

머신 러닝 알고리즘을 이용한 COVID-19 Risk 분석 및 Safe Activity 지원 시스템

COVID-19 Risk Analytics and Safe Activity Assistant System with Machine Learning Algorithms

전도영¹ 송명호¹ 김수동^{1*}
DoYeong Jeon Myeong Ho Song Soo Dong Kim

요약

최근 COVID-19으로 인하여 전세계적으로 수많은 감염자와 사망자가 발생하였다. 아직까지도 효과적인 COVID-19에 대한 백신의 개발은 성공하지 못한 상태이다. 따라서 사람들은 이 질병의 감염에 크게 우려하고 있다. 그간 정부 공공기관이 제공한 감염 정보는 거의 단순한 합산 및 통계 숫자에 불과하다. 따라서, 개인이나 개인이 있는 장소의 구체적인 위험도는 판단하기 어렵다. 본 논문에서는 머신러닝 알고리즘 기반 COVID-19의 위험도 분석과 안전 활동에 대한 정보 제공에 대한 방법을 제안한다. 이 논문은 COVID-19 감염 및 사망 위험도와 관련된 포괄적인 메트릭 체계를 제안하고, 이를 통해 개인 및 그룹에 대한 위험도를 정량적으로 제공하는 기법을 제시한다. 제시된 시스템은 개인 및 지역 정보와 특성을 반영한 한 클러스터링 알고리즘 등 효과적인 SW 기법들을 활용한다.

☞ 주제어: COVID-19, 머신 러닝, 클러스터링, Risk 분석, Safe Activity 지원

ABSTRACT

COVID-19 has recently impacted the world with the large numbers of infected and deaths. The development of effective COVID-19 vaccine has not been successful. Hence, people have a high concern on the infection of this disease. The infection information from the governmental public organizations are mainly based on simple summary statistics. Consequently, it is hard to assess the infection risks of individual person and the current location of the person. In this paper, we present a machine learning-based software system that analyzes COVID-19 infection risks and guidelines for safe activities. This paper proposes a suite of risk factors regarding COVID-19 infection and deaths and methods to quantitatively measure the individual and group risks using the proposed metrics. The proposed system utilizes a clustering algorithms and various software approaches that reflect the information and features of individuals and their geographical locations.

☞ keyword : COVID-19, Machine Learning, Clustering, Risk Analytics, Safe Activity Assistance

1. 서론

2019년 COVID-19가 전세계적으로 확산된 이후, 팬데믹 (Pandemic) 상황이 지속되고 있다 [1-3]. 각 국가에서는 질병의 감염 및 전파를 막기 위해 마스크 착용, 사회적 거리두기, 격리 등 조치를 취하고 있지만 확진자 수는 증가 추세를 보이고 있다. 최근 여러 국가 및 연구기관에서 백신 연구 성공 사례와 배급 및 접종 관련 일정이 발표되고 있지만 전세계적으로 안정적인 단계에 도달하기까지는

비교적 많은 시간이 소요될 것으로 예상된다.

현재 COVID-19의 위험성에 대해 리스크 [4-7], 감염경로 예측 및 분석 [8-19], 지역사회 및 구성원에 미치는 영향 [20-25] 등 다양한 주제로 연구들이 진행되고 있다. 하지만 COVID-19에 대한 위험도를 수치적으로 표현하거나 안전한 활동을 보장할 수 있는 연구는 드물게 진행되고 있다. 따라서 본 논문에서는 COVID-19의 위험도에 대해 영향을 미치는 요소들을 제시하고 해당 요소들을 통해 위험도를 나타내는 메트릭을 제안한다. 더불어 머신 러닝 알고리즘 중 클러스터링 기법을 이용하여 질병에 대해 환자를 구분하는 연구들 [26-29]을 COVID-19의 위험도 분석에 적용한다.

본 논문에서 제시된 COVID-19 위험도 측정 메트릭 및 분석 방법을 통해 사람들에게 COVID-19로부터 안전한

¹ Dept. of Software, Soongsil University, Seoul, 06978, South Korea.

* Corresponding author: sdkim777@gmail.com

[Received 21 July 2020, Reviewed 5 August 2020(R2 27 October 2020), Accepted 24 November 2020]

생활 활동에 보조하는데 큰 기여를 할 것으로 기대한다.

2. 관련 연구

Alaa의 연구는 인구 밀도를 분석하는 방법으로 COVID-19가 확산될 위험이 있는 구역을 식별하는 전략을 제시하였다 [30]. 하지만 분석 결과를 이용한 해결책을 제시하지 않았다.

Wang의 연구는 지역 사회에서 COVID-19의 출현과 확산을 모델링 및 분석 후 확산의 원인요소 도출하는 방법을 제시하였다 [31]. 하지만 위험요소를 활용하여 COVID-19의 확산을 통제하는 방안에 대해서는 제시하지 않았다.

Lee의 연구는 COVID-19가 현재 한국의 지역 사회에 미친 정도에 대해 분석하고 향후 발전 방향을 제안하였다 [32]. 하지만 관공서 및 언론사 등의 특정 집단에 대한 한정된 방안들을 제시하였다.

마지막으로 Song의 연구는 대화형 웹 기반 매핑 플랫폼을 통해 사회적 거리두기를 강화할 수 있는 방안을 제시하였다. 하지만 사용자들이 COVID-19에 대한 위험도를 직접적으로 느낄 수 있는 방안에 대해 제시되지 않았다 [33].

현재 진행된 연구들을 고려하였을 때 질병의 위험도에 관련된 분석에 대한 연구들이 대다수이다. 또한 분석 결과를 이용하여 사람들이 안전한 행동을 할 수 있도록 제안하거나 판단할 수 있는 연구는 활발히 진행되지 않았다. 따라서 본 논문에서는 COVID-19리스크에 연관된 여러 메트릭들과 이를 이용하여 리스크 분석 및 안전한 활동을 지원할 수 있는 방안을 제시한다.

3. COVID-19 리스크 평가 메트릭

본 절에서는 COVID-19 리스크 평가 및 필요성을 설명하며 COVID-19 리스크 평가를 위한 요소들과 메트릭을 제안한다.

3.1 COVID-19 리스크 평가 정의

COVID-19 리스크 평가는 COVID-19 관련 데이터를 이용하여 사람과 장소에 대한 위험성을 수치화하고 등급을 산정하는 것을 의미한다. 이 평가는 COVID-19의 위험성에 둔감해지는 사회 구성원들에게 경각심의 고취를 기대할 수 있다. 이러한 COVID-19 리스크 평가는 장소에 대한 GSR (Group Safety Risk) 값과 사람에 대한 ISR (Individual Safety Risk) 값, 두 가지로 이루어진다. 다음

절에서 GSR과 ISR을 산출하기 위한 요소 및 메트릭을 제안한다.

3.2 그룹 안전 리스크 (GSR) 평가 메트릭

GSR에서 그룹 (Group) 은 시, 군, 구처럼 행정구역을 의미하는 것뿐만 아니라 건물 등 작은 단위 장소도 포함하며, GSR 값은 이 그룹에 대해 위험한 상태를 나타낸다. GSR값의 범위는 [0, 1]이며 0이면 무결한 장소를 의미하고 1이면 감염성이 매우 높은 위험한 상태를 의미한다.

