

2024년도

융복합지식학회

추계학술대회 논문집

THE SOCIETY OF CONVERGENCE KNOWLEDGE

- 일 시 : 2024년 11월 15일(금) 12:00 ~ 18:00
- 장 소 : 성결대학교 학술정보관 6층 아림국제회의실
- 주 최 : (사)융복합지식학회
- 주 관 : 성결대학교, 융복합사업단, 산업통상자원부
- 후 원 : (주)키삭, (주)아이티콤, (주)KES, 이오코퍼레이션,
(주)OpenLinkSystem, (사)한국노벨과학상포럼,
(주)솔빛아이텍, 서경대학교 환경보건센터,
법무법인 유스티아, (주)뷰티코스메틱사이언스
(주)EKM



(사)융복합지식학회

The Society of Convergence Knowledge

목 차

[구두 세션]

- **클리디안 거리 공식을 이용한 2D 게임 속 AI 봇 연구**
..... 임재선, 김찬영, 박영훈, 박지호, 신현지, 유승민, 최유진, 한동연, *박승균 (신한대학교) ▶ 1
- **비콘 센서 기반 실내 전시시설의 방문객 편의성 증대 시스템 설계**
..... 김유진, 원서연, *서지훈 (강남대학교) ▶ 4
- **객체인식을 활용한 개문사고 방지 시스템 에 관한 연구**
..... 김서희, 최민준, 송효은, 안제홍, 강지훈, *이동규 (신한대학교) ▶ 9
- **IoT를 활용한 인텔리전트 빌딩 기반의 실내 조명 제어 플랫폼 개발**
..... 송채빈, 오영은, 정지혜, *서지훈 (강남대학교) ▶ 12

[포스터 세션]

- **텍스트 마이닝을 통한 소비자의 ESG 인식과 접근성 연구: 점자 표기를 중심으로**
..... 김준환 (서울시립대학교), *임석진 (성결대학교) ▶ 17
- **인간 혈청 프로테오믹스를 이용한 근감소증 진단을 위한 SI를 활용한 바이오마커 개발**
..... 노시형, 임동욱 (원광대학교 근감소증토달솔루션센터), *정창원 (원광대학교병원) ▶ 20
- **디지털 디바이스 활용 능력에 따른**
..... *임석진 (성결대학교) ▶ 24
- **프라이버시 인식에 관한 연구**
..... 윤채연, *문석재 (광운대학교) ▶ 24
- **분산 환경에서 메타데이터 관리를 위한 MapReduce 기반의 Fabricator**
..... 유동환 , 홍정민, 강민석, *한석준 (신한대학교) ▶ 26
- **처방전 자동 입력을 위한 복용 관리 앱 설계**
..... *임석진 (성결대학교) ▶ 28
- **암호화 기반 하이브리드 데이터 방송을 위한 프레임워크 설계**
..... 이가빈, *문석재 (광운대학교) ▶ 31
- **빅데이터 기반의 공공인프라 공간정보 플랫폼 개발**

..... 서지훈, 김유진, 원서연 (강남대학교), *주길홍 (경인교육대학교)	▶ 35
▪ 근감소증 표준 데이터 관리 플랫폼 개발	
... 임동욱, 노시형 (원광대학교 근감소증토탈솔루션센터), 김지희, *정창원 (원광대학교 병원)	▶ 39
▪ 사서의 인공지능 교육을 위한 학습 과정 정보 추천 시스템 설계	
..... *서지훈 (강남대학교), 황태준 (인천대학교), 배성민 (강남대학교), 정환웅 (강남대학교)	▶ 41
▪ 딥러닝을 활용한 메타유전체 분류 도구의 정확도 향상	
..... 장세인, *문석재 (광운대학교)	▶ 44
▪ 소셜 네트워크 데이터 기반의 트렌드룩에 대한 오피니언마이닝 분석	
..... 송채빈, 이서영, *서지훈 (강남대학교)	▶ 47
▪ 대형 생활 폐기물 분류 애플리케이션	
..... 홍기동 , 김도영, 박태영, *김유희 (신한대학교)	▶ 50
▪ 분산 클라우드 환경에서 딥러닝을 위한 파라미터 시스템에 관한 연구	
..... 김수연, *문석재 (광운대학교)	▶ 54
▪ 삼킴장애 환자의 흡인과 침투 검출 알고리즘 개발	
..... 임동욱, 노시형 (원광대학교 근감소증토탈솔루션센터), *문희경 (원광대학교)	▶ 56
▪ 데이터 패브릭 아키텍처를 통한 빅데이터 관리 및 분석 전략 개선	
..... 이사랑, 윤채연, *문석재 (광운대학교)	▶ 58
▪ 원달러 환율 변동성 예측모형 추정 연구	
..... *안창호 (서경대학교)	▶ 60
▪ 뉴욕시 택시 데이터 분석을 통한 승객 수요 및 요금 특성 탐색	
..... 신지은, *문석재 (광운대학교)	▶ 62
▪ 체험형 힐링프로그램이 대학생 마음건강에 미치는 영향	
..... - 이해정, 한덕관 (원광대학교)	▶ 64
▪ 류마티스 관절염 조기진단을 위한 관절손상 평가를 위한 인공지능	
정종혁 이창훈 이명수 (원광대학교 병원), 임동욱 노시형 (원광대학교 근감소증토탈솔루션센터), *문희경(원광대학교)	▶ 66
▪ 의미적 분석과 연구자 관심 분야를 이용한 자동 추천 시스템	
..... *홍광희 (여주대학교)	▶ 68
▪ 메타버스 공간을 활용한 하이플렉스 수업 적용연구	
..... 이영미, *김휴찬 (제주한라대학교)	▶ 72

2024년도 (사)융복합지식학회

추계학술발표대회 구두 세션

유클리디안 거리 공식을 이용한 2D 게임 속 AI 봇 연구

임재선, 김찬영, 박영훈, 박지호, 신현지, 유승민, 최유진, 한동연, 박승균
 신한대학교
 Shinhan Univ

Research on AI Bots in 2D Games Using the Euclidean Distance Formula

Lee Soon Shin, Im Jae Seon, Kim Chan Young, Park Yeong Hun, Park Ji Ho, Shin Hyeon Ji, Yu Seung Min, Choi Yu Jin, Han Dong Yeon, Park Seung Gyun
 Shinhan Univ

요약

현대 게임에서는 인공지능(AI)이 핵심 요소이다. 게임 AI의 추적 기술은 게임의 몰입도와 긴장감을 높인다. 이 연구에서는 AI의 효율적인 추적을 위해 유클리디안 거리 공식을 사용한다. 유클리디안 거리 공식은 n 차원 공간에서 두 점 사이의 거리를 계산하는 데 사용되며, 2차원 공간에서의 유클리디안 거리 공식은 피타고라스 공식과 유사하다. 이 공식은 게임에서 객체 간의 거리 계산과 같은 거리 알고리즘과 인공지능 등 다양한 분야에 사용된다. 적 AI가 플레이어를 추적할 때 유클리디안 거리 계산을 사용하여 두 객체 간의 거리를 수치화시키고, 설정된 수치 이내일 경우 적 AI가 플레이어를 추적하도록 함수를 작동시킨다. 적 AI의 추적 방향은 플레이어와의 x축 비교만으로 결정하므로 구현이 간단하고 디버깅도 용이하다.

I. 서론

현대 게임 개발에서 인공지능(AI)은 플레이어와의 상호작용을 더욱 풍부하고 현실감 있게 만들어주는 핵심 요소로 자리 잡고 있다. 뱀파이어 서바이버즈 라는 게임으로 예를 들자면 이 게임은 직관적이고 조작성이 쉬운 간단한 구성의 게임이다. 간단한 구성의 게임으로 재미를 느끼게 만드는 것은 어렵다. 그럼에도 플레이어들에게 많은 사랑을 받았는데 그 이유에는 무작위성과 몰입도라고 말할 수 있다. 이 게임은 제한 시간 동안 무수히 많은 적들이 플레이어를 끊임없이 추적하며 게임에 긴장감을 불어넣는다.[1] 적 AI가 플레이어의 위치를 지속적으로 추적하는 기술은 게임의 몰입도를 높이고, 전략적 요소를 강화하는 데 중요한 역할을 한다. 이러한 위치 추적 기술은 플레이어의 행동에 실시간으로 반응할 수 있도록 하여, 게임의 긴장감을 극대화하고 몰입도를 높이는 데 기여한다. 본 연구에서는 AI의 플레이어 추적 알고리즘을 설계하고 구현하는 방법을 다룬다. 이를 위해 유클리디안 거리 계산 공식을 사용하여 두 지점 간의 거리를 효율적으로 측정하고, 감지 범위를 설정하여 플레이어를 인식하는 과정을 설명한다. 감지 범위는 적 AI가 플레이어를 인식할 수 있는 최대 거리로, 이 범위 내에서만 플레이어의 위치를 추적하도록 설정된다. 이러한 접근 방식은 AI의 자원을 효율적으로 관리하고, 게임의 성능을 향상시키는 데 기여한다. 적 AI가 플레이어를 향해 이동하는 방향은 플레이어와 적 간의 벡터 차이를 통해 계산된다. 이 벡터 계산을 통해 적이 플레이어를 향해 직선으로 이동시키고 속도 값 조절을 통해 보다

직관적이고 자연스러운 추적 행동을 가능하게 한다. 또한, 적이 플레이어에게 접근할 때마다 스프라이트의 방향을 조정하여 시각적으로도 플레이어를 향해 다가가는 모습을 실현한다. 이는 플레이어에게 더욱 몰입감 있는 경험을 제공하며, 게임의 시각적 표현을 강화하는 요소로 작용한다.

II. 본문

2.1 유클리디안 거리(Euclidean distance)

“유클리디안 거리(Euclidean distance)” 공식은 n차원 공간에서 두 점 사이의 거리를 알고 싶을 때 사용하는 공식이다. 이 중에서 2차원 공간에서의 공식은 피타고라스 공식과 매우 유사하다.[2] 피타고라스 공식은 다음과 같다. 꼭짓점이 각각 A, B, C이고 밑변이 a, 높이가 b, 빗변이 c인 직각삼각형이 있다면, 항상 $c^2 = a^2 + b^2$ 가 성립한다. 이것을 변형하면 $c = \sqrt{a^2 + b^2}$ 이 된다. 이 때 점 $A(x_1, y_1)$ 와 $B(x_2, y_2)$ 사이의 거리 $d(c)$ 를 알고 싶다면 a와 b를 구해 위 공식으로 해결할 수 있다. 이 도형은 2차원 평면좌표를 따르므로 밑변 a는 $x_1 - x_2$ 이고, 높이 b는 $y_1 - y_2$ 로 나타낼 수 있다. a와 b를 식 $c = \sqrt{a^2 + b^2}$ 에 대입하면 c는 d(거리)이므로 $d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ 이 된다. 즉 2차원 공간에서 두 점 사이의 거리를 구하고 싶다면, 두 점의 x좌표의 차이와 y좌표의 차이를 구하여 식에 적용하면 된다. 다른 n차원에서도 축의 개수를 늘리며 위 식을 변형하여 사용한다.

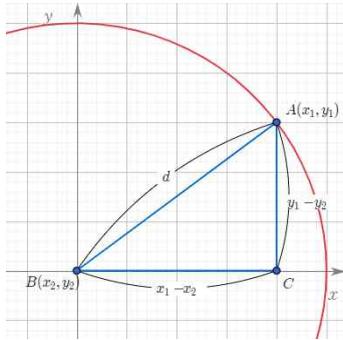
이러한 공식은 게임에서 물체 사이의 거리를 계산하거나 거리 계산 알고리즘, 인공지능 등 여러 분야에 사용할 수

있다. 적이 플레이어를 추적할 때 유클리디안 거리 공식을 이용한다. 이를 이용하여 적과 플레이어 사이의 거리를 수치화하고, 설정된 수치보다 거리가 가까울 때 적이 플레이어를 추적하는 함수를 작동시킨다.

