

TNM병기 클러스터링을 활용한 폐암 최종병기 결정

최철웅, 벨무루간 아레수 발라지, 김경백
전남대학교 전자컴퓨터공학부
e-mail : sentilemon02@gmail.com, arreshvnass@gmail.com,
kyungbaekkim@jnu.ac.kr

Lung cancer Final staging with TNM Staging clustering

Chulwoong Choi, Velmurugan Arresh Balaji, Kyungbaek Kim
Dept of Electronics and Computer Engineering,
Chonnam National University

요 약

폐암 환자들의 암 진행단계에 맞는 최종병기를 결정할 때 T, N, M병기를 기준으로 많이 사용한다. 컴퓨터과학분야에서 TNM병기로 최종병기를 결정할 수 있는지 알아보기 위해 클러스터링 알고리즘을 사용하였다. 두 곳의 병원에서 제공받은 폐암환자들의 T, N, M병기 데이터를 기준으로 각각 K-Means 클러스터링을 진행하고 ARI(Adjusted Rand Index), Homogeneity, Completeness와 V-measure를 통해 클러스터링 결과의 성능을 평가하였다. K값을 7로 설정하고 클러스터링 했을 때 ARI는 0.9, V-measure는 0.79로 가장 우수한 성능을 보였다.

1. 서 론

폐암 환자들의 치료방침 결정과 예후 결정을 하기 위해서는 최종병기를 정확하게 추정하는 것이 중요하다. 최종병기는 종양크기(T), 림프절 전이(N), 원격전이(M) 세 가지 요소를 포함하는 TNM분류 방법을 기반으로 한다[1].

TNM분류 방법은 미국암연합회(American Joint Committee on Cancer, AJCC)에서 작성한 암 병기 매뉴얼을 국제적으로 많이 사용한다. 주요 암별로 TNM 분류 방법이 상이하지만 폐암 또한 AJCC의 TNM 분류 방법을 기반으로 최종병기를 결정한다. AJCC의 TNM병기에 따른 폐암 환자 최종병기 분류는 표 1과 같다[2].

(표 1) TNM병기 분류 방법과 최종병기

최종병기	TNM 병기
1A	T1 N0 M0
1B	T2 N0 M0
2A	T1 N1 M0
2B	T2 N1 M0
	T3 N0 M0
3A	T3 N1 M0
	T1 N2 M0
	T2 N2 M0
	T3 N2 M0
3B	T4 N0 M0
	T4 N1 M0
	T4 N2 M0
	T1 N3 M0
	T3 N3 M0
	T4 N3 M0
4	Any T, Any N, M1

머신러닝은 비지도학습(Unsupervised learning) 방법 중 하나로 데이터들의 묶음을 클러스터라고 하며 클러스터간의 유사성과 연결법을 사용하여 관계를 분석하는 방법이다. 클러스터간의 유사성은 거리측정법을 사용하며 사전에 클러스터 수 K를 지정하느냐 안하느냐에 따라 계층적 클러스터와 분할 클러스터로 나뉜다. 대표적으로 가장 많이 사용하는 클러스터링 알고리즘으로는 K-Means가 있다[3].

클러스터링의 성능 평가는 분류와 달리 라벨이 없고 답이 정해져있지 않기 때문에 매우 어렵다. 이는 원래 데이터의 클러스터를 알려주는 정답이 있는 경우도 마찬가지다. 클러스터링의 성능 평가가 어렵다 보니 단일기준 보다는 다양한 성능평가 기준을 사용한다. 대표적으로 많이 사용하는 클러스터링 성능평가 기준으로는 ARI(Adjusted Rand Index)와 Homogeneity, Completeness와 V-measure가 있다.