이러한 GSR값을 산출하기 위한 요소들의 집합을 식 (1)과 같이 나타낸다.

$$GSRFactorSet_{GName} = \{GSRFac_1, GSRFac_2, \dots, GSRFac_n\} \quad (1)$$

예를 들어 어떤 소규모 그룹 g의 GSR값을 구한다고 가정한다. 필요한 요소가 해당 그룹의 면적 및 그룹에 머물렀던 사람들의 정보일 경우, $GSRFactorSet_g = \{VisitSevSet_g, GDensity_g\}$ 과 같이 표현될 수 있다. $GSRFactorSet_g$ 의 첫 번째 요소인 $VisitSevSet_g$ 은 그룹에 있었던 사람들의 위험도를 수치화한 값들의 집합을 나타내며 식 (2)와 같이 나타낸다.

$$VisitSevSet_g = \{severity_i \mid A \text{ severity value of visited person}\} \quad (2)$$

$VisitSevSet_g$ 의 원소인 $severity$ 의 범위는 [0, 1]이다. 0에 가까울수록 COVID-19에 대해 무결하고 1에 가까울수록 전염성이 높다는 것을 의미하며 완치되지 않은 감염자는 1의 값을 가진다. 집합 원소의 수는 방문한 사람 수와 같다. 이러한 $severity$ 값을 산출하기 위한 요소들은 식 (3)와 같으며 COVID-19의 연구결과에 따라 수정될 수 있다.

$$SevFacSet_{pID} = \{SevFac_1, SevFac_2, \dots, SevFac_n\} \quad (3)$$

$SevFacSet_{pID}$ 에서 pID는 식별을 위한 사람의 아이디를 의미한다. $SevFacSet_{pID}$ 의 요소로 감염자 혹은 접촉자와 접촉하고 경과한 날짜와 개인의 연령이라 가정하면 그룹 g를 방문한 사람 x에 대한 표현은 식 (4)와 같다.

$$SevFacSet_x = \{DaysIncurred_x, Age_x\} \quad (4)$$

$DaysIncurred$ 는 감염자 혹은 접촉자와 접촉 후 경과한 날을 의미하며 최대값은 COVID-19의 증상이 발현할 수 있는 최대 기간에 기인한다. Age 는 연령을 의미하며 범위는 [0, 100] 이다. 이 요소들로 $Severity$ 를 계산하는 메트릭

은 식 (5)과 같다.

$$Severity_{pID} = \prod_{n \in SevFacSet_{pID}}^{|SevFacSet_{pID}|} (W_n \times SevFacM_n) \quad (5)$$

$|SevFacSet_{pID}|$ 은 집합 $SevFacSet_{pID}$ 원소의 수를 의미한다. $SevFacM_n$ 는 $SevFacSet_{pID}$ 의 요소 n 이 Severity에 미치는 값을 계산하는 메트릭이다. W_n 는 요소 n 에 대한 가중치를 의미하며 모든 가중치의 합은 1이다. 즉, DaysIncurred에 대해 계산하는 메트릭은 $SevFacM_{DaysIncurred}$ 이며 Age값에 대해 계산하는 메트릭은 $SevFacM_{Age}$ 이다. DaysIncurred와 Age에 대한 가중치는 W_{days} 와 W_{age} 이다. $SevFacM_{DaysIncurred}$ 은 식 (6)과 같다.

$$SevFacM_{DaysIncurred} = 1 - \frac{(DaysIncurred)}{(MaxC19Days)} \quad (6)$$

DaysIncurred는 감염되거나 접촉한 후 경과한 날을 의미하며 MaxC19Days는 COVID-19의 증상이 나타날 수 있는 최대 기간을 의미한다. MaxC19Days는 COVID-19에 대한 연구 결과에 따라 달라질 수 있으며 WHO (World Health Organization)에 의하면 현재 MaxC19Days는 14이다 [34]. $SevFacM_{DaysIncurred}$ 결과의 범위는 [0, 1]이다. 0에 가까울수록 접촉한지 오래되었음을 의미하고 1에 가까울수록 최근에 접촉했음을 의미한다.

$SevFacM_{Age}$ 은 식 (7)과 같다.

$$SevFacM_{Age} = \frac{(Age)}{(MaxLifeLimit)} \quad (7)$$

MaxLifeLimit은 가능한 최대 연령을 의미하며 본 예에서는 식 (5)에서 Age의 범위에 따라 100으로 설정된다. 이러한 $SevFacM_{Age}$ 결과의 범위는 [0, 1]이다. 0에 가까울수록 어리며 1에 가까울수록 늙었음을 의미한다.

예를 들어, 70세의 노인이 7일 전에 감염자와 접촉하고 Age에 가중치를 70%, DaysIncurred에 가중치를 30%를 줄 경우, 해당 노인이 가지는 Severity값은 $0.7 \times 0.7 + 0.3 \times 7/14$ 로 약 0.64 가 된다.

GSRFactorSet_g 두 번째 요소인 GDensity_g는 그룹 내 사람들의 밀집도를 나타낸다. 범위는 [0, 1]이며 0에 가까울수록 공간이 넓거나 사람이 적음을 의미하고 1에 가까울수록 비좁은 공간에 사람이 많다는 것을 의미한다. 예를 들어 20명 정도가 적정인 피트니스 센터에 2배인 40명이 있다면 GDensity_g에 기반한 값은 20명일 때에 비해 높게 나온다.

그룹 g의 GSR값을 구하기 위해 VisitSevSet_g과 GDensity_g에 대한 계산이 선행되어야 한다. 먼저 VisitSevSet_g로 산출되는 값은 VisitSevSet_g 원소의 평균으로 구해지며 그 값은 AvgSeverity_g이다. GDensity_g을 기반으로 한 값 GDensityVal_g 식 (8)을 통해 구해진다.

$$GDensityVal_i = \min\left(\frac{(PeopleN_i) \times (MinSafeArea)}{(GroupTotalArea_i)}, 1\right) \quad (8)$$

PeopleN_i는 그룹 i에 있었던 사람 수를 의미한다. MinSafeArea는 COVID-19의 전염을 막기 위해 사람과 사람 사이에 유지되어야 하는 최소거리를 기반으로 계산한 면적을 의미한다. 이 거리는 COVID-19 연구에 따라 유동적으로 변할 수 있다. GroupTotalArea_i는 그룹 i의 총 면적을 의미한다. min()은 입력 값 중에 최소값을 출력하는 함수이다. 즉, 식 (8)에서 GDensityVal_i의 최대값은 1이 된다.

AvgSeverity와 GDensityVal을 기반으로 GSR값을 구하는 메트릭은 식 (9)과 같다.

$$GSR_{GName} = AvgSeverity_{GName} \times GDensityVal_{GName} \quad (9)$$

AvgSeverity의 범위는 [0,1]이고 GDensity의 범위가 [0,1]이므로 GSR의 범위는 [0,1]이다. 0에 가까울수록 COVID-19의 위험성이 낮고 무결한 장소를 의미하며 1에 가까울수록 고위험군이 있거나 많은 감염자들이 존재했음을 의미한다.

즉, 20명 정도가 적정인 인원인 그룹에 40명이 있고, 그 중 감염자가 2명일 때, SeveritySet은 {0.8, 0.7}이고 AvgSeverity는 0.75이다. 그리고 해당 그룹의 GroupTotalArea가 200, MinSafeArea가 10일 때 GDensityVal은 min() 함수에 의해 1이 된다.