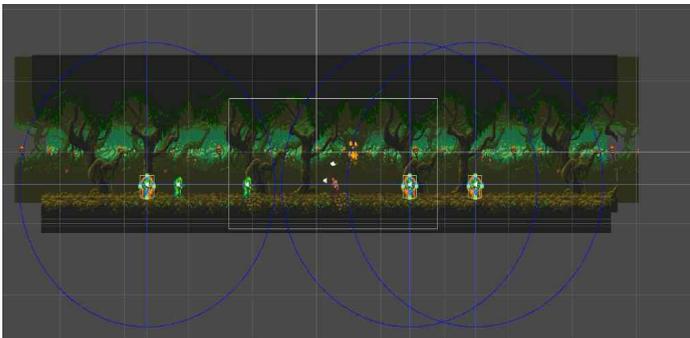
$distanceToPlayer = Vector2.Distance(transform.position, player.position)$

위 함수를 이용하여 플레이어와 적의 거리를 구할 수 있다. 각각 $distanceToPlayer$ 는 c(거리 또는 빛변)과 대응되며 $transform.position$ 과 $player.position$ 은 점A, B와 대응된다.

유클리디안 거리 공식은 두 점 사이의 위치만 알 수 있다면 거리를 쉽게 구할 수 있다는 장점을 가지고 있다. 2차원 평면상에서 아무런 조건 없이 순수 직선 상의 거리를 구할 때 다른 어떠한 거리공식보다 가장 직관적이고 적합한 공식이 유클리디안 공식이다.



(그림 1) Eculidean distance calculation method



(그림 2) Visual representation of 'DetrcionRange2D'

2.2 AI 감지 알고리즘(AI Detection Algorithmm)

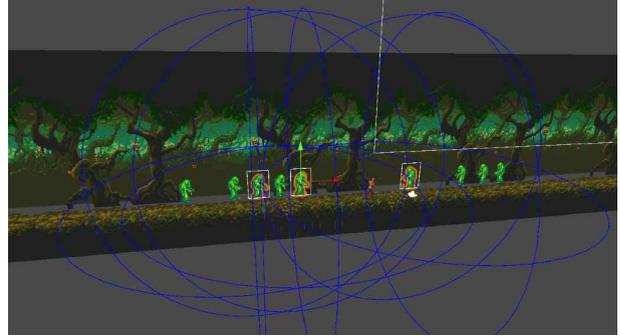
AI 감지 알고리즘은 적과 플레이어의 거리를 계산한 후, 적의 감지 범위 내에 있는지 확인하고 행동을 취하도록 하는 알고리즘이다. 이 알고리즘은 유클리디안 거리를 중심적으로 이용한다. 먼저 $detrcionRange$ 변수를 이용해 적의 감지 범위를 정의한 후, 적과 플레이어의 유클리디안 거리를 구하여 플레이어가 적의 감지 범위 내에 있다면 적이 플레이어를 향해 이동 및 모션을 취하게 한다.

첫 번째로 적이 플레이어를 감지하기 위해서, 먼저 $Vector2.Distance$ 함수를 사용하여 적과 플레이어 사이의 거리를 구한다. 해당 함수는 $transform.position$ (적의 현위치)와 $player.position$ (플레이어위치) 사이의 유클리디안 거리를 계산한 후, 계산값을 $distanceToPlayer$ 변수에 저장한다. 만약 $detrcionRange$ (적의 감지 범위)의 값보다

$distanceToPlayer$ (적과 플레이어 사이 거리)의 값이 크다면, 적의 감지 범위 내에 들어온 것이므로 $MoveTowardsPlayer(distanceToPlayer)$ 메소드를 호출한다. 이 메소드는 적이 플레이어를 향해 움직이는 AI 추적 함수이며, 일정 속도로 이동할 수 있도록 한다.

두 번째로 적의 범위 내에 플레이어가 인식되지 않는다면 $Rigidbody2D rb$ 의 현재 이동속도를 변경하지 않고 물리 속성을 그대로 유지하게 한다. [3~5]

해당 알고리즘은 적과 플레이어 사이의 거리 계산이 간단하고 조건문의 난이도 또한 높지 않아 구현하기 쉽다. 프레임 속도 역시 안정적으로 유지할 수 있다는 장점이 있다.



(그림 3) Visual representation of 'DetrcionRange3D'

2.3 AI 적 추적 알고리즘(AI enemy tracking algorithm)

AI 적 추적 알고리즘이란 AI 기능을 가진 적이 플레이어가 감지 범위 안에 들어왔을 때 추적 하고 또한 플레이어가 감지 범위를 벗어나게 된다면 추적을 정지하게 된다. 위 1번 알고리즘에 사용된 유클리드 거리 기반으로 설계한 캐릭터 추적하는 로직이다. 플레이어의 현재 시점에서 AI와의 거리를 인용한 로직을 설계했으며, $player.position.x$ (플레이어 위치)와 $transform.position$ (적의 현위치)를 비교하여 만약 0보다 클 경우 AI는 플레이어가 우측에 적이 있다고 판단하여 우측으로 이동하게끔 구현이 되었으며, 만약 0보다 작으면 좌측에 이동하게끔 추적 알고리즘을 설계하였다. 또한 $velocity.y$ 를 선언하여, y축의 고정값을 유지하여 적AI의 y축 이동을 방지한다. 게임의 코드를 분석해 보면 $moveSpeed$ 값을 이용하여 적을 일정한 속도로 이동시키는 방식이기 때문에, 복잡한 계산 없이도 자연스러운 이동을 구현할 수 있고 X좌표 값의 비교만으로 적의 이동 방향을 결정하기 때문에 구현이 매우 쉽고 복잡한 로직도 없을뿐더러, 디버깅에도 용이하다. 또한 오직 X축을 기준으로만 이동하기 때문에 Y축을 사용하지 않는 횡스크롤 게임에서 적당한 수준의 추적을 구현할 수 있어 매우 적합하며, 결론적으로 AI 적 추적 알고리즘은 간단한 AI를 가진 적 캐릭터를 구현하는 데에 있어 최적의 선택이며, 복잡한 패턴을 요구하지 않는 게임인 횡스크롤 게임에서 효과적인 방법이다.



(그림 4) Visual representation of 'AI enemy tracking'

$$(x_1 - x_2) > 0$$

$$\text{우측이동} : \vec{v} = (v_x, v_y) = (1 \times \text{moveSpeed}, v_y)$$

$$\text{좌측이동} : \vec{v} = (v_x, v_y) = (-1 \times \text{moveSpeed}, v_y)$$

(그림 5) AI enemy tracking

III. 결론

본 연구를 통해 유클리디안 공식과 AI 감지 알고리즘, AI 적 추적 알고리즘 3가지 기초적인 요소를 연구하고 효율적으로 구현하는 방법을 제시하였다. 또한 이를 이용해 기초적인 횡스크롤 게임을 만듦으로써 게임의 몰입도를 높일 수 있음을 확인하였다. 제시된 알고리즘은 간단한 계산을 통해 자원 소모를 최소화하여 게임 성능을 떨어뜨리지 않는 효율성을 제공함과 동시에, 적의 움직임을 자연스럽게 직관적으로 구현하였다. 특히 횡스크롤과 같이 간단한 게임에서 복잡한 패턴을 요구하지 않는 상황에서도 충분히 유용하고, 효과적인 결과를 도출해 내었다. 이에 따라 본 연구의 결과는 게임 개발에 있어 실질적인 기여와 함께 게임 개발자들에게 기초 자료가 될 것으로 기대된다.

ACKNOWLEDGMENT

※ 감사의 글 : “본 연구는 2023년도 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업 지원을 받아 수행되었음” (2023-0-00089)

참 고 문 헌

- [1] 서삼광, 「[글로벌 e게임] 심플한 몬스터 탄막 슈팅, '뱀파이어 서바이버즈」, 게임뷰, 2022.02.14, <https://m.gamevu.co.kr/news/articleView.html?idxno=22038>
- [2] 김현아. "비유클리드 평면상의 피타고라스정리 연구." 국내석사학위논문 제주대학교 대학원, 2013. 제주특별자치도
- [3] 이기연, 정형원, 김태규. (2018-11-09). 공포어드벤처게임의 추적 인공지능 패턴 모델 제시. 한국엔터테인먼트산업학회 학술대회 논문집, 대전.
- [4] 유광석(Kwangseok Yu), 김동욱(Dongk Kim), 김덕수(Deok-Soo Kim), and Chee Yap. "평면상에서 원형의 방해물의 존재시 유클리디안 최단경로 계산 알고리즘." 한국 CAD/CAM 학회 학술발표회 논문집 2004.2 (2004): pp.243-248.

- [5] 이영재. (2001). 컴퓨터게임을 위한 2D 충돌 감지 알고리즘 비교 분석에 관한 연구. 한국게임학회 논문지, 1(1), pp.42-48.

비콘 센서 기반 실내 전시시설의 방문객 편의성 증대 시스템 설계

김유진*, 원서연*, 서지훈**
강남대학교

cpyj126@kangnam.ac.kr, 202303109@kangnam.ac.kr, jihoon@kangnam.ac.kr

Beacon-Based System Design for Enhancing Visitor Experience in Indoor Exhibitions

Kim Yu Jin*, Won Seo Yeon*, Seo Ji Hoon**
Kangnam University*, Kangnam University**

요 약

기존의 실내 전시시설은 방문객들을 위해 인쇄된 종이 재질의 책자를 통한 아날로그 정보 제공 방식이 일반적이었다. 하지만 이와 같은 사례는 소지의 불편함과 정보의 갱신이 빠르지 않다는 치명적인 문제점들이 존재한다. 그러므로 본 연구는 비콘 디바이스를 기반으로 한 애플리케이션을 개발하여 방문객이 전시물에 대한 정보를 실시간으로 제공 받을 수 있도록 하는 다이렉트 전송 시스템과 실내 구획 내의 공간적 혼잡도를 파악할 수 있는 인프라 구축망을 설계하였다.

I. 서 론

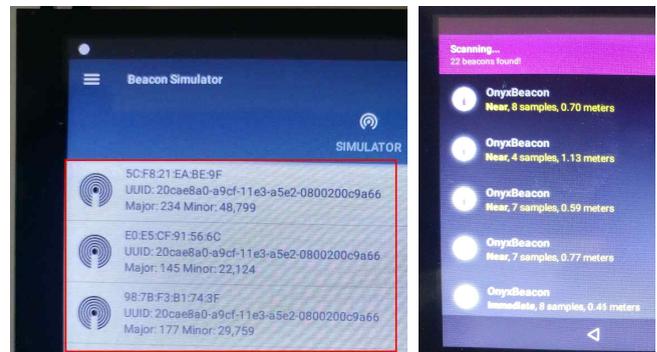
기존에 실내 전시시설을 방문할 경우, 주최 측이나 해당 부스(Booth)에서 제공하는 아날로그 형태의 종이 책자를 통해 정보를 얻을 수 있는 것이 일반적이었다. 하지만 종이 책자는 실시간으로 변화하는 전시 정보를 제공할 수 없으며 휴대의 불편함이 존재한다. 또한 일반적인 지도 애플리케이션에서는 실내에서 방문객의 위치를 파악할 수 없기 때문에 스마트폰을 통해 길을 찾을 수 없고, 원하는 전시물을 찾지 못하는 등의 한계가 존재하였다. 이러한 불편을 해소하고 방문객의 편의성을 높이기 위해 본 논문에서는 비콘 기술(Beacon technology)을 활용하여 실내 환경의 전시시설을 관람하는 방문객에게 적합한 맞춤형 인프라 정보(Infrastructure Information)를 제공하고자 한다. 본 연구의 인프라 구축 설계를 위해 활용한 비콘 센서의 4.0 버전은 초저전력(BLE:Bluetooth Low Energy)을 기반으로 동작하는 특징이 있으며 이에 따라 소비되는 전력을 절감하기 때문에 경제적인 손실을 최소화할 수 있다. 또한 비콘 센서를 중심으로 넓은 범위의 거리까지 통신이 가능하여 다양한 환경에서 구축이 가능하다는 장점이 있으며 사용자에게 수동적인 동작이 요구되지 않는 등 다양한 장점들이 있어 실내 측위, 마케팅, 농업과 축산 등 많은 분야에서 사용되고 있다. 그러므로 본 논문에서는 비콘 센서의 데이터 전송 기법과 위치정보 기능을 활용하여 실내 공간의 인프라망을 구축하고 이를 제어할 수 있는 다양한 기능의 애플리케이션 시스템을 설계하였다.

