 초기 사망률도

30%에 달하는 질병이 심근경색이다. 심근경색은 심장에 혈액을 공급하는 관상동맥에 혈전이 막아 심장근육으로 혈액을 공급하지 못하여 발생하는 질병이다.^[1] 높은 사망

*장안대학교 IT학부 인터넷정보통신과
접수일자 2019년 9월 16일, 수정완료 2019년 10월 3일
계재확정일자 2019년 10월 4일

Received: 16 September, 2019 / Revised: 3 October, 2019 /
Accepted: 4 October, 2019
*Corresponding Author: hjoyoon@jangan.ac.kr
Dept. of Internet Communication, Jangan University, Korea

를 보이고 있는 심근경색의 원인은 음주, 흡연, 스트레스, 만성질환 등 여러 원인이 있으나, 예측하기가 어렵고 발병 즉시 빠른 처치가 가장 우선이다.^[2] 심근경색이 발생하면 병원 도착 전 1/3이 사망에 이르게 되는 무서운 질병이다. 최근 사회적 현상으로 급성심근경색이 과거에 비해 증가하고 있는 추세에 있다.^[3] 이 논문에서는 심근경색에 대한 정확하고 빠른 진단을 위해 UCI Machine Learning Repository에서 제공하는 데이터를 가중퍼지소속함수를 이용한 신경망으로 정상과 비정상을 분류하는 실험을 하였다. 아래 그림1.은 실험의 흐름도를 나타낸 것이다.

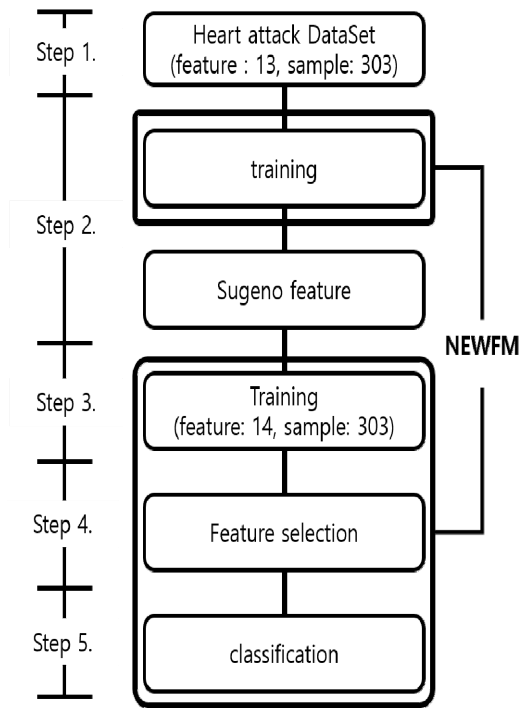


그림 1. 실험 흐름도
Fig. 1. Experiment of flowchart

step 1. UCI Machine Learning repository에서 제공하는 13개의 특징과 303개로 이루어진 심근경색에 대한 데이터를 입력한다.^[4] 제공하는 실험데이터의 속성은 다음 표1과 같다

step 2. 뉴럴네트워크인 NEWFM을 이용하여 데이터를 학습시키고 학습시킨 결과로 sugeno 값을 특징값으로 추출한다.

step 3. sugeno의 값을 특징값으로 포함하여 학습하였다.(14개의 특징)

step 4. 다시 뉴럴네트워크인 NEWFM을 이용하여 데이터를 학습시키고 특징 중 분류과정에 좋은 영향을 주는 특징과 좋지 않은 영향을 주는 특징을 비중복면적으로 특징추출을 한다.

step 5. 특징추출 후 심근경색에 대한 정상과 비정상으로 분류한다.