3.3 개인 안전 리스크 (ISR) 평가 메트릭

ISR은 개인이 COVID-19에 대해 위험한 정도를 보여주는 척도다. 결과값의 범위는 [0, 1]이며 0이면 COVID-19에 감염될 확률이 없는 상태를 의미하고 1이면 당장이 라고 감염될 수 있는 상태를 의미한다. ISR은 COVID-19에 대한 개인의 감염 여부나 가능성을 의미하는 Severity와 달리 주변 환경을 고려하여 현재 개인에게 얼마나 큰 위협이 있는지를 의미한다. ISR을 산출하는 데 필요한 요소들은 식 (10)와 같다.

$$ISRFactorSet_{pID} = \{ISRFac_1, ISRFac_2, \dots, ISRFac_n\} \quad (10)$$

COVID-19의 위험성과 감염성에 대한 연구는 현재 진행형이므로 ISR값 산출을 위한 요소는 추가 및 제거될 수 있다. 예를 들어 ISR을 산출하기 위한 요소로 개인의 Severity, 개인과 연관성 높은 그룹 타입, 그리고 주변 그룹에 대한 정보가 있을 경우 사람 y에 대한 ISRFactorSet_y를 식 (11)과 같이 나타낼 수 있다.

$$ISRFactorSet_y = \{Severity_y, RelatedGSet_y, SurroundGSRSet_y\} \quad (11)$$

RelatedGSet_y는 사람 y와 연관성이 있는 그룹의 집합을 의미한다. SurroundGSRSet_y는 개인의 주변 환경에 위치한 GSR의 집합을 의미하며 식 (12)와 같이 나타낼 수 있다.

$$SurroundingGSRSet_{personID} = \{GSR_{GroupName} \mid Surrounding GSR of person\} \quad (12)$$

사람 y의 ISR값을 산출하기 위해서 RelatedGSet_y과 SurroundGSRSet_y를 통해 유의미한 그룹을 선별하는 작업이 선행되어야 한다. 유의미한 그룹이란 SurroundGSRSet_y의 원소 중 그룹 타입이 RelatedGSet_y에 속하는 원소를 의미한다. 이러한 원소를 tGSR이라하고 이러한 원소들의 집합을 tGSRSet_y라하면 사람 y의 tGSRSet_y는 다음과 같다.

$$tGSRSet_y = \{tGSR_1, tGSR_2, \dots, tGSR_k\} \quad (13)$$

이러한 ISRFactorSet_y를 바탕으로 산출되는 ISR값은 식 (14)과 같다.

$$ISR_y = \frac{w_s \cdot Severity_y \times w_g \sum_{i \in tGSRSet_y} \left(\frac{LivingD - tGSRD_i}{LivingD} \times GSR_i \right)}{|tGSRSet_y|} \quad (14)$$

w_s와 w_g는 각각 Severity와 tGSRSet에 대한 가중치를 의미하며 두 가중치의 합은 1이다. |tGSRSet_y|은 집합 tGSRSet_y의 원소의 수를 의미한다. LivingD는 사용자의 생활 반경을 의미한다. tGSRD는 사용자와 tGSR과의 거리를 의미한다. LivingD와 tGSRD로 사용자로부터 거리에 따른 영향력을 조절할 수 있다.

4. 클러스터링 기반 COVID-19 안전활동 알고리즘

본 장에서는 3.2절과 3.3절에서 언급한 요소들과 메트

릭을 통해 나온 결과값으로 클러스터링을 진행한다. 그리고 본 논문에서는 일반 클러스터링 기법과 다른 가중치 기반 클러스터링 기법을 설명한다.

4.1 가중치 기반 클러스터링 기법

가중치 기반 클러스터링 기법이란 클러스터링에서 사용되는 거리함수에 가중치를 추가한 것을 의미한다. 가중치는 클러스터링에 사용되는 특징들에 적용되며 각 특징에 영향력을 조절한다. 이때 각 특징에 적용되는 가중치의 합은 1이다.

예를 들어 유클리드 거리함수를 이용하는 경우, 특징 집합 SetA를 가진 임의의 두 데이터 p, q에 대해 가중치를 부여한 메트릭은 식 (15)와 같다.

$$euc_dis_{pq} = \sqrt{\sum_{i \in SetA} w_i \cdot (i_p - i_q)^2} \quad (15)$$

이러한 가중치 기반 클러스터링 기법의 알고리즘은 표 1과 같다.

(표 1) 가중치 기반 클러스터링 기법 알고리즘
(Table 1) Algorithm of Weight Based Clustering

Algorithm 1. Algorithm of Weight Based Clustering	
Input:	algoName: Algorithm, cN: Number of Clusters, df: Distance Function Method, wList: Weight List of Features
Output:	clusterInstacnce: Instance of Cluster Algorithm
1	function createCluster(algoName, cN, df, wList):
2	newCluster := selectCluster(algoName)
3	newCluster.setFunction(df)
4	if cN inputted:
5	newCluster.setClusterNum(cN)
6	if wList inputted:
7	newCluster.setWeight(wList)
8	return clusterInstacnce

함수에 클러스터 알고리즘 이름, 클러스터의 수, 거리 함수, 가중치 리스트를 전달한다 (Line 1). selectCluster() 함수에 클러스터 알고리즘 이름을 전달하여 사용할 클러스터를 결정한다 (Line 2). 클러스터 알고리즘에 클러스터의 수를 입력해야 할 경우 클러스터의 수를 setClusterNum을 통해 클러스터 인스턴스에 전달한다 (Line 4-6). 그리고 거리 함수에 적용할 가중치 리스트를 입력했다면 setWeight() 함

수를 통해 클러스터 인스턴스에 전달한다 (Line 7~9). 그 후, 클러스터 인스턴스를 반환한다 (Line 10).

4.2 그룹 기반 클러스터링 기법

그룹 기반 클러스터링이란 사람을 이용하여 클러스터링하는 것을 의미한다. 본 절에서는 그룹 기반 클러스터링의 필요성, 알고리즘 및 활용 방안에 대해 설명한다.

4.2.1 그룹 기반 클러스터링의 필요성

현재 CDC (Centers for Disease Control and Prevention) 에 의하면 COVID-19는 사람들 간 호흡기 비말을 통해 전파되지만 [35] WHO에 따르면 특정 조건에서는 공기로 통한 전염이 충분히 가능하다고 경고한다 [36]. 따라서 감염자나 접촉자가 머문 장소 또한 분류될 필요가 있다. 그룹 기반 클러스터링은 이런 그룹에 대한 방역 활동에 일조할 수 있다.

4.2.2 그룹 기반 클러스터링 알고리즘

인구 기반 클러스터링과 마찬가지로 그룹 기반 클러스터링 알고리즘도 특징이 요구되며 사용자는 원하는 종류를 선택할 수 있다. 예를 들어 필수적으로 포함되는 GSR 값을 비롯하여 업종, 매일 방문하는 사람 수 등과 같은 특징을 선택하여 클러스터링을 진행할 수 있다. 예를 들어 GSR값과업종 정보를 가진 데이터로 클러스터링을 진행할 경우 그 결과는 표 2와 같다.

(표 2) 그룹 클러스터링 출력 데이터 예시
(Table 2) Example of Group Clustering Output Data

ID	GSR	BusinessCategory	ClusterID
1	0.33	Bar	1
2	0.89	Bar	2
3	0.09	Restaurant	3
4	0.11	Laundry	3
5	0.01	Apartment	3

입력 데이터에는 다섯개의 데이터가 있으며 각 데이터는 식별을 위해 ID 값을 가지고 있다. 각 데이터는 GSR값과 Business-Category 값을 가지고 있다. GSR값은 3.2절의 GSR값과 동일하다. BusinessCategory은 업종을 의미한다. 이 데이터에 대한 클러스터링 결과는 ClusterID 열과 같다. 해당 열은 그룹 기반 클러스터링을 통해 분류된 클러스터의 ID이다. ID 3, 4, 5는 비교적 낮은 GSR 값으로 3으

로 분류되었으며 ID 1, 2은 GSR값이 높은 편이지만 카테고리 차이가 다르기 때문에 ID 1은 ClusterID 1로 분류되었고 ID 2은 ClusterID 2로 분류되었다.