II. 본문

2.1 비콘 디바이스의 통신 방법에 대한 기능적 진화

비콘이 작동하려면, 송신기와 수신기가 필요하다. 비콘의 BLE(Bluetooth Low Energy) 기술을 통해 브로드캐스트

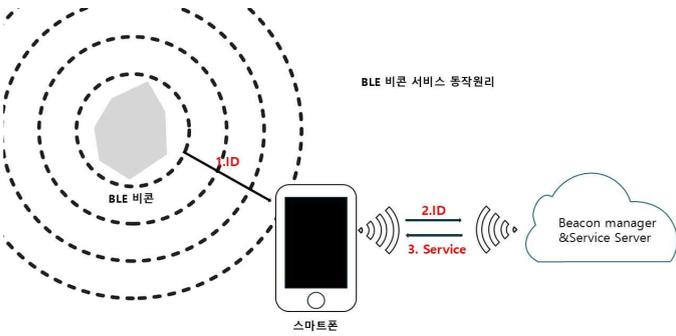
패킷 형태의 데이터를 주기적으로 송신한다. 이 패킷은 비콘의 고유 식별자인 UUID (Universally Unique Identifier) 값을 보내고 사용자의 애플리케이션과 같은 수신기는 UUID 값을 받아 서버에 보낸다. 수신 장치에서 BLE 신호를 확인하고 RSSI(Received Signal Strength Indicator)를 측정한다. RSSI 값의 강도 측정을 통해 비콘과 수신기 간의 거리의 대략적 파악이 가능하다[1].



(그림 1) BLE 비콘 디바이스의 UUID 수신 테스트

블루투스의 표준 규격은 기존 블루투스 및 저전력 블루투스를 모두 다루고 있다. 지난 수년간 많은 소비자를 대상으로 판매된 디바이스에서 일반적으로 잘 알려진 무선 규격의 기본 블루투스에 비해 더욱 발전된 저전력 블루투스는 이전 버전과 통신 될 수는 없지만 고도로 최적화된 새로운 기술로 시장에 진입하고 있다. 2010년 6월 발표된 블루투스 핵심 규격의 BLE 버전 4.0 이전 블루투스 디바이스에서는 저전력 지원이 불가능하고 BR/EDR만 지원 가능하지만 4.0 이후의 단일 모드(블루투스 스마트)에서는 BLE만 지원할 수 있다. 기존 블루투스 및 저전력 모두를 지원 가능하게 하는 것이 4.0 이후의 듀얼 모드(블루투스 스마

트 레디)이다. 단일 모드 디바이스로부터 얻은 데이터를 GSM이나 WiFi 무선을 통해 인터넷으로 전달할 수 있는 듀얼 모드 장치들은 BLE 센서가 더 많이 시장에 진입할수록 더 일반화되는 기능으로 확산하고 있다.

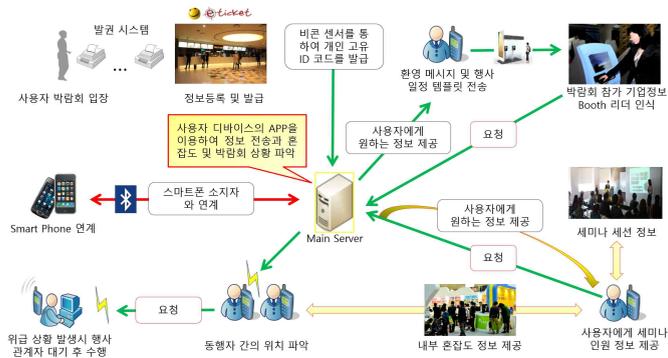


(그림 2) BLE 비콘 서비스 동작 원리

2.2 비콘 기반의 실내 전시시설 인프라망 구축

2.2.1 인프라망 모델 구성 시나리오

본 논문에서 제안하는 사용자 시나리오 모델의 절차는 다음과 같다.



(그림 3) 사용자 시나리오 모델

2.2.2 설계 환경 및 기기

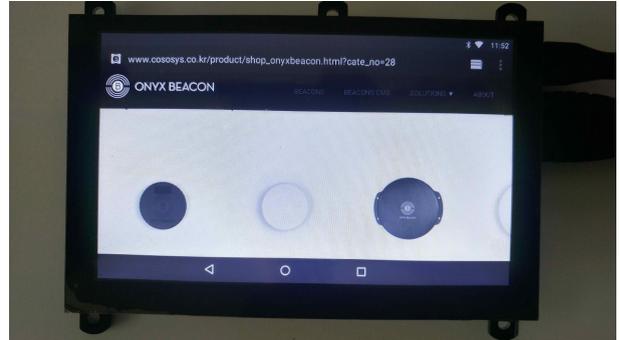
블루투스 모듈은 저전력 에너지 4.0 비콘 디바이스를 이용하였으며 안드로이드 기반으로 제어가 가능하도록 오드 로이드 장비를 이용하여 설계한다.



(그림 4) 비콘 센서 연동을 위한 안드로이드 셋팅

본 논문의 설계 단계에서 도입되는 비콘 디바이스는 일반적인 모델과 배터리의 수명주기가 높은 모델 두 가지를 동

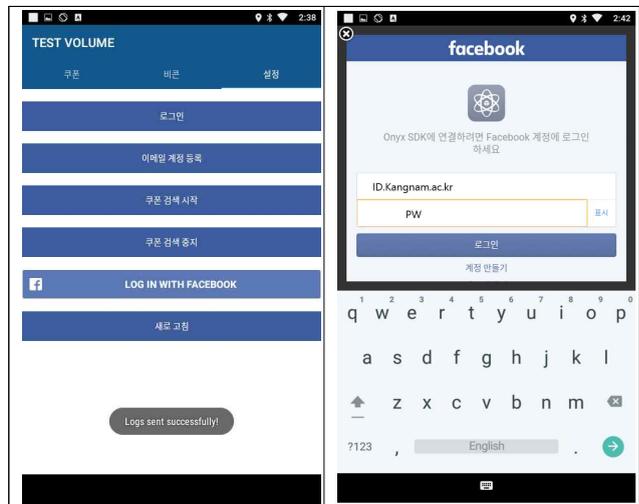
시에 이용하였다.



(그림 5) 비콘 디바이스 인식을 위한 환경 설정

2.2.3 시스템 설계

먼저, 전시시설 입구에 비콘 디바이스를 설치한다. 방문객 들은 입장 할 때 QR 코드를 통해 간편하게 애플리케이션 설치를 하고 블루투스를 필수적으로 활성화해야 한다. 그리고 보안 문제를 위해서 로그인 인증 절차를 진행하며, 서비스 이용에 동의해야 한다. 이를 통해 개인 정보 유출을 방지하며 맞춤형 서비스 제공을 할 수 있게 된다.



(그림 6) 초기 프로토타입 앱 UI와 인증 절차



(그림 7) 실제 로그인 인증 후 UISS 신호 수신 강도

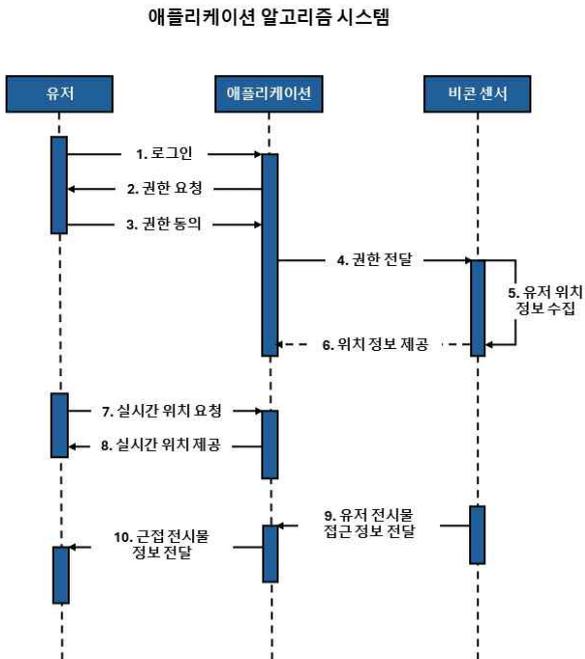
비콘 센서와 방문객의 스마트폰이 잘 연결되었다면 이제 부터 방문객은 더 이상의 조작 없이 전시물에 대한 정보를

수신받을 수 있다. 전시물들의 간격을 고려하여, 비콘 송신 범위를 너무 넓게 한다면 신호가 겹치는 문제가 발생할 수 있으므로 방문객의 탐지 범위는 5m로 제한한다.

비콘 센서가 전송할 수 있는 데이터의 양이 많지 않다는 단점이 있다. 음성 혹은 이미지로 이루어진 정보를 방문객에게 전송이 필요할 필요가 있을 수도 있는데, 이때는 비콘 자체에서 많은 용량의 정보를 직접 유저에게 전달하는 것이 아닌 애플리케이션의 기능을 통해 자세한 정보 전달이 가능하다. 비콘에서는 유저가 전시물에 근접했다는 신호를 애플리케이션에 발신하고, 이를 수신한 애플리케이션은 알맞은 정보를 팝업 및 알림을 통해 유저가 손쉽게 전시물 정보에 접근할 수 있도록 한다. 전시기관 내부 곳곳에 비콘 센서를 설치해 두고, 이를 통해 방문객들의 동선을 추적하고 전시관마다 혼잡도를 파악할 수 있다. 혼잡도를 직접 확인할 수 있는 것뿐만 아니라 방문객이 자신의 현재 위치도 파악할 수 있고, 전시관 내 혼잡도 데이터를 수집하여 학습시킨 애플리케이션 알고리즘 자체로 혼잡한 전시관을 피해 쾌적한 관람이 가능한 코스를 추천할 수 있다.

2.2.4 비콘 센서 제어를 위한 애플리케이션 동작 알고리즘

본 논문에서 제안하는 앱의 시퀀스 다이어그램은 다음과 같은 절차로 동작한다.