본 논문에서는 화순전남대학교병원 673명(CNUHH 데이터셋)과 [4]의 연구에서 사용한 데이터셋 419명(NSCLC 데이터셋)의 폐암환자 정보를 사용한다. K값을 최종병기에 맞게 변경해 가며 폐암환자들의 T, N, M 병기를 K-Means 알고리즘을 사용하여 클러스터링한다. 다양한 성능평가 기준을 사용하여 성능을 종합적으로 평가하고 두 데이터셋의 클러스터링 결과를 비교분석한다.

본 논문의 순서는 다음과 같다. 2장에서는 K-Means 알고리즘과 실험에 사용한 Scikit Learn 라이브러리 대해 설명 하고, 3장에서는 실험에 사용한 데이터셋 설명, 데이터 전처리 방법과 클러스터링 실험환경 및 실험결과에 대해 비교분석한다. 4장에서는 분석결과를 토대로 결론 및 향후 연구에 대해서 기술한다.

2. 관련 연구

2.1 K-Means

K-Means는 K개의 무작위 중심점을 찍은 후, 중심점에 서 각 점간의 거리의 합을 계산하며 중심점을 이동한다. 각 점간의 거리합이 가장 작을 때 해당 중심점을 기준으로 클러스터링 하는 방법이다. 중심점은 각 클러스터 데이터의 평균값을 위치로 가진다[5]. 점간의 거리 계산시 사용하는 거리측정 방법은 보편적으로 유클리드 거리를 많이 사용한다.

2.2 Scikit Learn

Python 프로그래밍 언어 기반의 머신러닝 라이브러리로 설치, 학습, 사용이 쉽고 분류, 회귀, 클러스터링, 차원축소, 모델선택, 전처리에 대해 다양한 알고리즘을 지원한다. 파이썬 이외의 언어에서는 사용할 수 없고 딥러닝과 강화학습을 지원하지 않는 단점이 있다[6].

3. 실험 및 결과 분석

3.1 실험 데이터셋

실험에 사용한 CNUHH 데이터셋과 NSCLC 데이터셋은 다음 표 2, 표 3과 같다. 본 논문에서는 T, N, M, 최종병기 4개의 컬럼만 사용한다.

(표 2) CNUHH 데이터셋

기간	2005년 ~ 2018년
컬럼 수	17
환자 수	673명
T 병기	T1, T2, T3, T4, Any T
N 병기	N0, N1, N2, N3, Any N
M 병기	M0, M1
최종병기	1A, 1B, 2A, 2B, 3A, 3B, 4

(표 3) NSCLC 데이터셋

컬럼 수	10
환자 수	419명
T 병기	1, 2, 3, 4, 5
N 병기	0, 1, 2, 3, 4
M 병기	0, 1
최종병기	1, 2, 3A, 3B

두 데이터셋 모두 T, N, M, 최종병기를 갖고 있지만 라벨이 조금 상이하다. T, N, M 병기에서 CHUNN는 범주형 타입 사용한 반면 NSCLC는 순서형 타입으로 라벨링 돼있다. 최종병기는 CHUHH 같은 경우 AJCC의 분류에 맞게 7개로 라벨 된 반면 NSCLC는 총 4개로 라벨링 되었고 4기 환자는 없는 것을 확인 할 수 있다. 최종병기별 환자

수는 표 4와 같다.

(표 4) CNUHH와 NSCLC의 최종병기별 환자 수

최종병기		환자 수	
		CNUHH	NSCLC
1	1A	12	93
	1B	24	
2	2A	27	39
	2B	15	
3	3A	116	112
	3B	107	175
4	4	372	0
전체		673	419

CHUHH는 3기와 4기가 전체 데이터의 88%를 차지하고 있으며 NSCLC는 3기가 전체 데이터의 68%를 차지하고 있다. 두 데이터셋 모두 3기 이상의 병의 정도가 심한 환자들이 대부분 분포하고 있는 것을 확인할 수 있다.