II. 가중퍼지소속함수(Neural Network Weighted Fuzzy Membership Function(NEWFM))

퍼지(fuzzy)란 인간의 애매모호한 언어를 컴퓨터 언어로 표현하고자 한 이론이다.^[5] 즉, 모호한 상태를 수식화하여 시스템을 구축한다. 가중퍼지소속함수는 퍼지소속함수에 가중치를 적용한 이론이다. 뉴럴네트워크인 NEWFM(Neural Network Weighted Fuzzy Membership Function)은 학습데이터를 통해서 학습을 시킨 후 가중퍼지소속함수를 이용하여 테스트데이터의 클래스를 분류를 할 수 있다.^[5] 학습 과정은 그림2와 같이 특징 당 클래스 개수만큼 퍼지함수셀(FSet)을 생성한다. 퍼지 함수셀을 조정하면서, 멤버쉽함수를 조정한다. 생성된 퍼지함수셀(FSet)을 학습시킨 후 클래스 분류에 좋은 영향을 주는 특징(BEST Feature)과 클래스 분류에 나쁜 영향을 주는 특징(Worst Feature) 특징선택을 한다. 특징선택의 방법은 average of weighted 방법으로 그림3과 같이 각 클래스에 대해 중복면적과 비중복면적을 구하여 비중복면적에 대하여 표준편차를 구하여 그 값에 따라 특징의 순서를 결정한다.

아래 그림2는 멤버쉽함수를 이용하여 각 특징마다 만들어지는 퍼지셀(FSet)을 보여 주고 있다.^[6] 특징선택으로 좋은 영향을 주는 특징들로 다시 학습을 시킨 후 클래스를 판별한다. 클래스를 판별하는 방법은 average of weight 방법과 sum of matched 방법이 있다.

average of weight 방법은 한 레코드에서 각 특징들의 입력값에 대한 비중복면적의 값을 구해서 더한다. 비중복면적의 값은 모든 특징에 대해 클래스1에 대해 구하여 합하고 클래스2에 대해 구해서 합한 후 평균을 내어 더 큰 값의 클래스가 된다.

표 1. 실험 데이터 속성.^[4]

Table 1. Simulation Parameters.^[4]

	속성	설명
1	age	age in years
2	sex	sex (1 = male; 0 = female)
3	cp	chest pain type -- Value 1: typical angina -- Value 2: atypical angina -- Value 3: non-anginal pain -- Value 4: asymptomatic
4	trestbps	resting blood pressure (in mm Hg on admission to the hospital)
5	chol	serum cholesterol in mg/dl
6	fbs	fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
7	restecg	resting electrocardiographic results -- Value 0: normal -- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) -- Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
8	thalach	maximum heart rate achieved
9	exang	exercise induced angina (1 = yes; 0 = no)
10	oldpeak	ST depression induced by exercise relative to rest
11	slope	the slope of the peak exercise ST segment -- Value 1: upsloping -- Value 2: flat -- Value 3: downsloping
12	ca	number of major vessels (0-3) colored by flourosopy
13	thal	3 = normal; 6 = fixed defect; 7 = reversable defect
14	num	diagnosis of heart disease (angiographic disease status) -- Value 0: < 50% diameter narrowing -- Value 1: > 50% diameter narrowing (in any major vessel: attributes 59 through 68 are vessels)

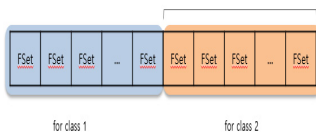
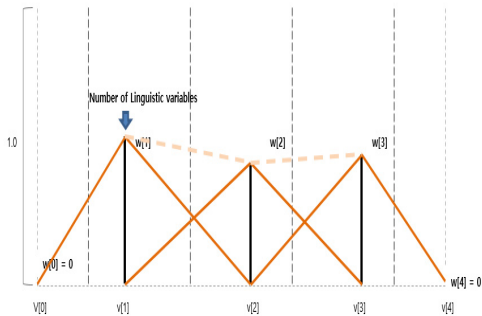


그림 2. 멤버쉽함수를 이용하여 만들어진 퍼지셋
 Fig. 2. FuzzySet using membership function

sum of matched 방법은 샘플데이터 학습 시 각 특징들에 대해 구해진 퍼지셋에 해당하는 클래스의 합을 구한다. 클래스1과 클래스2를 비교하여 많은 클래스로 분류는 하는 방법이다.^[7]

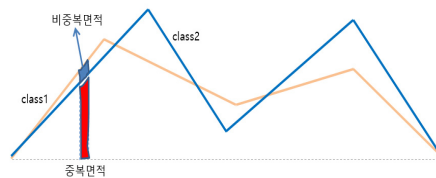


그림 3. 비중복면적
 Fig. 3. Non-weighted Area

III. 실험 및 평가

14개의 특징과 303개의 데이터로 심근경색을 분류하는 실험에서 뉴럴네트워크인 NEWFM의 실험 조건은 표 2와 같다.