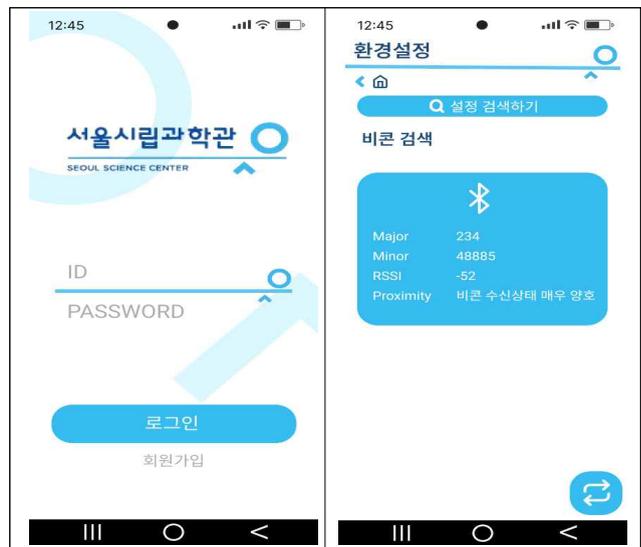
(그림 8) 애플리케이션 알고리즘 시스템

(표 1) 애플리케이션 동작 알고리즘

애플리케이션 동작 알고리즘
STEP. 1 전시기관 입구에 설치된 비콘과 방문객의 스마트폰을 연동하여 방문객의 인원수를 파악한다.
STEP. 2 전시기관 내부에 위치한 비콘에 방문객이 가까이 접근하면, 방문객에게 전시물에 관한 정보를 송신한다.
STEP. 3 곳곳에 설치된 비콘을 통해 방문객의 위치를 알아내고, 혼잡도를 파악하여 애플리케이션으로 이 점을 알려준다.

2.2.5 애플리케이션 UI 설계 및 CMS 구축

다음은 비콘 기술을 활용한 애플리케이션 UI 설계다. 자체 애플리케이션 개발이 되지 않은 전시기관 중 많은 방문객이 이용하는 서울시립과학관을 예시로 들었다. 시스템 설계에 따라 안드로이드를 기반으로 애플리케이션 UI를 설계하였다. 서울시립과학관의 로고와 메인 색상을 해치지 않는 선에서 사용자 접근이 쉽도록 깔끔한 디자인을 유지하였다.



(그림 9) 애플리케이션 로그인 실행 화면 및 센서 연동 UI

본 애플리케이션에는 로그인 기능이 탑재되어 있다. 위와 같은 인증 절차는 개인정보보호를 위해 실시한다. 실시간 위치 정보를 비콘으로 받기 때문에 사용자의 동의가 필수적이며 다른 사용자에게 제공되는 문제를 피하기 위함이다. 로그인을 통해 내 위치 제공 동의 절차를 거치고 무분별한 정보 제공을 피할 수 있다. 로그인을 한 후 사용자는 환경설정 화면을 통해서 비콘 센서와 스마트폰이 잘 연결되었는지 확인할 수 있다.

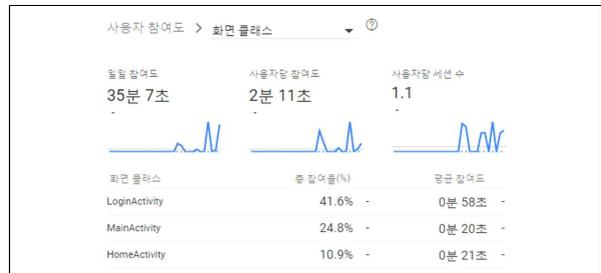


(그림 10) 내부 시설 지도 UI

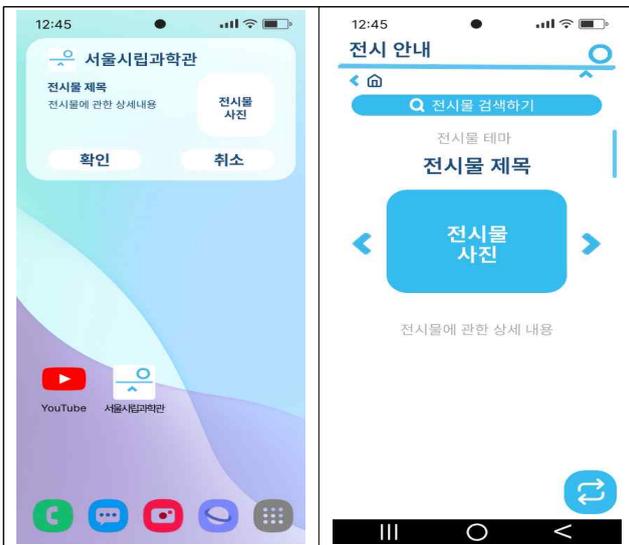
이모지를 통해 내부 시설의 혼잡도를 나타낸 화면이다. 또한 나의 위치가 지도 앱에 표현되어 보다 쉽게 목적지를 찾아가는 것이 가능하다.

Name	Protocol	Identifier	Tags	Location
Beacon1	iBeacon	00000000-0000-0000-000000001234 / 1 / 2	Beacon1	
Beacon2	iBeacon	20CAE8A0-A9CF-11E3-A5E2-0800200C9A66 / 1 / 2		
Beacon3	iBeacon	00000000-0000-0000-0000000012345 / 1 / 2		

(그림 12) 비콘 센서 정보 관리



(그림 13) 화면 클래스에 대한 관리자 정보



(그림 11) 정보수신 화면

애플리케이션을 켜지 않더라도 전시물에 가까이 다가갔을 때, 전시물에 관한 상세 내용이 팝업 형식으로 뜨는 화면이다. 상세 내용이 길거나 사진을 포함할 경우, 확인을 누르면 더 자세한 내용을 볼 수 있다. 또한 본 시스템은 CMS를 이용하여 데이터 공유에 대한 모니터링 정보를 파악할 수 있기 때문에 실용적인 특징을 가지고 있다.

III. 결론

본 논문에서는 비콘 센서를 기반으로 하여 실내 전시시설의 방문객들이 더욱 편리하게 관람 가능한 방법에 대해 설계해 보았다. 비콘 기술을 활용한 이 애플리케이션을 더 발전시킨다면 박람회 현장 등에도 적용할 수 있다. 팜플렛, 명함 등의 정보를 비콘을 통해 전달하여, 방문객들은 스마트폰으로 모든 정보를 확인할 수 있다. 박람회 특성상 전달해야 하는 홍보물이 많기 때문에, 이 애플리케이션을 통해서 방문객들은 더 효율적이고 편리한 관람이 가능할 것이다. 박람회에 참가하는 단체와 기업의 입장에서 팜플렛, 명함의 인쇄 비용 절감이 가능하다. 과거 '2024 서울국제도서전'의 사례를 예로 살펴보면 총 365개의 참가사가 있었고, 한 부스당 2종류의 팜플렛을 1,000장씩 제작했다고 가정할 경우, 인쇄 업체에 건적을 넣어본 결과, 1,000장당 약 15만 원의 비용이 발생한다. 이에 따라 한 부스당 30만 원 즉, 전시회 전체 규모로 따지면 대략 10,950 천 원의 비용 절감 효과가 있다. 그러므로 향후 본 논문에서 제시하는 비콘 인프라망 기반의 시스템이 다양한 실내 전시시설에 활용되어 더 나은 전시 환경을 제공할 수 있기를 기대한다.

참 고 문 헌

- [1] 채건희, 최성민, 설지환, 이재홍, “사물인식과 비콘을 활용한 모바일 내부정보 유출방지 시스템”, 한국인터넷방송통신학회 논문지, 18(5), 17-23, 2018.
- [2] 김형진, “비콘·센서와 스마트폰을 활용한 안전관리 서비스 구현”, 융복합지식학회논문지, 5(1), 25-31, 2017.

- [3] 변성준, 조용한, 최상근, 조봉래, 이건원, 민병학, “위치 기반(LBS) 쇠퇴지역 재난재해 위험성 예측 시스템 구축”, 한국산학기술학회논문지, 21(11), 570-576, 2020.
- [4] 박상현, 임현국, “폭이 좁은 건물 구조 하에서 GPS와 Beacon을 이용한 실내 측위 및 최단 경로 탐색 시스템의 설계 및 구현”, 한국컴퓨터정보학회논문지, 28(3), 11-16, 2023.
- [5] 서지훈, 주길홍, “근거리 통신망을 이용한 센서 기반의 캠퍼스 인프라 구축 시스템” 인공지능연구 논문지, 5(1), 36-44, 2024.

객체인식을 활용한 개문사고 방지 시스템에 관한 연구

최민준1), 김서희2), 송효은3), 안제홍4), 강지훈5), 이동규6)

신한대학교 소프트웨어융합과

{chlalswns916, rok618, 030419s, burindol3, 1rkdwlgn1}@naver.com

dglee@shinhan.ac.kr

Study on the the prevention system for door accidents using object detection

Choi minjun, Kim seohee, Song hyoeun, Ahn jaehong, Kang jihoon, Lee

donggyu

Shinhan Univ

요 약

본 논문은 갑작스러운 개문 사고로 크게 다치거나, 뒤에 오는 차량을 미처 확인하지 못한 운전자와 안전 운전을 했으나 급작스러운 개문으로 피할 수 없는 사고에 휘말린 운전자 모두 억울한 상황이 발생하는 등의 문제를 해결하고자 객체 인식 프로그램인 YOLOv5를 활용해 개문사고 예방 시스템을 구현하였다. 이 시스템은 YOLOv5를 통해 높은 정확성과 빠른 속도로 객체를 탐지하고 거리 인식까지 한번에 처리하는 효율성을 가지고 있다. 프레임 속도, 해상도, 픽셀 수 등을 조정해 가며 시스템의 정확성과 객체 인식 속도를 높였다.

II. 본 문

2.1 기존 개문사고 방지 시스템 분석

기존 개문사고를 위해 사회에서 고안된 방지 시스템들이 존재하고 있으며 우리가 일상 속에서 쉽게 눈에 보이는 시스템으로는 모든 차량에 붙어있는 백미러로 직접 수동으로 확인하는 수단이 있고 택시 같은 경우엔 승객들도 확인이 가능한 전용 후방미러가 있지만 오로지 사람이 보고 판단해야 하는 영역이기에 버스같이 사람이 많이 이용하는 대중교통의 경우 이것만으로는 사고를 방지하는데 한계가 있다.

그래서 고안된 시스템으로는 버스 같은 경우 기사 1명이 판단하는데 힘들 수 있기에 운전 중에는 절대 도어가 열리지 않도록 도어장치와 주차브레이크의 작동과 연관되도록 하여 무조건 브레이크를 밟아 정차한 상태여야만 열리는 시스템이 존재한다[2]. 과거에는 안내양같이 승객들이 내리는 문을 직접 관리하는 사람이 있었지만 하차벨과 자동문으로 시스템이 기계로 바뀌되로는 사실상 승객들을 관리하는 건 운전기사 한 명 밖에 존재하지 않는다. 그래서 차량이 멈춰 있지 않은 한 절대 문을 열지 못하도록 하고 안전한 상태에서 문을 열어 기사와 승객이 안전을 지킬 수 있도록 시스템이 되어있는 상태다.

지하철의 경우 일반 차량보다 더 많은 문이 있지만 과거엔 승강장과 열차 사이에 안전 장치가 따로 없었고 사고가 일어날 가능성이 높았기에 이를 방지하기 위하여 승강장에 고정벽과 자동문을 추가로 설치하고 오고 가는 차량과 신호를 연결하여 출입문과 스크린도어의 개폐문이 동시에 열리도록 하는 ‘스크린 도어’ 시스템을 도입하였다[3].

이를 통해 스크린 도어가 없을 때 아무런 관리가 안 되어 사고를 방지할 수가 없었지만 도입된 이후에는 사고 사례도 현저히 줄어들었다.

I. 서 론

1.1 연구 배경

개문사고란 주정차 중인 차량의 문이 갑자기 열려 그 옆을 지나던 자전거, 오토바이, 자동차 혹은 보행자가 부딪히는 사고를 뜻한다. 이러한 사고가 발생하는 이유는 차 안에서 문을 열기 전에 주변의 상황을 충분히 확인하지 않았을 때 발생하며 주로 도심 지역이나 좁은 차도에서 많이 발생한다.