4.3 인구 기반 클러스터링 기법

인구 기반 클러스터링이란 사람을 이용하여 클러스터링하는 것을 의미한다. 본 절에서는 인구 기반 클러스터링의 필요성, 알고리즘 및 활용 방안에 대해 설명한다.

4.3.1 인구 기반 클러스터링의 필요성

COVID-19 감염의 경우 감염자와의 접촉 뿐 만 아니라 개인의 건강 상태, 나이, 영양 상태 등에 의해 위험도가 달라질 수 있다. 따라서 여러 요소를 고려하여 COVID-19에 대한 위험도를 사람들에게 대해 분류하는 작업이 필요하다. 해당 클러스터링을 통해 막연한 위험이 아닌 개인의 상태를 고려한 위험도 산출이 가능하다.

4.3.2 인구 기반 클러스터링 알고리즘

인구 기반 클러스터링에서 사용자는 원하는 특징을 선택할 수 있다. 예를 들어 필수적으로 포함되는 ISR값을 비롯하여 나이, 성별, 관련 질환, 거주 지역 중 일부를 선택하여 클러스터링 결과를 볼 수 있다.

(표 3) 사람 클러스터링 출력 데이터 예시
(Table 3) Example of People Clustering Output Data

ID	ISR	Age	ClusterID
1	0.01	27	1
2	0.05	43	1
3	0.03	71	2
4	0.65	22	3
5	0.89	22	3

예를 들어 ISR값과 나이 정보를 가진 데이터로 클러스터링을 진행할 경우 그 결과는 표 3와 같다. 입력 데이터에는 다섯개의 데이터가 있으며 각 데이터는 식별을 위해 ID 값을 가지고 있다. 그리고 각 데이터는 ISR값과 Age값을 가지고 있다. ISR값은 3.3절의 ISR 값과 동일하다. Age는 나이를 의미하며 0이상의 정수로 표시된다. 이 데이터에 대한 클러스터링 결과는 ClusterID 열과 같다. 해당 열은 인구 기반 클러스터링을 통해 분류된 클러스터의 ID이다. ID 4, 5는 비교적 높은 ISR 값으로 3으로 분류되었으며 ID 1, 2, 3은 ISR값이 낮지만 Age의 편차가 커

서 ID 1, 2는 ClusterID 1로 분류되었고 ID 3은 ClusterID 2로 분류되었다.

4.4 안전활동을 위한 클러스터링 평가 및 검증 기법

인적 데이터로 안전 활동을 위한 클러스터링 알고리즘을 이용함에 따라 본 절에서 언급한 그룹 기반, 인구 기반 클러스터링의 결과를 평가하고 검증하기 위한 기법이 필요하다. 따라서 본 논문에서는 최적의 클러스터 수를 찾기 위한 기법으로 Elbow 기법과 Silhouette 기법을 사용한다.

Elbow 기법은 한 축이 클러스터의 수인 꺾은선 그래프의 팔꿈치 (Elbow)에 위치한 클러스터의 수를 최적의 클러스터 수로 선택하는 기법이다. 본 논문에서는 WCSS (Within-Cluster Sum of Square) 지표를 사용한다. WCSS 값을 산출하기 위한 메트릭은 다음과 같다.

$$WCSS = \sum_{i=0}^n \min_{\mu_j \in C} (\|x_i - \mu_j\|^2) \quad (16)$$

n 은 데이터의 수를 의미한다. j 는 클러스터의 색인을 의미하며 μ_j 은 클러스터의 중심 좌표를 의미한다. x_i 는 데이터의 좌표를 의미한다. C 는 클러스터의 집합을 의미한다. WCSS의 값의 범위는 $[0, \infty)$ 이며 0에 가까울수록 클러스터 내 데이터들이 잘 응집되어 있다는 것을 의미하며 값이 클수록 데이터들이 클러스터 중심으로부터 더 산재되어 있다는 것을 의미한다.

Silhouette 기법은 각 데이터의 계산된 Silhouette 계수 (Coefficient) 값을 통해 최적의 클러스터 수를 결정하는 기법이다. Silhouette 계수는 클러스터 내의 한 데이터에 대해 가장 근접한 다른 클러스터의 데이터들과 얼마나 멀리 떨어져 있는지를 의미한다. Silhouette 계수를 계산하기 위한 메트릭은 다음과 같다.

$$s_i = \frac{b - a}{\max(a, b)} \quad (17)$$

s_i 는 데이터 i 의 Silhouette 계수를 의미한다. a 는 i 가 속한 클러스터 내 데이터들에 대한 거리의 평균이다. b 는 i 가 속한 클러스터에 가장 근접한 클러스터의 데이터들에 대한 거리의 평균이다. Silhouette 계수의 범위는 $[-1, 1]$ 이며 1에 가까울수록 해당 데이터가 근처 군집과 떨어져 있음을 의미하고 0에 가까울수록 클러스터들의 구분이 명확하지 않은 경계선이 위치하고 있음을 의미한다. -1에 가까울수록 데이터가 현재 속한 클러스터가 아닌 다른

클러스터에 더 근접해 있음을 의미한다.

5. 실험 및 평가

본 절에서는 그룹 기반 및 인구 기반 클러스터링 알고리즘을 적용하여 다양한 시나리오를 제시하고 실험을 진행한다.

5.1 실험 환경

5.1.1 데이터셋 구성

해당 실험에 사용되는 데이터는 사람에 대한 데이터와 지역에 대한 데이터로 구분된다. 사람에 대한 데이터는 10,000명의 인적 데이터로 구성되어 있다. 이 데이터의 속성은 연령, 감염 및 접촉 날짜, COVID-19 Status (감염, 접촉, 무결), 대한민국 내 거주지, 면적 데이터 및 이동경로로 구성된다. 지역에 대한 데이터로는 각 지역의 위치, 지역의 종류 (시, 도, 건물 등), 해당 지역의 GSR 값 등으로 구성되어 있다. 각 데이터에는 ISR 및 GSR 메트릭을 통해 얻어진 ISR 값 및 GSR 값도 포함하고 있다.

5.1.2 사용 클러스터링 구성

본 실험은 제시한 그룹 기반과 인구 기반 클러스터링을 이용하여 아래의 다섯가지 시나리오에 맞춰 진행한다.

- 소규모 시설 단위 위험도 그룹 분류
- 행정 구역 단위 위험도 그룹 분류
- 업종 단위 위험도 그룹 분류
- 선택적 고위험 그룹 분류
- 그룹 변동에 따른 위험 그룹 예측

각 시나리오에 맞게 데이터셋에서 데이터 전처리 후 클러스터링을 적용한다. 이 때 각 클러스터링 알고리즘에서 사용되는 중심 개수는 5개로 설정한다.

5.2 실험 시나리오 적용 및 평가

5.2.1 소규모 시설 단위 위험도 그룹 분류

본 실험은 건물 등의 소규모 시설에 대한 그룹 기반 클러스터링을 적용한다. 해당 시나리오는 작은 단위의 지역의 COVID-19에 대한 위험도를 측정하고 해당 지역을 분류하여 주변 사람들에게 경각심을 줄 수 있다. 본 실험에서는 감염자 이 때 사용되는 메소드는 표 4과 같다.

