|  |
| --- |
| COVID-19 Risk Analytics and Safe Activity Assistant with Clustering and Classification Algorithms  ABSTRACT  (abstract)  Keywords:  클러스터링 및 분류 알고리즘을 이용한 COVID-19 위험 분석 및 안전 활동 보조  요 약  요약 내용  키워드: |

# 서 론

# 관련 연구

관련 연구 내용

# COVID-19 안전 위험 평가의 설계

COVID-19 안전 위험 평가는 COVID-19 데이터를 이용하여 감염자, 접촉자(감염자와 접촉한 사람) 혹은 장소에 대한 위험성을 계산하고 안전성에 대해 평가하는 것을 의미한다. 이 평가의 목적은 사회에서 완화되어가는 COVID-19 전염병의 위험도를 수치화하여 보여줌으로써 개인이 사회의 일원으로서 놓치고 있었던 막연한 경각심을 일깨워 방어적인 행동을 더 유발하게 해 공익에 좀 더 기여할 수 있는데 있다.

COVID-19 안전 위험 평가는 두 가지 방식으로 산출할 수 있다. GSR(Group Safety Risk, 집단 안전 위험)과 ISR(Individual Safety Risk, 개인 안전 위험)이다. 그리고 이 두 방식에 공통적으로 들어가는 요소인 Severity(위험도)가 필요하다. 이번 장의 하위 항목에서 GSR과 ISR을 산출하기 위한 요소 및 수식을 제안한다.

## 안전 위험 평가 요소

COVID-19 안전 위험 평가를 계산하기 위해 필요한 요소는 크게 감염자나 접촉자의 Severity 계산을 위한 요소와 GSR 값 계산을 위한 요소 그리고 ISR값 계산을 위한 요소로 나뉜다.

### Severity의 요소

Severity는 감염자 혹은 접촉자가 얼마나 위험한지 보여주는 척도이다. 이러한 Severity값을 산출하기 위한 요소는 아래와 같다.

현재 COVID-19의 위험성과 감염성에 대해 규명하고 분석하고 있는 중이므로 GSR값 산출을 위한 요소는 추가 및 제거될 수 있다.

*IncurredDate*는 특정 인물이 감염된 날짜를 의미한다.

### GSR의 요소

GSR에서 Group은 시, 군, 구처럼 행정구역을 의미하는 것뿐만 아니라 건물 혹은 편의시설처럼 국소적인 장소를 의미한다. 즉 GSR은 해당 Group이 얼마나 위험한 상태인지 보여주는 척도다. GSR값의 범위는 [0, 1]이며 0이면 무결한 장소를 의미하고 1이면 감염성이 매우 높은 위험한 상태를 의미한다.

이러한 GSR값을 산출하기 위한 요소는 아래와 같다.

현재 COVID-19의 위험성과 감염성에 대해 규명하고 분석하고 있는 중이므로 GSR값 산출을 위한 요소는 추가 및 제거될 수 있다.

*SeveritySet*은 Group, 즉 지역, 건물 혹은 편의시설에 있었던 사람들의 위험도의 집합을 나타낸다. *SeveritySet*의 원소인 *severityi*의 범주는 [0, 1]이며 0이면 건강한 상태를 의미하며 1에 가까울수록 그 사람이 위험하다는 것을 의미한다. *Density*는 해당 지역, 건물 혹은 편의시설의 사람들의 밀집도를 나타낸다. 범주는 [0, 1]이며 0에 가까울수록 공간이 넓거나 사람이 적음을 의미하고 1에 가까울수록 비좁은 공간에 사람이 많다는 것을 의미한다.

### ISR의 요소

ISR은 개인이 현재 얼마나 위험한 상태인지 보여주는 척도다. 결과값의 범위는 [0, 1]이며 0이면 무결한 상태를 의미하고 1이면 매우 위험한 상태를 의미한다. ISR을 산출하는 데 필요한 요소들은 아래와 같다.

현재 COVID-19의 위험성과 감염성에 대해 규명하고 분석하고 있는 중이므로 ISR값 산출을 위한 요소는 추가 및 제거될 수 있다.

*Severityindivisual*은 측정하고 하는 사용자의 Severity를 의미한다. *Severityindivisual*의 범주는 [0, 1]이며 0이면 건강한 상태를 의미하며 1에 가까울수록 그 사람이 위험하다는 것을 의미한다. *RelatedPlaceSet*은 개인과 연관성이 깊은 Group의 집합을 의미하며 원소인 *rPlace*는 구체적인 특정 장소를 의미한다. 일례로 술집의 코로나 확산은 현지 직장인보다 현지 중고등학생에게 더 적게 영향을 끼칠 것이며, 대구의 코로나 확산은 태백의 시민보다 서울 시민에게 더 민감하게 작용할 것이다. *SurroundingGSRSet은* 개인의 주변 환경에 위치한 GSR의 집합을 의미하며 원소인 *GSRgroupName­*은 *groupName*의 *GSR*값을 의미한다. 이때 가능한 *groupName*의 범주는 거주지 주변 환경과 현재 위치한 장소의 주변 환경 정보를 포함한다.

## Metrics for Safety Risk

### Metric for Severtiypeoele

### Metric for GSR

### Metric for ISR

마지막으로 COVID-19 안전 위험 평가에 사용될 요소들에 적용할 가중치가 있다. 클러스터링 알고리즘을 사용할 때 거리 함수(Distance Function)를 사용하게 된다. 이때 거리 함수에 각 요소별로 가중치를 다르게 줘서 COVID-19 안전 위험 평가에 요소가 미치는 영향력을 조절할 수 있다. 또한 각 가중치의 합을 1로 고정함으로써 요소가 미치는 영향력의 비율을 조절할 수 있다.

Weight 값들 고려한 계산 공식 추가

# Clustering of People and Groups

## Clustering People using Factors

클러스터링이란 거리 함수를 이용하여 주어진 데이터들 간의 거리를 계산하고 여러 개의 클러스터를 생성하여 데이터를 클러스터별로 분류하는 것을 의미한다. 즉, ISR을 이용한 클러스터링이란 ISR Factor Set을 이용하여 클러스터링하는 것을 의미한다.

People Clustering 필요성

Clustering Algorithm

Clustering Results

Clustering 결과의 활용

## Clustering Groups using Factors

Group Clustering 필요성

Clustering Algorithm

Clustering Results

Clustering 결과의 활용

# Design of Safe Activity Assistant

Query processing

….

….

# Experiments and Assessment

# Conclusion

References

1. …
2. …
3. …