개문사고는 다양한 형태로 나타나며 그로 인한 사회적, 경제적 손실도 상당하다. 도시교통정보센터에 나타난 최근 3년간 개문사고에 대한 통계는 5,733명 발생 8명 사망 6,273명 부상으로 상당한 수준이다[1]. 사고로 인해 개인과 개인 사이의 감정을 악화시키는 사회적 영향이 있고 사고 처리 비용, 보험료 인상, 의료비 등 경제적인 영향도 끼친다.

1.2 연구 목적

본 연구는 개문사고를 예방하기 위해 시작되었다. 개문사고는 보통 차에서 차 밖의 상황을 인지하지 못하고 차 문을 여는 것에서부터 시작하기 때문에 이를 방지하기 위해 기술적인 방법이 필요하다고 생각했다.

1.3 연구 범위 및 방법론

본 연구는 개문사고에 대한 사회적, 경제적 영향을 고려하고 개문사고를 충분히 줄일 수 있는 방법에 대해 연구하는 것이다. 이를 위해 기본적인 예방 방법을 생각해 예방을 실현시킬 수 있는 현실적인 방안이 의거하여 연구하였다.

2.2 관련 기술 및 이론

위에서 설명한 개문사고 방지 시스템들을 관련된 기술들을 살펴보면 먼저 우리가 어떤 걸 볼 때 눈으로 판단하는 것처럼 시스템에서 우리들을 보고 판단하는 센서 기술이 가장 대표적이다.

이러한 기술에는 적외선 센서나 마이크로파 센서가 있으며 이러한 센서는 사람과 차량을 측정하고 이를 시스템과 연결시킬 수 있다. 허나 모든 것이 완벽한 건 아닌 듯 말 그대로 보이는 것으로 판단되기 때문에 만약 추운 겨울에 따듯한 실내같이 온도 차이가 심한 환경이면 오작동이 있을 확률이 있고 마찬가지로 태양열로 인해서도 오류가 발생할 수도 있으며 또한 감지속도에 있어서도 그렇게 빠르지 않은 점이 단점이다.

그래서 이를 보완하기 위해 오류를 없애진 못하더라도 초음파 센서를 추가로 도입하여 탐지 범위를 늘이고 탐지에 대한 속도를 높여 보완하기 위한 수단도 나왔다[4].

이처럼 개문 사고를 방지하는 것은 이러한 시스템도 중요하지만 더 이론적 배경으로 들어가면 먼저 인식 이론이 중요하다. 앞서 말한 것처럼 지하철이나 버스같은 대중교통의 경우 운전기사 말고는 거의 모든 것이 기계화 되어있고 이를 제어하는 것이 곧 운전기사이기에 기사의 역할이 중요하다고 볼 수 있는데 이러한 운전자의 인식 과정과 주의 분산을 이해하고 이와 관련하여 시스템을 보완하면 더 효과적인 시스템이 개발될 수 있다.

또한 과거에 있었던 사고를 분석하고 통계하여 그저 보완하는 것이 아닌 그동안 있었던 것을 빅데이터로 특정 패턴을 분석하고 그것의 경향을 이용하여 시스템을 짜면 더 효과적인 방안이 나올 수도 있다.

2.1.1 시스템 기능적 요구사항

이 시스템은 라즈베리파이5에 연결된 웹캠으로부터 영상을 실시간으로 불러와 캡처하고, YOLOv5 라는 객체인식 프로그램을 활용해 객체를 인식하고 분류하는 것에서부터 시작한다. 특정 거리 내에서 YOLOv5가 객체를 사람, 오토바이, 자전거, 자동차로 객체를 분류할 경우 LED 경고등이 켜짐과 동시에 무선 스피커에 “자동차가 다가오고 있습니다. 주의하세요” 와 같은 경고 음성이 출력되도록 구성하였다.

2.1.2 시스템 비기능적 요구사항

우선, 이 시스템의 목적은 후방에서 다가오는 객체로부터 충돌을 예방하는 것을 목적으로 설계된 것이므로 객체 인식의 속도를 높이는 것이 가장 중요하였다. 이를 위해 프레임 속도를 조정해 가며 객체 인식에 소요되는 시간을 최대한 단축하고 실시간으로 인식하는 것에 중점을 두었다.

두 번째로, 정확성을 높이기 위해 객체 인식 거리 설정에 집중하였다. 이 시스템은 20m 내로 객체가 인식될 경우 경고 신호를 출력하여 사용자가 위험을 인식하도록 설계 되었다. 이 거리는 객체의 움직임 속도가 각기 다르기에, 가장 빠른 속도를 보이는 자동차가 접근할 때를 고려해 설정한 거리이다.

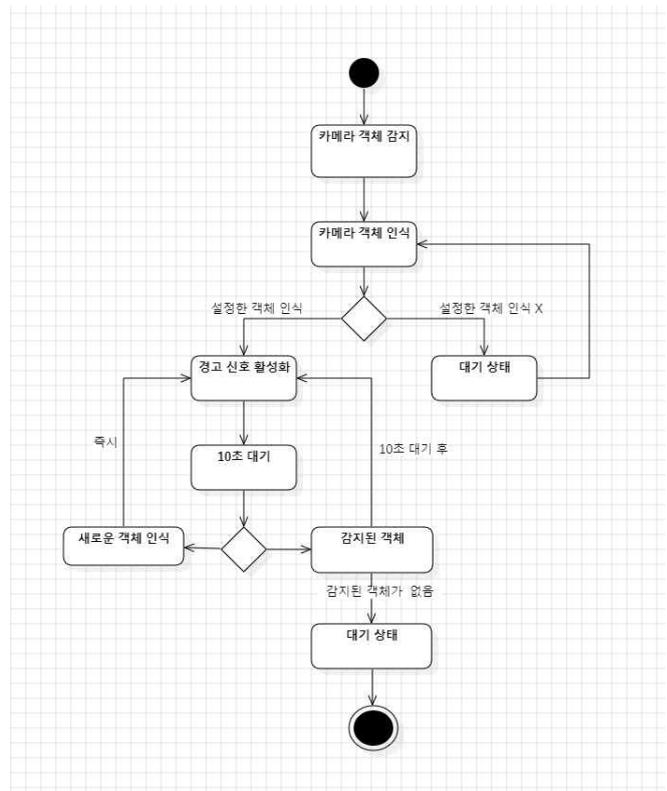
2.2.1 시스템 하드웨어

본 시스템의 하드웨어로는 라즈베리파이5가 이용되었다. 라즈베리파이5를 통해 LED를 경고등을 점등함과 동시에 무선 스피커를 통해 경고 음성을 함께 출력할 수 있고 다른 장치와도 호환성이 뛰어나 이번 시스템의 요구 사항을 만족시키는데에 효율적이라고 판단하였다.

2.2.1 시스템 소프트웨어

소프트웨어로는 실시간 객체 인식을 위하여 YOLOv5를 사용하였다. YOLOv5는 마치 사람처럼 생각하는 딥러닝을 이용한 기술로 높은 정확성과 빠른 속도를 이용해 객체를 인식할 수 있다는 큰 장점이 있다. 뿐만 아니라 YOLOv5 내에 자체적으로 거리를 탐지하는 기능이 탑재되어 있기에 별도의 초음파 센서나, 인체 감지 센서가 필요하지 않아 기능적으로나 비용적으로나 쉽게 거리를 탐지 할 수 있다. 그 결과 전체적으로 시스템의 효율성을 향상 시킬 수 있다.

3.1.1 프로토타입 개발



프로젝트 특성상 빠르게 위험을 인식하고 알려야 하기에 이미지 전체를 한 번에 처리하여 객체를 인식하는 것이 신속하고 높은 정확도를 가진 YOLO를 기반으로 개발했다. YOLO는 객체에 대한 정보를 학습시켜 정확도를 높일 수 있다는 장점을 활용해 정확도를 지속적으로 개선할 수 있었다. 이 시스템은 웹캠을 통해 실시간으로 위험을 탐지하고 사용자에게 위험을 알릴 수 있도록 개발했다. 이 프로젝트는 실시간으로 위험을 인식하고 사용자에게 경고를 알리는 것을 목표로 프로토타입을 개발했다.

3.1.2 데이터 수집 및 처리

위험을 인식하는 데이터 처리 과정을 크게 두 가지 단계로 설명할 수 있다. 첫 번째 단계로 YOLO 모델을 활용해 실시간으로 영상을 처리하고 객체를 인식한다. YOLO는 이미지 전체를 한 번에 처리하기 때문에 객체를 인식하는 속도가 빠르다. 그리고 YOLO에 객체에 대한 정보를 학습시켜 인식의 정확도를 높였다.

두 번째 단계로 인식한 객체에 대한 거리 측정을 위한 방식을 초음파 센서 대신 객체의 크기에 기반해 거리를 측정하는 방식을 채택했다. 프로젝트의 소음이나 온도 변화의 환경을 고려해 초음파 센서보

다 카메라를 통해 거리를 측정하는 방법이 낫다고 판단했다. 카메라로 실시간 영상 처리와 거리 계산을 수행할 수 있는 장점이 있고 영상의 프레임률 지속적으로 개선해 효율성과 정확도를 향상시킬 수 있었다.

위험을 인식하는데 YOLO와 크기 기반 거리 측정을 이용하였고 위험을 인식한 후 사용자에게 경고를 위해 경고 시스템을 구현했다. 시스템이 사용자의 안전을 확인하고 즉각적으로 경고를 하는데 중점을 두어 직관적으로 위험을 알릴 수 있고 사용자가 쉽게 인식할 수 있는 LED와 음성 파일을 재생할 수 있는 스피커를 활용했다.

III. 성능 평가

4.1.1 객체 인식 정확도

사람, 자동차, 자전거, 오토바이와 같은 객체를 상대로 수행하였으며, 객체의 움직임이 사람인지 물체인지 구별하여 포착하는 것도 가능했다. 보다 정확한 평가를 위해서 실제 도로에서 수행해 보았으며, 자동차나 사람, 오토바이 등 객체를 인식하였을 때 위치나, 종류, 크기 등 객체 정보 검출이 용이하게 나타났고 인식 정확도로는 평균적으로 90% 이상의 인식률을 나타내었다.

4.1.2 거리 측정 정확도

객체 탐지에서 거리 측정은 객체와의 거리를 감지하고 학습하는 것으로, YOLOv5 객체 인식 모델을 사용하여 수행하였다. 객체(보행자, 차량, 오토바이 등)의 크기와 위치를 파악하고, 객체와의 거리를 측정하였다. 초음파 센서를 사용해 보았으나 모든 객체를 측정하는 데에 있어서 한계가 있었다. 따라서, 객체 탐지와 객체 성능을 종합적으로 평가하여, YOLOv5 객체 인식 모델을 사용하는 것으로 결정하였다. 크기와 해상도를 고려하여 거리 정보를 실시간으로 추정할 수 있는 장점을 가지고 있고, 테스트 과정에서 보행자, 차량 등 여러 객체를 인식하였으며, YOLOv5 객체 인식 모델이 인식한 정보와 위치를 기반으로 실제 거리와 인식한 거리를 동일하게 인식하는 높은 정확성을 보였다.

IV. 결론

초기 테스트 결과, 카메라가 반응하는 속도와 객체를 인식하는 반응 속도가 미흡하다 판단하였다. 피드백을 반영하여 우선, 객체와의 거리 인식을 기존 10m에서 20m로 연장하였다. 이는 특히 오토바이 같이 속도가 빠른 객체들의 변수가 있을 수 있어 인식 거리를 조금 더 확보한 것이다.