표 2. NEWFM의 실험조건

Table 2. Experiment condition of NEWFM

조건(condition)	값(value)
class Decision	average of weight
Epoch	30
Rate for Weights	0.04
Training Count	50000

실험결과 14개의 특징 중 6번 특징(fbs)과 7번 특징 (restecg)이 실험에 좋지 않은 영향을 미쳐 제거가 된 특징이다. 또한 11번(slope), 13번(sugeno), 14번(thal)은 좋은 영향을 미치는 특징으로 나왔다. 특징추출에서 제거가 된 특징의 비중복면적의 그래프는 그림4, 그림5와 같은 그래프를 얻었으며, 실험에 좋은 영향을 미치는 11번, 13번, 14번 특징에 대한 비중복면적 그래프는 그림6, 그림7, 그림8과 같은 결과가 나왔다. 그림에서 보는 바와 같이 클래스1과 클래스2가 비중복면적이 넓을수록 좋은 영향을 주는 특징이고, 비중복면적이 좁을수록 분류에 좋지 않은 영향을 주는 것으로 나타났다. 또한 sugeno값의 특징이 정확도를 높인데 좋은 영향으로 나타난 것을 알 수 있다.

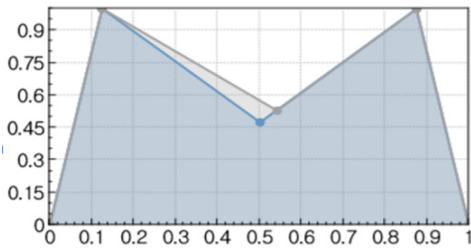


그림 4. 특징 6번의 비중복면적의 결과
Fig. 4. Non-Weighted Result Graph of Feature 6

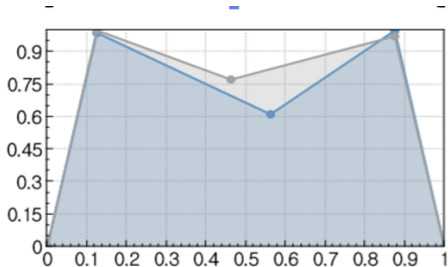


그림 5. 특징 7번의 비중복면적의 결과
Fig. 5. Non-Weighted Result Graph of Feature 7

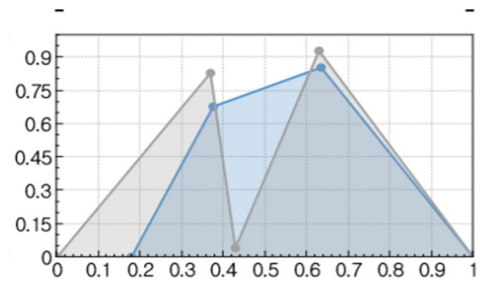


그림 6. 특징 11번의 비중복면적의 결과
Fig. 6. Non-Weighted Result Graph of Feature 11

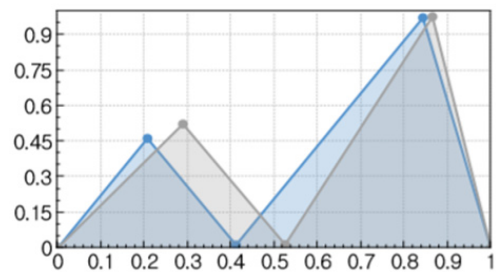


그림 7. 특징 13번의 비중복면적의 결과
Fig. 7. Non-Weighted Result Graph of Feature 13

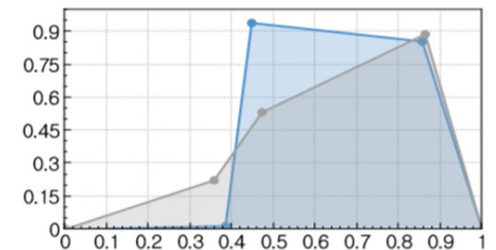


그림 8. 특징 14번의 비중복면적의 결과
Fig. 8. Non-Weighted Result Graph of Feature 14

표 3. NEWFM의 실험 결과 비교

Table 3. Experiment Result of NEWFM

Method	accuracy
Navee Bayes ^[10]	72.5%
Our proposal Method	86.77%

IV. 결 론

가중퍼지소속함수를 이용하여 심장마비의 조기 진단을 위한 실험을 하였다. NEWFM을 이용한 분류에서는

특징선택을 하여 14개의 특징 중 12개의 특징만으로 분류를 하였다. 실험결과 표3과 같이 Naive Bayes의 77.4%보다 높은 86.77%의 정확도를 보였다.^{[8][9]}

[10] M Akhil Jabbar, B L Deekshatulu Priti Chandra, "Classification of Heart Disease using Artificial Neural Network and Feature Subset Selection", Global Journal of Computer Science and Technology Volume XII Issue III Version 1, 2013.

References

- [1] Heart attack from <http://wikipedia.org>
- [2] HeeTack Lee, YouSik Hong, SangSuk Lee, "AED System using Fuzzy Rules", The Journal of The Institute of Internet, Broadcasting and Communication, Vol. 13, No. 4, pp. 215-220, Month 08.