(표 4) 소규모 시설 그룹 분류 알고리즘
(Table 4) Algorithm for Classifying Facility Group

Algorithm 2. Algorithm for Classifying Facility Group

Input: *ipd*: Infected People Data, *cpd*: Contacted People Data, *fgd*: Facility Group Data, *cluster*: Instance of Cluster

Output: *clusterResult*: List of Cluster ID, *clusterInfo*: Information of Cluster

1	function classifyFacilityGroup(<i>ipd</i> , <i>cpd</i> , <i>fgd</i> , <i>cluster</i>):
2	<i>integratedData</i> := incorporateData(<i>ipd</i> , <i>cpd</i> , <i>fgd</i>)
3	<i>clusterIDList</i> := cluster(<i>integratedData</i>)
4	<i>clusterResult</i> := incorporateData(<i>integratedData</i> , <i>clusterIDList</i>)
5	<i>clusterInfo</i> := extractClusterInfo(<i>clusterResult</i>)
6	return <i>clusterResult</i> , <i>clusterInfo</i>

함수에 입력된 감염자 데이터, 접촉자 데이터, 그리고 소규모 시설 그룹 데이터를 *incorporateData()* 함수로 통합한다 (Line 2). 통합된 데이터로 클러스터링 후 결과로 나온 Cluster ID 리스트까지 통합한다 (Line 3~4). 통합 데이터를 이용하여 클러스터링을 통한 클러스터들의 정보를 추출한다 (Line 5). 그 후, 통합 데이터와 클러스터 정보를 반환한다 (Line 6).

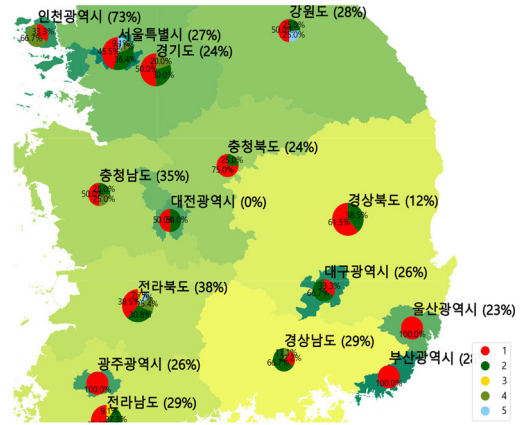
위의 알고리즘을 적용하여 주택가와 대학가가 인접하게 위치한 곳에 대해 클러스터링을 진행하였다. 주택가의 경우, 비교적 낮은 위험도로 클러스터링이 되었고 대학가의 경우 주택가보다 높은 위험도의 클러스터 분류가 되었다. 특히, 많은 사람들이 이용하는 버스와 지하철 정류장 주변은 가장 높은 위험도에 속하는 클러스터로 구분이 되었다.

5.2.2 행정 구역 단위 위험도 그룹 분류

본 실험에서는 5.2.1에서 사용된 범위보다 큰 범위에 대한 위험도 그룹을 측정한다. 해당 실험을 통해 각 행정 구역 별로 위험도를 측정할 수 있고, 각 지역구별로 상황에 맞는 정책을 제시할 수 있다.

해당 실험에서는 모든 사람 데이터를 각 시도별로 그룹을 묶어 사용하였다. 이 데이터를 통한 클러스터링 결과는 Figure 3처럼 표현된다.

(그림 1)의 원그래프의 크기는 해당 그룹의 속한 인적 데이터의 수를 의미하며 각 클러스터에 따라 5가지 색으로 구분된다. 그룹 명에 포함된 퍼센티지는 전체 인적 데이터 중에 감염자 혹은 접촉자 수를 의미한다. 클러스터

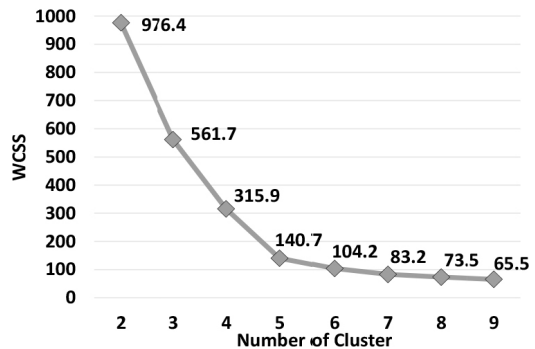


(그림 1) 큰 행정구역 그룹 별 클러스터링 결과
(Figure 1) Result of Clustering by Province Group

색인이 1에서 5로 갈수록 COVID-19에 대해 더 취약한 클러스터를 의미한다.

(그림 1)을 보면 4번 및 5번 클러스터 군의 비율이 많은 강원도, 서울특별시, 인천광역시 등은 타 행정구역보다 COVID-19에 대한 위험도가 높다. 반면 울산광역시, 부산광역시, 광주광역시 등은 위험도가 낮음을 알 수 있다.

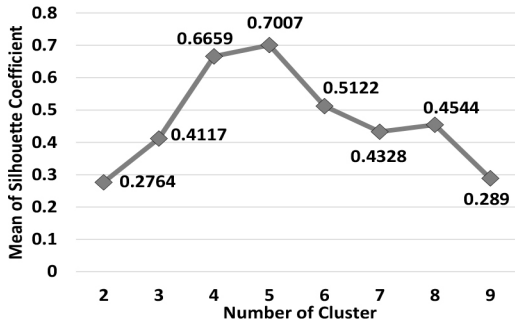
해당 클러스터링 결과의 타당성을 위해 Elbow 기법과 Silhouette 기법을 적용하였다.



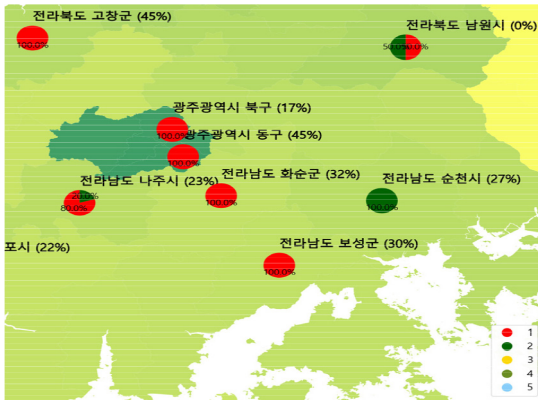
(그림 2) 클러스터 수에 따른 Elbow 기법 결과
(Figure 2) Elbow Result for Different Number of Clusters

클러스터 수가 2에서 5까지 증가하는 동안 WCSS 값이 크게 감소하다가 클러스터 수가 5일 때 WCSS 변화량이 이전에 비해 크게 감소하였다.

그림 3에서 모든 경우를 비교해 볼 때 클러스터 수가 4와 5일 때 평균 Silhouette 계수값이 높다. 따라서 Elbow 기법과 Silhouette 기법을 통한 결과를 고려할 때 최적의 클러스터 수는 5이고 현재 클러스터링된 결과가 타당한 것을 확인할 수 있다.



(그림 3) 클러스터 수에 따른 Silhouette 값
(Figure 3) Silhouette Coefficient Values for Different Number of Clusters



(그림 4) 업종 단위 그룹 별 클러스터링 결과
(Figure 4) Result of Clustering by Small Province Group

5.2.3 업종 단위 위험도 그룹 분류

해당 실험은 각 지역 내 소규모 단위의 그룹들을 업종 별로 묶고 각 업종에 대한 위험도를 분류한다. 본 실험을 통해 각 지역의 해당 업종에 대한 위험도를 판단할 수 있고 사람들에게 해당 업종 그룹에 대한 방문하지 않는 것을 권고할 수 있다.