두 번째로 높은 해상도로 인해 버벅거림과 반응속도가 문제가 되었다. 실시간 객체를 탐지하는 데에 있어서 높은 해상도로 인해 카메라가 인식하는 객체가 5초 딜레이 나는 상황이 발생하였다. 따라서 해상도를 최적화하였으며, 픽셀의 수를 줄이고 프레임 속도를 30FPS 고정으로 유지하게 했다. 또한, 해상도를 320 x 240까지 낮추어서 처리 속도를 높였다. 이와 같이 성능과 처리 속도를 개선하여 성능을 다 담지는 못했지만 0.5초라는 딜레이가 나왔고 성능과 속도의 균형을 맞춰 전체적으로 시스템의 신뢰성을 높여보았다.

ACKNOWLEDGMENT

※ 감사의 글 : “본 연구는 2023년도 과학기술정보통신부

및 정보통신기획평가원의 SW중심대학사업 지원을 받아 수행되었음” (2023-0-00089)

참 고 문 헌

- [1] 도시교통정보센터 “사고처리정보 중 개문사고”, <https://www.utic.go.kr/additional/newAccidentCase.do>
- [2] 류병곤, “공기압을 이용한 버스의 개문발차 방지시스템”, 2008, 10-2008-0045386, 2008-05-16, 2009-10-12.
- [3] 장변선, “지하철 승강장내에 스크린도어 도입배경 및 국내현황”, pp.17 - 21, 2003.
- [4] 김기두, “초음파 센서를 이용한 자동문의 물체인식 성능개선”, pp.97 - 107, 2017

IoT를 활용한 인텔리전트 빌딩 기반의 실내 조명 제어 플랫폼 개발

송채빈*, 오영은*, 정지혜*, 서지훈**
강남대학교

{202204226, kou527, 202304308, jihoon}@kangnam.ac.kr

Development of an indoor lighting control platform based on intelligent building utilizing IoT

Song Chae Bean*, Oh Young Eun*, Jeong Ji Hye*, Seo Ji Hoon**
Kangnam University*, Kangnam University**

요 약

본 연구에서는 안드로이드 앱과 Wi-Fi 모듈이 장착된 H/W 컨버터를 연동하여, 실내 조명을 그룹별로 제어하고 조도를 조절할 수 있는 스마트 조명 제어 시스템을 성공적으로 구현하였다. 이를 통해 건물의 에너지 효율을 높이고, 사용자의 편의성을 증대시켰다. 또한, 에너지 소비 절약 예측 알고리즘을 통해 사용자 기반의 환경을 조성할 수 있다. 본 시스템은 중·소형 건물뿐만 아니라 대규모 인텔리전트 빌딩에도 적용 가능하며, Green-IT에 기여하는 솔루션으로서의 가능성을 확인하였다.

I. 서 론

21세기 들어서 에너지 절감 및 효율성 증대는 중요한 화두가 되고 있다. 특히 빌딩 자동화 시스템(BAS, Building Automation System)에서 스마트 조명 제어는 에너지 사용을 효율적으로 관리하는 중요한 요소로 자리 잡았다. 본 논문은 국내 스마트폰 점유율이 가장 높은 안드로이드 기반 애플리케이션과 RTC(Real Time Clock) 모듈, Wi-Fi 모듈이 내장된 하드웨어 컨버터를 연동하여 LED 조명을 제어하는 시스템을 개발하고자 한다. 해당 시스템은 건물의 LED 조명을 그룹으로 분류하고, 사용자가 원하는 조도로 조정하며, 이를 통해 에너지 절감을 추구하는 것을 목표로 한다. 본 연구는 크게 세 가지 기능군으로 나누어졌으며, 각각의 기능은 스마트한 조명 제어, 이력 관리, 그리고 사용자 설정 기능으로 구성되어 있다. 이 시스템은 Green-IT에 기여하고, 보다 인간 중심적인 스마트 실내 조명 제어 시스템의 구축을 목표로 한다. 특히, 건물 규모에 상관없이 조명을 유연하게 관리할 수 있는 점에서 실용적인 솔루션을 제공한다.

II. 본문

2.1 기능의 요구조건

본 논문에서 제시하는 기술은 안드로이드 기반의 컨트롤러를 활용하여 지능형 고층빌딩의 실내 구획에 설치된 조명을 제어할 수 있는 기술로써 다음과 같은 항목을 고려하였다. 첫 번째, 스마트폰을 이용하여 LED 조명을 제어할 수 있는 인간 중심형 지능 시스템으로 구축되어야 한다. 즉, 홈 서비스 시스템 인프라에 적용이 가능하고 극장 관람 조명 스케줄링 시스템에도 적용이 가능해야 한다. 또한 공공기관과 학교 강의실, 기업 사무실 환경 시스템 등 다양한 곳에서 다양한 상황에 적용이 가능해야 한다. 두 번째, 스마트폰 앱을 이용하여 간편하게 LED 조명 제어

시스템을 구축하기 때문에 공공 서비스 및 산업용 시스템 구축에도 적용이 가능해야 한다. 스마트팜 기반의 농작물 재배 조명 시스템, 일반 자영업, 스마트 LED 광고 산업, 주차 관리 시스템 및 경기장 조명 시스템, 박람회 부츠 조명 시스템 등 다양한 공공 분야에 적용이 가능해야 한다. 세 번째 지능형 고층 빌딩의 전 구역의 제어가 가능해야 하고 비상시 실내 특별피난 계단 및 이동 경로 LED 유도 등의 서비스에 적용이 가능해야 한다. 또한 건설현장 근로자의 위치 지시등 LED 이동 경로, 지하철 LED 이동 경로 표시판 운영 서비스 등 다양한 시스템에 적용이 가능해야 한다.

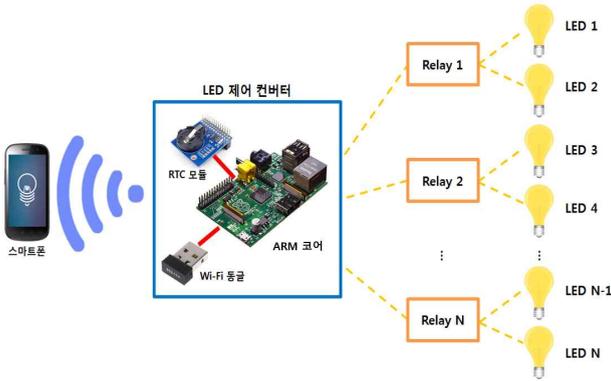
2.2 시스템 제어 기능

본 논문에서 제시한 “조명 제어 플랫폼”은 보급률이 높은 스마트폰을 이용해 실내 환경에 적합한 LED 조명의 밝기를 설정할 수 있고, 에너지 절감을 위한 절전모드 등의 제어를 실용적으로 활용할 수 있는 지능형 리모트 컨트롤러 시스템을 구축하였다. 본 기술에서 특화된 항목은 앱과 컨버터를 연동하여 IoT 기반의 실시간 조명 제어 시스템을 구축한다. 하나의 앱 컨트롤러로 고층 건물의 전체 실내 조명을 제어할 수 있기 때문에 관리자들이 효율적으로 운영할 수 있다. 또한 LED 조명의 전력 및 밝기를 실용적으로 관리할 수 있는 무선 시스템을 구축하고, 사용자 중심의 설정 기능을 부여하여 실내의 쾌적한 환경을 제공할 수 있는 LED 조명 제어 시스템을 개발하였다.

2.3 시스템 아키텍처

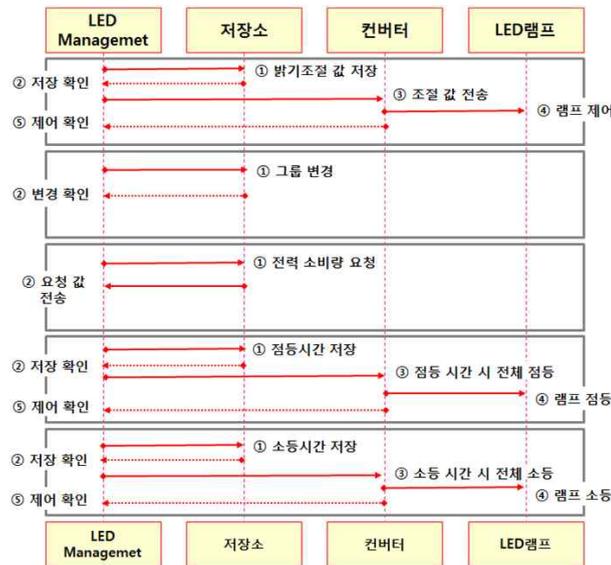
시스템은 크게 안드로이드 앱, Wi-Fi 모듈이 탑재된 H/W 컨버터, 그리고 LED 조명 장치로 구성된다. 안드로이드 앱은 사용자가 직관적으로 그룹별 조명을 제어하고, 밝기 조절 및 상태 확인을 할 수 있는 인터페이스를 제공

하며, Wi-Fi 모듈을 통해 H/W 컨버터와 통신한다.



(그림 1) 전체 코어 흐름도

H/W 컨버터는 Wi-Fi 신호를 통해 명령을 받아 LED 조명을 제어하고, 상태정보를 전달한다. 사용자는 앱을 통해 LED 조명을 그룹으로 나누어 관리할 수 있다. 예를 들어, 건물의 각 층이나 방을 하나의 그룹으로 설정하고, 그룹별로 개별 조작이 가능하다. 이는 사용자가 특정 구역에서만 조명이 켜지거나 꺼지도록 설정할 수 있도록 하며, 이를 통해 에너지를 효율적으로 사용할 수 있다.



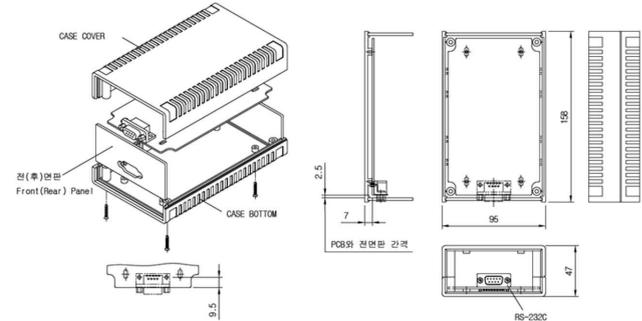
(그림 2) 그룹관리 기반 제어 시퀀스 다이어그램

LED 조명의 밝기는 사용자가 원하는 대로 조절할 수 있다. 앱에서 슬라이더 인터페이스를 통해 밝기 값을 조정하면, Wi-Fi를 통해 해당 명령이 H/W 컨버터로 전달되고, LED 조명의 밝기가 조절된다. 또한 현재 조명의 상태를 실시간으로 확인할 수 있는 기능도 제공한다. 사용자는 각 그룹의 조명이 켜져 있는지, 꺼져 있는지, 그리고 밝기 설정이 어떤지 확인할 수 있다. 이는 에너지 사용량을 직관적으로 파악할 수 있게 한다. 또한 고층 빌딩 등의 대형 건물에서 LED 조명 제어 이력 정보를 저장하는 기능을 제공한다. 조명이 켜지거나 꺼지는 시간, 밝기 조절에 대한

상세한 기록 등을 저장하여, 에너지 사용량을 분석할 수 있다. 이러한 데이터를 기반으로 향후 더 효율적인 조명 제어 계획을 세울 수 있다.

2.4 하드웨어 및 외부 구성

본 논문에서 제시하는 조명 제어 플랫폼에서 프로토콜 컨버터와 안드로이드 간의 데이터 전송 프로토콜의 기본사항 및 패킷의 구조는 다음과 같이 적용된다.



(그림 3) 컨버터 외관 전면부 상세

본 논문에서 제시한 기술에 대한 프로토콜의 전반적인 정보는 다음과 같다.