DOI: <http://dx.doi.org/10.7236/IIBC.2013.13.4.215>.
- [3] Cheol-Soo Bae, Young-Cheol Park, Kee-Hwan Nam, Yong-Seok Kang, "Direct Controller for Nonlinear System Using a Neural Network" AED System using Fuzzy Rules", KIIECT, Vol. 5(1), No. 4, pp. 7-12, 2012.-2
DOI: <http://www.kiiectsys.or.kr/sobis/kiiect.jsp>
- [4] Heart attack dataset form
<https://archive.ics.uci.edu/ml/index.php>
- [5] B.H.Wang, J.W.Lim and J.S.Lim, "Gene regulatory network identification from the yeast cell cycle based on a neuro-fuzzy system", Genet.Mol.Res. 15(3): gmr.15039002, August 30, 2016.
DOI: <http://www.dx.doi.org/10.4238/grm.15039002>
- [6] Son SY, Lee SH, Lim JS, "Feature selection for daily pea load forecasting using a neuro-fuzzy system", Multimedia Tools Appl. 74:2321-2336, 2015.
DOI: <http://springer.com/article/10.1007/s11042-1943-0>
- [7] S.Florence, N.G.Bhuvanewari Amma, G. An napoorani, K.Malathi, "Predicting the Risk of Heart Attacks using Neural Network and Decision Tree", IJIRCC Vol. 2, Issue 11, November 2014.
DOI: <http://www.ijirccce.com/2320-9801>
- [8] S.Sherlin, D.S halini Devi, T.Vetriselvi, "An Automated and Efficient Mining of the Healthcare Data for the Prognosis of Heart Attack using the HUI Miner and Naive Bayes Classifier", International Journal of Engineering Research & Technology ISSN:2278-0181, 2017.
DOI: <http://www.ijert.org/2278-0181>
- [9] K.Srinivas, B.Kavihta Tani, "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks". International Journal on ComputerScience and Engineering Vol.02, No. 02, 2010, 250-255, 2010.
DOI: <http://www.kiiectsys.or.kr/sobis/kiiect.jsp>

저 자 소 개

윤 희 진(정회원)



- 2001년 2월 : 동국대학교 컴퓨터공학과 (공학석사)
- 2015년 8월 : 가천대학교 일반대학원 전자계산과 (공학박사)
- 2013년 3월 ~ 현재 : 장안대학교 IT학부 인터넷정보통신과 조교수

※ 장안대학교 2019년 연구지원비로 작성된 논문임.