본 실험에서는 전라남도 주변 지방의 문화 센터에 대

한 데이터를 생성하고 클러스터링을 진행한다. 클러스터링 결과는 그림 4와 같다.

그림 4의 원그래프의 크기는 해당 그룹의 속한 문화센터의 수를 의미하며 면적은 5개의 클러스터에 문화센터의 클러스터에 따라 나뉜다. 그룹 명에 포함된 퍼센티지는 해당 행정 구역에서 위험군에 포함된 사람들의 비율을 의미한다. 클러스터링 ID가 1에 가까울수록 위험도가 낮은 그룹이고, 5에 가까울수록 위험도가 높은 그룹이다.

클러스터링의 결과를 통해 클러스터 2로 분류되는 문화센터가 있는 순천시, 남원시, 나주시의 문화센터는 피하는 것이 좋으며 클러스터 1에 속한 그룹 중에서도 위험군에 속한 사람들의 비율이 가장 낮은 광주광역시 북구 지역의 문화센터가 안전하다는 결론을 얻을 수 있다.

5.2.4 선택적 고위험 그룹 분류

해당 실험은 사용자가 상황에 맞게 필요한 요소에 대해 가중치를 제어하여 클러스터링에 적용하는 실험이다. 본 실험을 통해 각기 다른 사람들의 환경에 맞춰 적합한 클러스터링 결과를 만들어 낼 수 있어 적합한 방법 방법을 적용할 수 있다. 이를 위해 표 5의 알고리즘을 적용한다.

(표 5) ISR 값 데이터로 분류하기 위한 알고리즘
(Table 5) Algorithm to classify by ISR value data

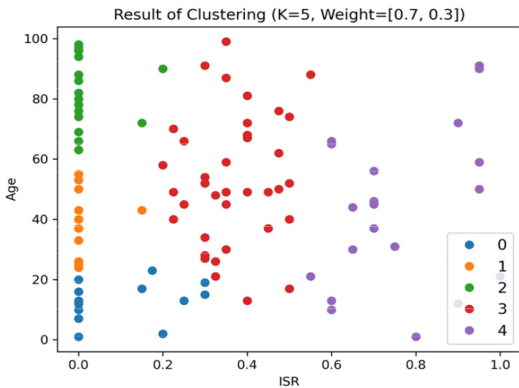
Algorithm 3. Algorithm to Classify by ISR Value Data	
Input:	<i>wList</i> : Weights for Features, <i>algoName</i> : Algorithm, <i>cN</i> : Number of Clusters, <i>df</i> : Distance Function, <i>ipd</i> : Infected People, <i>cpd</i> : Contacted People, <i>upd</i> : User Data, <i>gd</i> : Group Data, <i>weightedCluster</i> : Cluster Instance
Output:	<i>clusterResult</i> : List of Cluster ID, <i>clusterInfo</i> : Information of Cluster
1	function classifyIndividualStatus(<i>wList</i> , <i>algoName</i> , <i>cN</i> , <i>df</i> , <i>ipd</i> , <i>cpd</i> , <i>upd</i> , <i>gd</i>):
2	<i>weightedCluster</i> := createCluster(<i>wList</i> , <i>algoName</i> , <i>cN</i> , <i>df</i>)
3	<i>integratedData</i> := incorporateData(<i>ipd</i> , <i>cpd</i> , <i>upd</i> , <i>gd</i>)
4	<i>clusterResult</i> := <i>weightedCluster</i> (<i>integratedData</i>)
5	<i>clusterResult</i> := incorporateData(<i>integratedData</i> , <i>clusterResult</i>)
6	<i>clusterInfo</i> := extractClusterInfo(<i>clusterResult</i>)
7	return <i>clusterResult</i> , <i>clusterInfo</i>

가중치 리스트, 클러스터 알고리즘, 클러스터 수, 거리 함수를 매개변수로 가중치가 부여된 클러스터 인스턴스

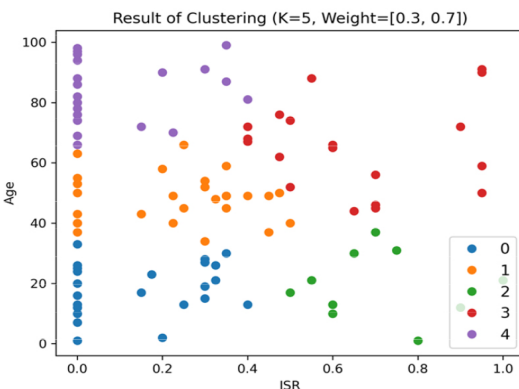
를 생성한다 (Line 2). 함수에 입력된 감염자 데이터, 접촉자 데이터, 그리고 소규모 시설 그룹 데이터를 incorporateData() 함수로 통합한다 (Line 3). 통합된 데이터로 클러스터링 후 결과로 나온 Cluster ID 리스트까지 통합한다 (Line 4~5). 통합 데이터를 이용하여 클러스터링을 통한 클러스터들의 정보를 추출한다 (Line 6). 그 후, 통합 데이터와 클러스터 정보를 반환한다 (Line 7).

해당 알고리즘을 이용하여 ISR값과 연령에 대해 다르게 가중치를 준 후 클러스터링을 진행하였다. 그림 5는 ISR에 0.7, 연령에 0.3을 부여하여 클러스터링한 결과를, 그림 6은 ISR에 0.3, 연령에 0.7을 부여하여 클러스터링한 결과를 나타낸다.

해당 클러스터링 결과는 클러스터의 ID가 높아질수록 위험도가 증가한다. 그림 5에서는 클러스터링이 ISR에 민



(그림 5) 사람 클러스터링 결과 (ISR: 0.7, 연령: 0.3)
(Figure 5) People Clustering Result (ISR:0.7, Age:0.3)



(그림 6) 사람 클러스터링 결과 (ISR: 0.3, 연령: 0.7)
(Figure 6) People Clustering Result (ISR:0.3, Age:0.7)

감하게 클러스터링을 구성한 모습을 보여주며, 반대로 그림 6에서는 Age에 민감하게 클러스터링을 구성한 모습을 보여준다.

5.2.5 그룹 변동에 따른 위험 그룹 예측

해당 실험은 사용자가 새로운 그룹으로 이동하였을 때 발생할 수 있는 해당 그룹에 대한 위험도 측정을 위한 실험이다. COVID-19 관련 데이터를 이용한 그룹 기반 클러스터링과 인구 기반 클러스터링의 특징을 이용하여 지역 사회 구성원의 미래 행동에 대한 결과를 예측하여 보여줄 수 있다. 해당 실험에서는 표 6의 알고리즘을 이용한다.