(표 1) 프로토콜 기본사항

통신방법	소켓통신
Server	프로토콜 컨버터
Client	안드로이드
포트번호	60XX
byte order	big endian

본 논문에서 진행하는 프로토콜 컨버터와 안드로이드 간의 데이터 전송 프로토콜에서 적용될 세부 항목들은 다음과 같다.

(표 2) 프로토콜 세부항목

기능	기능에 대한 세부 내용	
Write Dimming	요청(Client -> Server)	응답(Server -> Client)
Query Status	요청(Client -> Server)	요청(Server -> Client)
Get Log	요청(Client -> Server)	응답(Server -> Client)
Protocol Convertor	확인(Client -> Server)	-
해당 IP를 가진 Protocol Convertor	확인(Server -> Client)	응답(Server -> Client)
error	-	응답(Server -> Client)

(표 3) Paket 구조

STX	PW	CMD	length	data[]	check sum	ETX
1byte	4bytes	1byte	1byte	length bytes	1byte	1byte

2.5 에너지 소비 절약을 위한 예측 알고리즘 적용

본 논문의 조명 제어 플랫폼의 일부 기능은 일반적인 조명이나 과거 선행연구에서 진행된 아이디어도 포함한다. 그러나 이러한 시스템은 사용자의 수동적인 입력에 의존하기 때문에, 건물 내 인원 수 또는 외부 환경 변화를 반영하는 것에 한계가 있다. 따라서 본 논문에서는 에너지 소비 패턴을 학습하고, 이를 기반으로 에너지 소비량을 예측하여 선행연구보다 스마트하고 진보한 조명 제어를 가능하게 하는 인공지능 기반의 에너지 소비 예측 알고리즘을 추가적으로 설계하였다. 본 모델은 에너지 소비 패턴을 학습할 수 있는 데이터 수집 및 처리 시스템의 구현과 예측 모델을 기반으로 조명을 자동으로 제어하는 기능을 통하여 실내외 환경 요소를 반영한 에너지 소비 최적화를 위함이다. 본 연구에서 에너지 소비 패턴에 영향을 미치는 주요 데이터를 수집하였으며 항목은 다음과 같다.

(표 4) 에너지 소비 패턴의 영향을 미치는 데이터

no	항목	내용
1	시간대	조명이 켜지는 시간대는 주간, 야간, 새벽에 따라 소비 패턴이 달라진다.
2	요일	주중과 주말의 소비 패턴은 다르며, 공휴일 등의 특수 일정도 고려해야 한다.
3	계절	계절에 따라 실내외 조도 차이가 발생하므로, 이 또한 에너지 소비에 영향을 미친다.
4	내외부 기온 및 날씨	날씨가 흐리거나 비가 오는 경우, 자연광의 부족으로 인해 더 많은 조명이 필요할 수 있다.
5	실내 인원	실내에 있는 사람의 수가 많을수록 조명 필요량이 증가한다.
6	과거 소비 기록	과거의 조명 사용 기록은 미래의 소비 패턴을 예측하는 데 중요한 자료로 사용된다.

영향을 미치는 데이터 수집에 대한 입력 데이터 벡터 X는 각 변수 x_i 의 집합으로 구성되며, 이를 일반화하면 다음과 같다.

$$X = x_1, x_2, \dots, x_n \quad (1)$$

여기서 x_1 은 시간대, x_2 는 요일, x_3 는 실내 인원 수 등을 나타내며 이들은 모두 에너지 소비량에 영향을 미치는 변수이다.

에너지 소비 예측 알고리즘은 다음과 같이 정의한다.

$$\hat{E} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \quad (2)$$

여기서 에너지 소비량 E는 독립 변수 X와 종속 변수 E 간의 관계로 예측된 에너지 소비량을 표현하고 있다.

2.6 플랫폼 제어 테스트

앱 콘트롤러를 이용한 프로토타입 테스트를 진행하였다. 컨버터에 LED 조명을 연결하고 앱과 통신을 연동하여 조명의 제어가 정상적으로 동작하는지 실험을 하였으며 최종적으로 정상적으로 기능들이 동작하는 것을 확인할 수 있었다. 본 플랫폼은 초고층 지능형 빌딩의 실내 구역 환경에서 관리되도록 설계되었으며 최종적으로 기능의 안정화와 UI의 개선, 네트워크 통신 장애 문제 최소화 등을 점검하여 양산이 가능할 것으로 예상된다.



(그림 4) 컨버터와 LED 조명 연결 및 테스트 셋팅



(그림 4) 그룹 추가 및 테스트

III. 결론

본 논문에서는 조명 제어 시스템에 인공지능 기반 에너지 소비 예측 알고리즘을 추가하여, 에너지 효율성을 높이고 관리 체계를 향상시키는 방법을 제시하였다. 즉, 다양한 환경 데이터를 수집하고 알고리즘을 적용하여 미래의 에너지 소비량을 예측함으로써, 실시간으로 조명을 자동으로 제어하는 시스템을 개발하였다. 본 연구를 기반으로 건물 내부 에너지 사용량을 줄이고, Green-IT에 기여할 수 있을 것으로 예상된다.

참 고 문 헌

[1] 이윤주, 서지훈, 최진탁. (2014). 모바일 APP 플랫폼 기반 스마트 그리드 솔루션 통합 체계에 관한 연구. 융복합지식학회, 2(2) 21-27.

- [2] 조혜진, 이윤주, 최진탁. (2014). 사물인터넷통신을 기반으로 한 LED 커버터 제어 및 관리 시스템에 대한 설계. 융복합지식학회, 3(1) 33-39.

2024년도 (사)용복합지식학회

추계학술발표대회 포스터 세션

텍스트 마이닝을 통한 소비자의 ESG 인식과 접근성 연구: 점자 표기를 중심으로

김준환, 임석진*
 서울시립대학교, *성결대학교
 kjh715@uos.ac.kr, *imseokjin@gmail.com

A Study on Consumer Perception and Accessibility of ESG through Text Mining: Focused on Braille Labeling

Kim Joon Hwan, Im Seokjin*
 University of Seoul, *Sungkyul Univ.

요 약

본 연구는 텍스트 마이닝과 설문조사를 활용하여 비장애인 소비자의 ESG(Environmental, Social, Governance) 인식과 점자 표기를 통한 정보 접근성에 대한 반응을 분석하는 것을 목적으로 한다. 이를 바탕으로 소비자의 긍정적인 ESG 인식이 정보 접근성 개선, 특히 점자 표기 도입의 필요성과 중요성에 대한 인식을 높이는 데 기여하는지 확인하고자 한다. 본 연구에서는 온라인 데이터를 텍스트 마이닝 기법으로 분석하여 ESG 관련 주요 키워드와 감정적 반응을 도출하고, 설문조사를 통해 소비자의 심층적인 태도를 조사할 예정이다. 이러한 연구결과는 기업과 정부가 ESG 정보를 보다 포괄적으로 제공하고, 점자 표기 등 접근성 향상 방안을 도입하는 데 실질적인 시사점을 제시할 것이다. 또한 기업이 사회적 책임을 다하고 브랜드 이미지를 강화할 수 있는 전략적 방안을 모색할 수 있을 것으로 기대한다.

I. 서론

ESG(Environmental, Social, Governance)는 기업이 환경적, 사회적 책임을 다하고 투명한 경영을 통해 지속 가능한 성장을 추구하는 경영 철학이다. 현대 사회에서 기업들은 환경, 사회, 지배구조(ESG) 경영의 중요성을 인식하고, 이를 통해 지속 가능한 성장을 도모하고 있다. 특히, 취약계층에 대한 접근성 확대는 ESG 경영의 중요한 부분 중 하나로 부각되고 있다. 그 중에서도 시각장애인을 위한 점자 표기는 기업의 사회적 책임을 보여주는 주요 수단으로 자리 잡고 있으며, 점자 표기의 의무화가 일부 제품군에서 점차 확대되고 있다. 예를 들어, 2017년부터 시각장애인의 정보 접근성을 향상시키기 위해 ‘아이스시 8.0’과 ‘칠성사이다’에 브랜드명을 점자로 표시하는 노력이 시작되었다[1]. 또한 최근에는 화장품 용기나 포장에 점자 및 음성·수어영상변환용 코드를 표시하는 것을 의무화하는 법안이 발의되었다[2]. 이러한 점자 표기는 시각장애인에게 정보를 제공하는 중요한 도구로 기능하며, 사회적 배려와 포용성을 상징하는 요소로서의 의미를 가지고 있다. 그러나 ESG 인식 및 수용 정도는 다양한 계층과 사회적 배경에 따라 상이할 수 있으며, 특히, 시각 장애인을 비롯한 비장애인 소비자의 접근성을 고려한 연구는 미비한 실정이다.

점자 표기는 시각 장애인을 위한 중요한 정보전달 수단으로, 이들의 정보 접근권을 보장하는 데 중요한 역할을 한다. 따라서 본 연구는 점자 표기를 ESG관련 정보전달 방식으로 활용하고, 이를 비장애인 소비자들 어떻게 인식하고 수용하는지를 텍스트 마이닝 분석을 통해 알아보고자

한다. 이를 통해 소비자의 ESG에 대한 인식의 범위와 수용성에 관한 새로운 시사점을 제공할 수 있을 것으로 기대된다. 이러한 연구는 기업이 사회적 책임을 다하는 동시에 비장애인 소비자의 요구를 충족시키기 위한 마케팅 전략을 수립하는 데 중요한 시사점을 제공할 것이다.



그림 1. 용기제품 점자 표시(출처: 오투기 홈페이지)

II. 관련연구

2.1 ESG의 개념 및 연구동향

환경, 사회 및 거버넌스(ESG) 원칙은 다양한 부문에서 지속 가능성을 증진하기 위한 필수 프레임워크로 점점 더 인식되고 있다. ESG는 환경 관리, 사회적 책임, 효과적인 거버넌스를 통합하여 조직을 지속 가능한 관행으로 이끌고 경쟁력을 강화하고 있다[3]. 이와 관련하여 ESG 경영과 관련된 연구들은 주로 투자자 관점에서 ESG 요소가 기업 가치에 미치는 영향을 분석하거나, 기업의 지속 가능성 보고서와 같은 자료를 활용하여 ESG 전략이 실제로 기업에 미

치는 영향을 다루고 있다[4]. 한편, 소비자 관점에서 ESG 인식에 대한 연구는 점차 증가하고 있지만, 특정 취약 계층을 대상으로 한 연구는 부족한 실정이다. 이는 점차 표기와 관련된 연구는 시각 장애인의 정보 접근성을 중심으로 진행된 경우가 대부분이며, 이를 ESG와 접목한 연구는 거의 이루어지지 않았다.



그림 2. ESG 경영의 핵심요소

2.2 점자 표기의 개념 및 접근

점자는 시각장애인의 의사소통에 필수적인 촉각 문자 체계로, 양각된 점을 통해 읽고 쓸 수 있다. 또한 최근 기술의 발전으로 점자 학습 및 인식 기능이 크게 향상되었다. 예를 들어, 심층 컨볼루션 신경망(DCNN)과 같은 딥러닝 기법은 점자 문자를 인식하는 데 있어 정확도가 99.63%에 달할 정도로 높은 정확도를 달성하였다[5]. 이와 더불어 혁신적인 시각 및 촉각 지각 기법이 개발되어 점자 인식에 더욱 도움이 되고 있다[6]. 이러한 발전은 기술을 점자 교육 및 인식에 통합하여 시각 장애인의 포용성과 독립성을 증진하는 것이 얼마나 중요한지 나타냈지만, 교육 환경에서 효과적으로 사용할 수 있도록 이러한 기술과 교육에 대한 광범위한 액세스를 보장하는 데에는 여전히 과제가 남아 있다.