3.2 데이터 전처리

실험에 사용할 CNUHH 데이터셋의 T, N, M 병기 컬럼은 T1, N1, M0과 같은 명목형 데이터로 구성되어 있다. Scikit Learn 라이브러리에서 제공하는 K-Means 알고리즘을 사용하기 위해서는 범주형 데이터를 순서형 데이터로 변경해야한다. T, N, M 병기 값들은 병이 심해질수록 숫자가 커짐에 따라 T1, T2, T3, T4, Any T를 1~5, N0, N1, N2, N3, Any N을 0~4, M0과 M1을 0과 1로 변경하여 CNUHH와 NSCLC 데이터셋의 T, N, M 병기 라벨링을 동일하게 전처리 후 실험을 진행한다.

3.3 실험 환경

Scikit Learn 라이브러리의 K-Means 라이브러리를 사용하였고 거리측정은 유클리드거리를 사용하였다. 자세한 실험환경은 표 5와 같다. CNUHH 데이터셋은 AJCC에서 분류한 폐암환자 최종병기 7(1A기, 1B기, 2A기, 2B기, 3A기, 3B기, 4기)개와 최종병기를 기수별로 나눈 4(1기, 2기, 3기, 4기)개로 클러스터링 하기 위해 K값은 4와 7을 사용하였고 NSCLC 데이터셋의 K값은 4(1기, 2기, 3A기, 3B기)와 3(1기, 2기, 3기)을 사용하였다.

(표 5) 클러스터링 실험 환경

머신러닝	클러스터링
알고리즘	K-Means
거리측정	유클리드 거리
라이브러리	Scikit Learn
분석 도구	Python, Jupyter Notebook
K 값	3, 4, 7

3.4 성능평가 척도

비지도 학습방법 중 하나인 클러스터링은 다양한 성능평가 기준을 복합적으로 사용하여 성능을 평가한다. 표 6은 클러스터링 성능평가지 가장 많이 사용하는 ARI(Adjusted Rand Index), Homogeneity, Completeness와 V-measure 성능평가 방법을 설명하고 있다. 표 6의 모든 기준은 0.0부터 1.0의 값을 갖으며 1에 가까울수록 좋은 성능을 뜻한다 [7].

(표 6) 클러스터링 성능평가 기준

ARI(Adjusted Rand Index)	두 클러스터간의 모든 쌍을 비교하며 유사성을 평가한다. Rand Index의 클러스터 수가 많아지면 점수가 높아지는 단점을 확률적으로 조정하여 보완한 값이다.
Homogeneity	각 클러스터마다 단일 클래스의 멤버인 데이터 포인트로만 모였는지를 평가한다.
Completeness	단일 클래스의 멤버인 모든 데이터 포인트가 하나의 클러스터로 모였는지를 평가한다.
V-measure	Homogeneity와 Completeness가 얼마나 성공적으로 충족되었는지 명시적으로 측정하는 엔트로피 기반 평가 방법이다. Homogeneity와 Completeness의 조화평균을 계산한다.

3.5 실험 결과

표 7은 CNUHH와 NSCLC 데이터셋을 사용하여 각각 K-Means 클러스터링을 수행하고 성능을 평가한 결과이다. 모든 항목의 성능평가 결과는 소수점 셋째자리에서 버림 하였다.

NSCLC 데이터셋은 K값 3, 4 모두 클러스터링이 잘 되지 않았다. ARI 값이 각 0.45, 0.36으로 기존 클러스터의 라벨과 예측 클러스터의 라벨의 유사성이 절반도 되지 않는 것을 알 수 있다. 다만 K=3일 때 Homogeneity 점수가 0.74이지만 Completeness가 0.50인걸로 보아 클러스터내에 동일한 라벨의 데이터가 모였지만 클러스터가 여러개로 나누어진 것을 확인할 수 있다.