(표 6) 그룹 변동에 따른 ISR 예측 알고리즘
(Table 6) Algorithm for ISR Changing Group

Algorithm 4. Algorithm to predict ISR when moving to another group	
Input:	uISR: User ISR Data, rgd: Current Group Data, vg: Group Data to Visit
Output:	fISR: Predicted ISR Value, fUserClusterID: Predicted User Cluster ID, fGroupClusterIDList: Changed Groups of Future Cluster ID of Group User Inhabits
1	function predictISRValues(uISR, rgd, vg):
2	fISR := computefISR(uISR, vg)
3	fUserClusterID := computefClusterID(fISR, rgd)
4	for aGroup in range number of vg:
5	fGCID := computefGCID(fISR, aGroup)
6	if fGCID ≠ current GroupClusterID:
7	fGroupClusterIDList.append(aGroup, fGCID)
8	return fISR, fUserClusterID, fGroupClusterIDList

함수에 개인의 ISR 데이터, 개인 거주지 주변의 그룹 데이터, 그리고 방문할 곳의 그룹데이터를 입력한다 (Line 1). 한 개인이 방문할 그룹의 GSR에 따라 현재 ISR을 기반으로 새 ISR을 계산한다 (Line 2). 변경된 ISR 값에 따라 갱신될 개인의 클러스터 ID를 계산한다 (Line 3). 변경된 ISR값 및 개인의 클러스터에 따라 갱신되는 개인 거주지 주변의 그룹 정보와 클러스터 ID를 추출한다 (Line 4~9). 이때 변경되지 않는 그룹의 정보는 추출하지 않는다 (Line 6). 그 후, 갱신될 개인의 ISR값, 클러스터 ID, 그리고 거주지 주변 그룹의 클러스터 ID 리스트를 반환한다 (Line 10).

해당 알고리즘을 이용하여 ISR값이 낮고 지방에 거주하는 한 사람이 유동인구가 많은 지역으로 이동한다 가

정하였을 때 해당 지역의 클러스터링 결과는 변하지 않았지만 해당 사람의 ISR값이 증가하는 것을 확인할 수 있었다. 또한 감염자가 낮은 위험도를 나타내는 지역으로 이동한다 가정하여 진행하였을 때, 해당 사람에 대한 ISR 값은 변하지 않았지만 해당 지역에 대한 위험도 클러스터링 결과는 높은 쪽으로 분류되었다.

6. 결 론

COVID-19의 확산으로 인해 수많은 사람들의 삶의 질이 크게 저하되었을 뿐만 아니라 목숨까지 위협한 상황에 처해있다. 본 논문에서는 COVID-19에 대해 영향을 미치는 요소들을 이용하여 수치적으로 위험도를 나타내는 매트릭들을 제시하였다. 해당 매트릭들을 통해 주변의 건물 뿐 만 아니라 도시, 국가 등의 단위에서 해당 지역의 위험도와 한 개인의 처해진 상황과 해당 사람의 정보를 기반으로 한 위험도를 계산할 수 있음을 보였다.

더불어 클러스터링 알고리즘을 통해 COVID-19의 위험도를 인구 기반과 그룹 기반의 클러스터링 방식을 제안하였다. 또한 이를 이용하여 사람들에게 직접적으로 영향을 줄 수 있는 활용 방안 5가지를 제시하였다.

본 논문의 연구를 통해 실제 사람들에게 COVID-19의 위험에 대한 경각심을 주고 위험도를 평가하고 안전한 곳을 찾도록 하는 사람들의 요구를 충족시킬 수 있으며 지역 사회의 방역에 기여할 수 있다.

References

- [1] Joseph T. Wu, Kathy Leung, et al., "Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China," *Nature Medicine*, Vol.26, No.10, pp.506-510, March 2020.
<https://doi.org/10.1038/s41591-020-0822-7>
- [2] CDC COVID-19 Response Team, "Severe outcomes among patients with coronavirus disease 2019 (COVID-19) — United States, February 12–March 16, 2020," *CDC Stacks*, March 2020.
<http://dx.doi.org/10.15585/mmwr.mm6912e2>
- [3] Eunha Shim, Amna Tariq, et al., "Transmission potential and severity of COVID-19 in South Korea," *International Journal of Infectious Diseases*, Vol.93, pp.339-334, April 2020.
<https://doi.org/10.1016/j.ijid.2020.03.031>
- [4] G. Stewart, K. Heusden and G. A. Dumont, "How control theory can help us control Covid-19," *IEEE Spectrum*, Vol.57, No.6, pp.22-29, June 2020.
<https://doi.org/10.1109/MSPEC.2020.9099929>
- [5] M. Jain, P. K. Bhati, et al., "Modelling Logistic Growth Model for COVID-19 Pandemic in India," in *Proceedings of 5th International Conference on Communication and Electronics Systems (ICCES 2020)*, Coimbatore, India, pp.784-789, July 2020.
<https://doi.org/10.1109/ICCES48766.2020.9138049>
- [6] B. Wang, Y. Sun, et al., "Risk-Aware Identification of Highly Suspected COVID-19 Cases in Social IoT: A Joint Graph Theory and Reinforcement Learning Approach," *IEEE Access*, Vol.8, pp.115655-115661, June 2020.
<https://doi.org/10.1109/ACCESS.2020.3003750>
- [7] Pakpour, A.H. and Griffiths, M.D., "The fear of COVID-19 and its role in preventive behaviors," *Journal of Concurrent Disorders*, Vol.2, No.1, pp.58-63, April 2020.
<https://concurrentdisorders.ca/2020/04/03/the-fear-of-covid-19-and-its-role-in-preventive-behaviors/>
- [8] V. Chamola, V. Hassija, et al., "A Comprehensive Review of the COVID-19 Pandemic and the Role of IoT, Drones, AI, Blockchain, and 5G in Managing its Impact," *IEEE Access*, Vol.8, pp.90225-90265, May 2020.
<https://doi.org/10.1109/ACCESS.2020.2992341>
- [9] M. Abdel-Basset, R. Mohamed, et al., "A Hybrid COVID-19 Detection Model Using an Improved Marine Predators Algorithm and a Ranking-Based Diversity Reduction Strategy," *IEEE Access*, Vol.8, pp.79521-79540, April 2020.
<https://doi.org/10.1109/ACCESS.2020.2990893>
- [10] A. K. Nandi, "Data Modeling With Polynomial Representations and Autoregressive Time-Series Representations, and Their Connections," *IEEE Access*, Vol.8, pp.110412-110424, June 2020.
<https://doi.org/10.1109/ACCESS.2020.3000860>
- [11] E. Hernández-Orallo, P. Manzoni, et al., "Evaluating How Smartphone Contact Tracing Technology Can Reduce the Spread of Infectious Diseases: The Case of