CNUHH 데이터셋은 K값 4, 7 모두 우수한 성능으로 클러스터링 되었다. Homogeneity와 Completeness 모두 높은 점수를 보이기 때문에 동일한 값들이 하나의 클러스터로 묶였고 해당 클러스터에 이질적인 값이 매우 적은 것을 알 수 있다. 특히 K값이 7인 경우 ARI값이 0.91로 실제 클러스터와 예측 클러스터가 매우 유사하게 군집된 것을 알 수 있다.

(표 7) CNUHH, NSCLC 클러스터링 성능평가

데이터셋	CNUHH		NSCLC	
	K=4	K=7	K=3	K=4
ARI	0.81	0.91	0.45	0.36
Homogeneity	0.79	0.82	0.74	0.50
Completeness	0.69	0.77	0.55	0.47
V-measure	0.74	0.79	0.63	0.49

4. 결론 및 향후 연구

폐암 진단을 받은 환자들에게 올바른 치료와 예후 결정을 통해 수명을 늘리기 위해서는 정확하고 올바른 최종진단이 무엇보다 중요하다. 현재는 환자의 최종병기를 판단할 때 TNM병기를 기준으로 판단하고 있다. 본 논문에서는 데이터과학분야에서 폐암 환자의 TNM병기 데이터셋과 K-Means 알고리즘을 사용하여 실제 TNM병기만으로 최종병기에 맞게 클러스터링 되는지 분석하였다. 성능 측정결과 NSCLC 데이터셋은 대부분의 평가 결과가 좋지 않았다. 반면에 CNUHH 데이터셋은 높은 점수로 최종병기에 맞게 클러스터링되었다. 두 데이터셋 모두 T, N, M병기, 최종병기 및 동일한 라벨을 사용하였고 다른 점은 병기별 데이터의 분포와 환자 수였다. 상이한 결과가 나온 이유는 CNUHH 데이터셋이 TNM병기 및 최종병기가 정교하게 라벨링이 잘 돼있고 양질의 데이터이기 때문에 NSCLC 데이터셋보다 우수한 성능을 보인 것으로 생각된다. TNM병기로 폐암 환자의 최종병기를 판단할 수 있지만 데이터셋이 정교하거나 양질의 데이터를 갖지 못하면 데이터셋에 따라 성능이 유동적으로 변하는 것을 알 수 있다. TNM병기 이외의 폐암환자의 최종병기에 미치는 특성을 찾는 노력이 필요하다.

향후 연구로는 최종병기 기수별로 생존분포가 어떤 특성에 따라 달라지는지 그리고 최근에 많이 연구되고 있는 Radiomics 이미지 파일의 특성을 추가했을 때 클러스터링의 성능이 올라가는지 알아보고자 한다.

Acknowledgements

"이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단 바이오·의료기술개발사업의 지원을 받아 수행된 연구임(NRF-2019M3E5D1A02067961)."

참고문헌

- [1] Dengyuan Wu, Charles Yang, Stephen Wong, Jon Meyerle, Bowu Zhang, Dechang Chen, "An examination of TNM staging of melanoma by a machine learning algorithm", International Conference on Computerized Healthcare. IEEE, 2012.
- [2] 김혜영, "폐암의 병기 결정", 『대한의사협회 대한의사협회지』 제51권 제12호, 2008. pp.1118-1124.
- [3] 홍의석, 박미경, "대표적인 클러스터링 알고리즘을 사용한 비감독형 결함 예측 모델", 『정보처리학회 논문지/소프트웨어 및 데이터 공학』 제3권 제2호, 2014. pp.57-64.
- [4] Aerts, Hugo JW, et al. "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach." Nature communications 5.1, 2014. pp.1-9.
- [5] Kaur, Harjot, and Prince Verma. "Comparative Weka Analysis of Clustering Algorithm's." International Journal of Information Technology and Computer Science 9.8,

2017, pp.56-67.

[6] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." *Journal of machine Learning research* 12, 2011. pp.2825-2830.

[7] Rosenberg, Andrew, and Julia Hirschberg. "V-measure: A conditional entropy-based external cluster evaluation measure." *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*. 2007.