- COVID-19," *IEEE Access*, Vol.8, pp.99083-99097, May 2020.
<https://doi.org/10.1109/ACCESS.2020.2998042>
- [12] E. Montes-Orozco et al., "Identification of COVID-19 Spreaders Using Multiplex Networks Approach," *IEEE Access*, Vol.8, pp.122874-122883, July 2020.
<https://doi.org/10.1109/ACCESS.2020.3007726>
- [13] M. Small and D. Cavanagh, "Modelling Strong Control Measures for Epidemic Propagation With Networks—A COVID-19 Case Study," *IEEE Access*, Vol.8, pp.109719-109731, June 2020.
<https://doi.org/10.1109/ACCESS.2020.3001298>
- [14] S. S. Arun and G. Neelakanta Iyer, "On the Analysis of COVID19 - Novel Corona Viral Disease Pandemic Spread Data Using Machine Learning Techniques," *4th International Conference on Intelligent Computing and Control Systems (ICICCS 2020)*, Madurai, India, pp.1222-1227, May 2020.
<https://doi.org/10.1109/ICICCS48265.2020.9121027>
- [15] N. Zheng et al., "Predicting COVID-19 in China Using Hybrid AI Model," *IEEE Transactions on Cybernetics*, Vol.50, No.7, pp.2891-2904, July 2020.
<https://doi.org/10.1109/TCYB.2020.2990162>
- [16] Wei Xu and Chongyang Chen, "Research on the Influencing Factors and Management Countermeasures of College Students' Sense of Security under the Environment of Big Data-an Empirical Analysis based on the Event of COVID-19," in *Proceedings of the 2020 The 3rd International Conference on Big Data and Education (ICBDE 2020)*, Machinery, New York, USA, p21-25, April 2020.
<https://doi.org/10.1145/3396452.3396465>
- [17] K. El Emam, "Seven Ways to Evaluate the Utility of Synthetic Data," *IEEE Security & Privacy*, Vol.18, No.4, pp.56-59, July 2020.
<https://doi.org/10.1109/MSEC.2020.2992821>
- [18] Soon Ae Chun, Alen Chih-Yuan Li, et al., "Tracking Citizen's Concerns during COVID-19 Pandemic" in *Proceedings of 21st Annual International Conference on Digital Government Research (dg.o 2020)*, New York, USA, p322-323. June 2020.
<https://doi.org/10.1145/3396956.3397000>
- [19] A. Ashok, M. Guruprasad, et, al., "A Machine Learning Approach for Disease Surveillance and Visualization using Twitter Data," in *Proceedings of International Conference on Computational Intelligence in Data Science (ICCIDS 2019)*, Chennai, India, pp.1-6, Feb. 2019.
<https://doi.org/10.1109/ICCIDS.2019.8862087>
- [20] R. B. Duffey and E. Zio, "Analysing Recovery From Pandemics by Learning Theory: The Case of CoVid-19," *IEEE Access*, Vol.8, pp.110789-110795, June 2020.
<https://doi.org/10.1109/ACCESS.2020.3001344>
- [21] Andrea Remuzzi, Giuseppe Remuzzi, "COVID-19 and Italy: what next?," *The Lancet Haematology*, Vol.7, No.5, pp.1225-1228, May 2020.
[https://doi.org/10.1016/S0140-6736\(20\)30627-9](https://doi.org/10.1016/S0140-6736(20)30627-9)
- [22] P. Staszkiwicz, I. Chomiak-Orsa and I. Staszkiwicz, "Dynamics of the COVID-19 Contagion and Mortality: Country Factors, Social Media, and Market Response Evidence From a Global Panel Analysis," *IEEE Access*, Vol.8, pp.106009-106022, June 2020.
<https://doi.org/10.1109/ACCESS.2020.2999614>
- [23] A. Khattar, P. R. Jain and S. M. K. Quadri, "Effects of the Disastrous Pandemic COVID 19 on Learning Styles, Activities and Mental Health of Young Indian Students - A Machine Learning Approach," in *Proceedings of 4th International Conference on Intelligent Computing and Control Systems (ICICCS 2020)*, Madurai, India, pp.1190-1195, May 2020.
<https://doi.org/10.1109/ICICCS48265.2020.9120955>
- [24] R. F. Sear et al., "Quantifying COVID-19 Content in the Online Health Opinion War Using Machine Learning," *IEEE Access*, Vol.8, pp.91886-91893, May 2020.
<https://doi.org/10.1109/ACCESS.2020.2993967>
- [25] S. Greenstein, "Uncomfortable Economic Waters," *IEEE Micro*, Vol.40, No.4, pp.134-136, July. 2020.
<https://doi.org/10.1109/MM.2020.3001464>
- [26] El-Atem N, Irvine KM, Valery PC, et al. "Identifying areas of need relative to liver disease: geographic clustering within a health service district," *Australian Health Review*, Vol.41, No.4, pp.407-418, Aug. 2017.
<https://doi.org/10.1071/AH15225>

- [27] Poon, A. F. Y., "Impacts and shortcomings of genetic clustering methods for infectious disease outbreaks," *Virus Evolution*, Vol.2, No.2, pp.1-9, July 2016.
<https://doi.org/10.1093/ve/vew031>
- [28] Athanasios Tsanas and Siddharth Arora, "Large-scale Clustering of People Diagnosed with Parkinson's Disease using Acoustic Analysis of Sustained Vowels: Findings in the Parkinson's Voice Initiative Study," in *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies*, Valletta, Malta, pp.369-376, Jan. 2020.
<https://doi.org/10.5220/0009361203690376>
- [29] L. Li, S. Xu, S. Wang and X. Ma, "The Diseases Clustering for Multi-source Medical Sets," in *Proceedings of International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI 2016)*, Beijing, China, pp.294-298, Oct. 2016.
<https://doi.org/10.1109/IIKI.2016.37>
- [30] A. A. R. Alsaedy and E. K. P. Chong, "Detecting Regions At Risk for Spreading COVID-19 Using Existing Cellular Wireless Network Functionalities," *IEEE Open Journal of Engineering in Medicine and Biology*, Vol.1, pp.187-189, June 2020.
<https://doi.org/10.1109/OJEMB.2020.3002447>
- [31] B. Wang, S. Xu and M. Mansouri, "Modeling the emergence of COVID-19: a systems approach," in *Proceedings of IEEE 15th International Conference of System of Systems Engineering (SoSE)*, Budapest, Hungary, pp.445-450, June 2020.
<https://doi.org/10.1109/SoSE50414.2020.9130555>
- [32] Moo-Sik Lee, "Overcoming the COVID-19 Epidemics with Communities in Korea," *Journal of agricultural medicine and community health*, Vol.45, No.1, pp.41-46, March 2020.
<https://doi.org/10.5393/JAMCH.2020.45.1.041>
- [33] Song Gao, Jinmeng Rao, et al., "Mapping county-level mobility pattern changes in the United States in response to COVID-19," *SIGSPATIAL Special*, Vol.12, No.1, pp.16-26, March 2020.
<https://doi.org/10.1145/3404820.3404824>
- [34] World Health Organization, "Considerations for quarantine of contacts of COVID-19 cases: interim guidance," World Health Organization, August 19, 2020.
<https://apps.who.int/iris/handle/10665/333901>
- [35] "How to Protect Yourself & Others | CDC," Centers for Disease Control and Prevention. last modified April 24, 2020, accessed July 21, 2020,
<https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html>
- [36] "Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations," World Health Organization. last modified 29 March, 2020, accessed July 21, 2020,
<https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations>

❶ 저 자 소개 ❶



전 도 영

2020년 연세대학교 원주캠퍼스 컴퓨터공학과 졸업 (학사)
2020년~현재 숭실대학교 소프트웨어학과 석사과정
관심분야: Reinforcement Learning-based Intelligent SW System, SW Design
e-mail: corea940@gmail.com



송 명 호

2017년 연세대학교 원주캠퍼스 컴퓨터공학과 졸업 (학사)
2018년~2020년 숭실대학교 컴퓨터학과 졸업 (석사)
2020년~현재 숭실대학교 소프트웨어학과 박사과정
관심분야: 머신 러닝 기반 자율 제어 시스템 (Machine Learning based Autonomous System),
의료 영상 분석 시스템 (Medical Image Analytics System)
e-mail: songmho@gmail.com



김 수 동

1984년 Northeast Missouri State University 전산학 (학사)
1988년/1991년 The University of Iowa 전산학 (석사박사)
1991년~1993년 한국통신 연구개발단 선임연구원
1994년~1995년 현대전자 소프트웨어연구소 책임연구원
1995년 9월~현재 숭실대학교 소프트웨어학부 교수
관심분야: SW Architecture, Machine Learning Systems, Medical Analytics Software
e-mail: sdkim777@gmail.com