자음	ㄱ	ㄴ	ㄷ	ㄹ	ㅁ	ㅂ	ㅅ	ㅇ	ㅈ	ㅊ	ㅋ	ㆁ	ㅅ	ㅇ
초성	●○	●●	●○	○●	○●	●●	●●	○●	○●	○●	○●	○●	○●	○●
중성	○●	○●	○●	○●	○●	○●	○●	○●	○●	○●	○●	○●	○●	○●
종성	○●	○●	○●	○●	○●	○●	○●	○●	○●	○●	○●	○●	○●	○●

그림 3. 한글 자음의 점자(출처: 점자세상)

한편, 점자 표기는 시각장애인에게 제품 정보와 사용 방법을 전달하는 중요한 수단이다. 이는 단순한 정보 제공을 넘어, 시각장애인에게 사회적 소외감을 줄이고 일상생활에서의 자립을 지원하는 역할을 한다. 더불어, 기업이 점자 표기를 도입하는 것은 시각장애인을 배려하는 포용적 사회를 만들기 위한 노력이자, 기업이 사회적 책임을 이행하고 있다는 강력한 신호로 받아들여질 수 있다. 소비자에게도 이러한 점자 표기의 도입은 기업의 사회적 가치를 인식하게 하며, 기업에 대한 긍정적인 이미지를 강화할 수 있다.

III. 연구방법

본 연구에서는 텍스트 마이닝 분석과 설문조사를 결합하여 비장애인 소비자의 ESG 인식과 점자 표기 관련 반응을 종합적으로 분석하고자 한다. 다양한 온라인 데이터를 분

석하여 소비자의 ESG 인식 및 접근성에 대한 의견을 도출하고, 설문조사를 통해 심층적인 인식 및 접근성에 대한 태도 차이를 보완적으로 분석하는 방식으로 진행할 것이다.

구체적으로 텍스트 데이터는 ESG와 관련된 온라인 뉴스 기사, 블로그, 소셜 미디어 게시물 등에서 수집할 것이다. 특히, 점자 표기와 관련된 이슈를 포함한 게시물과 댓글을 중심으로 데이터를 수집하여 소비자의 반응을 분석할 예정이다. 수집된 데이터는 'ESG', '점자 표기', '접근성' 등 관련 키워드를 포함한 내용을 중심으로 크롤링 방식으로 수집하며, 데이터 수집 기간은 최근 3년간의 자료를 대상으로 한다. 수집된 데이터는 자연어 처리(NLP) 기법을 사용하여 불필요한 단어 제거(불용어 처리), 텍스트 정규화(소문자 변환, 특수문자 제거 등), 토큰화 및 어간 추출 등의 과정이 포함된다. 또한 전처리된 데이터를 바탕으로 키워드 분석을 진행하여 소비자가 ESG와 관련하여 가장 많이 언급하는 주제 및 핵심 키워드를 파악할 것이다. 이를 통해 ESG에 대한 주요 관심사와 인식을 확인할 수 있다.

추가적으로 감정분석(sentiment analysis)을 통해 비장애인 소비자의 ESG 정보와 점자 표기에 대한 감정적 반응(긍정, 부정, 중립)을 분류하여 수용성에 대한 전반적인 감정 흐름을 파악하고, 군집분석(cluster analysis) 기법을 사용하여 소비자의 반응을 그룹화하고, 각 그룹의 특성을 분석할 것이다. 이를 바탕으로 사회적 배경, 관심사 등에 따른 소비자의 ESG 인식 및 점자 표기에 대한 반응 차이를 도출할 수 있을 것이다.

IV. 결론

본 연구는 텍스트 마이닝 기법을 활용하여 소비자의 ESG 인식과 점자 표기를 통한 접근성에 대한 수용도를 분석하는 것을 목적으로 진행하고자 한다. 본 연구결과를 바탕으로 ESG 정보 제공 및 접근성 개선에 대한 몇 가지 제언을 다음과 같이 제시할 수 있을 것이다.

첫째, 기업과 정부는 ESG 관련 정보를 제공할 때 모든 계층이 쉽게 접근할 수 있도록 점자 표기와 같은 접근성 향상 방안을 적극적으로 도입해야 한다. 특히, 웹사이트 및 보고서 등 다양한 ESG 자료에 접근성 옵션을 포함하는 것이 필요하다. 이를 통해 시각 장애인을 포함한 취약 계층이 ESG 정보를 손쉽게 접할 수 있도록 해야 한다. 둘째, ESG는 단순히 기업의 이미지 제고를 위한 수단이 아니라, 모든 이해관계자를 포용하는 전략으로 접근해야 한다. 따라서 기업은 ESG 경영에 있어 취약 계층의 목소리를 반영하는 포용적 전략을 수립해야 하며, 이를 위해 소비자 설문조사나 의견 청취와 같은 방법을 통해 실질적인 피드백을 수집하고 이를 정책에 반영해야 할 것이다. 셋째, ESG 정보 제공의 표준화와 접근성 개선을 위한 정부 차원의 정책적 지원이 필요하다. 정부는 ESG 경영을 촉진하는 동시에, 정보 접근성 향상을 위한 가이드라인을 마련하여 기업들이 이를 준수할 수 있도록 해야 한다. 이를 기반으로 ESG 정보의 공정한 제공과 접근성이 보장될 수 있을 것이다.

따라서 기업은 점자 표기를 단순히 사회적 책임의 이행

수단으로만 볼 것이 아니라, 제품 디자인과 기능성을 고려하여 전략적으로 도입할 필요가 있다. 점자 표기를 통해 기업의 ESG 경영 가치를 극대화하면서도 소비자들의 다양한 요구를 충족시키는 마케팅 전략을 수립하는 것이 중요하다. 이러한 접근은 기업의 사회적 이미지를 강화하고, 장기적으로 비장애인 소비자의 브랜드 충성도를 높이는 데 기여할 수 있을 것으로 기대한다.

참 고 문 헌

- [1] <https://v.daum.net/v/20210420204317807>
- [2] <https://v.daum.net/v/20240924184448361>
- [3] 허종호, 이가은, 김소연, 김윤, 박지민, 최고은. “마케팅 분야 ESG 연구의 동향분석 및 향후 연구방향”, Entrepreneurship&ESG 연구, 제4권, 제1호, pp. 1-23, 2024.
- [4] Patel, A., & Gloor, P. A. “Creating a systematic ESG (Environmental Social Governance) scoring system using social network analysis and machine learning for more sustainable company practices”, In Handbook of Social Computing (pp. 265-278). Edward Elgar Publishing, 2024.
- [5] Rahmat, R. F., Purnamawati, S., Mardianto, W., Faza, S., Sulaiman, R., Nadi, F., & Lubis, A. R. “Braille letter recognition in deep convolutional neural network with horizontal and vertical projection”, Bulletin of Electrical Engineering and Informatics, 13(5), pp. 3380-3391, 2024.
- [6] Park, B. S., Im, S. M., Lee, H., Lee, Y. T., Nam, C., Hong, S., & Kim, M. G. “Visual and tactile perception techniques for braille recognition”, Micro and Nano Systems Letters, 11(1), 23, 2023.

인간 혈청 프로테오믹스를 이용한 근감소증 진단을 위한 AI를 활용한 바이오마커 개발

노시형¹, 임동욱¹, 정창원^{1,2}

¹원광대학교 근감소증토탈솔루션 센터

²원광대학교병원 정보관리실 스마트 사업팀

{nosij123, dwl316, mediblu}@wku.ac.kr

Development of AI-Based Biomarkers for Sarcopenia Diagnosis Using Human Serum Proteomics

SiHyeong Noh¹, Dong-Wook Lim¹, Chang-Won Jeong^{1,2}

¹Sarcopenia Total Solution Center, Wonkwang University,

²Smart Business Team of Wonkwang University Hospital

요 약

근감소증(Sarcopenia)은 근육량과 근기능의 감소가 진행되는 골격근 질환으로, 낙상, 기능 저하, 허약, 그리고 사망률 증가와 같은 부작용을 동반한다. 본 연구에서는 임상 현장에서 근감소증의 진단 기준에 따라 정상군(n=25), 근감소증군(n=23), 중증 근감소증군(n=24)으로 환자를 분류하였다. 해당 환자의 혈액 분석을 통해 총 5,420개의 단백질 데이터를 수집하였으며, 이를 바탕으로 근감소증 진단에 유의미한 단백질 패턴을 탐색하고, 잠재적 바이오마커를 발굴하기 위해 5가지 인공지능(AI) 모델을 적용하였다. 아울러, 생성된 AI 모델의 성능을 평가하고, 모델 개발 과정에서 발생한 오류와 한계를 분석하여 이를 해결하기 위한 방안을 제시한다.

I. 서 론

근감소증(Sarcopenia)은 근육량과 근력의 감소를 특징으로 하는 질환으로, 노화와 밀접한 관련이 있으며, 고령화 사회에서 주요한 건강 문제로 대두되고 있다. 근감소증은 삶의 질 저하, 신체적 장애, 그리고 심혈관 질환 및 대사 질환과 같은 다양한 건강 문제를 유발할 수 있어 조기 진단과 효과적인 치료가 매우 중요하다. 그러나 근감소증의 진단은 현재 European Working Group on Sarcopenia in Older People (EWGSOP), 아시아 근감소증 작업그룹(AWGS) 등의 국제 가이드라인에 따라 진단지표를 따르고 있다. 그러나 주관적인 요소가 개입될 수 있고 표준화된 진단 기준이 부족하다는 문제가 있다[1-3].

이러한 한계를 극복하기 위해, 단백질체학(Proteomics) 기술을 활용한 혈청 기반의 바이오마커 개발이 주목받고 있다. 특히, 혈청 프로테오믹스는 혈액 내 단백질의 변화를 정밀하게 분석하여 질병의 발병 메커니즘을 이해하고, 조기 진단에 활용할 수 있는 잠재력을 가지고 있다[4]. 인간 혈청은 질병 상태에 따라 단백질 구성과 농도가 변화하기 때문에, 이를 분석함으로써 근감소증의 특이적 단백질을 식별하고, 바이오마커로서의 가능성을 확인할 수 있다.

본 연구에서는 인간 혈청 프로테오믹스 데이터를 기반으로 근감소증을 조기 진단할 수 있는 AI 기반 바이오마커를 개발하고자 한다. AI 기술은 복잡한 데이터 패턴을 분석하고 예측할 수 있는 강력한 도구로, 프로테오믹스 데이터에서 의미 있는 바이오마커를 효과적으로 발굴하고 진단

모델을 구축하는 데 용이하다. 특히, 머신러닝 및 딥러닝 알고리즘을 활용하여 대규모의 단백질 데이터를 분석하고, 근감소증과 관련된 중요한 단백질 프로파일을 자동으로 식별할 수 있다.

본 연구의 목표는 인간 혈청 프로테오믹스 데이터를 활용하여 근감소증의 조기 진단을 위한 신뢰성 높은 AI 기반 바이오마커를 개발하고자 한다. 이를 통해 근감소증 진단의 정확성을 향상시키고, 조기 개입 및 맞춤형 치료가 가능하도록 하여, 고령화 사회에서 중요한 건강 문제 해결에 기여하고자 한다.

본 논문에서는 5,420종의 혈액 분석 데이터를 기반으로 근감소증 진단을 위한 주요 바이오마커를 식별하고, 이를 활용한 인공지능(AI) 모델을 개발하는 것을 목표로 한다. 이를 위해 5가지 AI 알고리즘을 적용하여 모델의 성능을 평가하고, 각 모델의 성능과 특징을 비교·분석하여 최적의 진단 모델을 도출하고자 한다. 또한, 개발된 AI 모델의 임상적 적용 가능성과 향후 연구 방향에 대해 고찰할 예정이다.

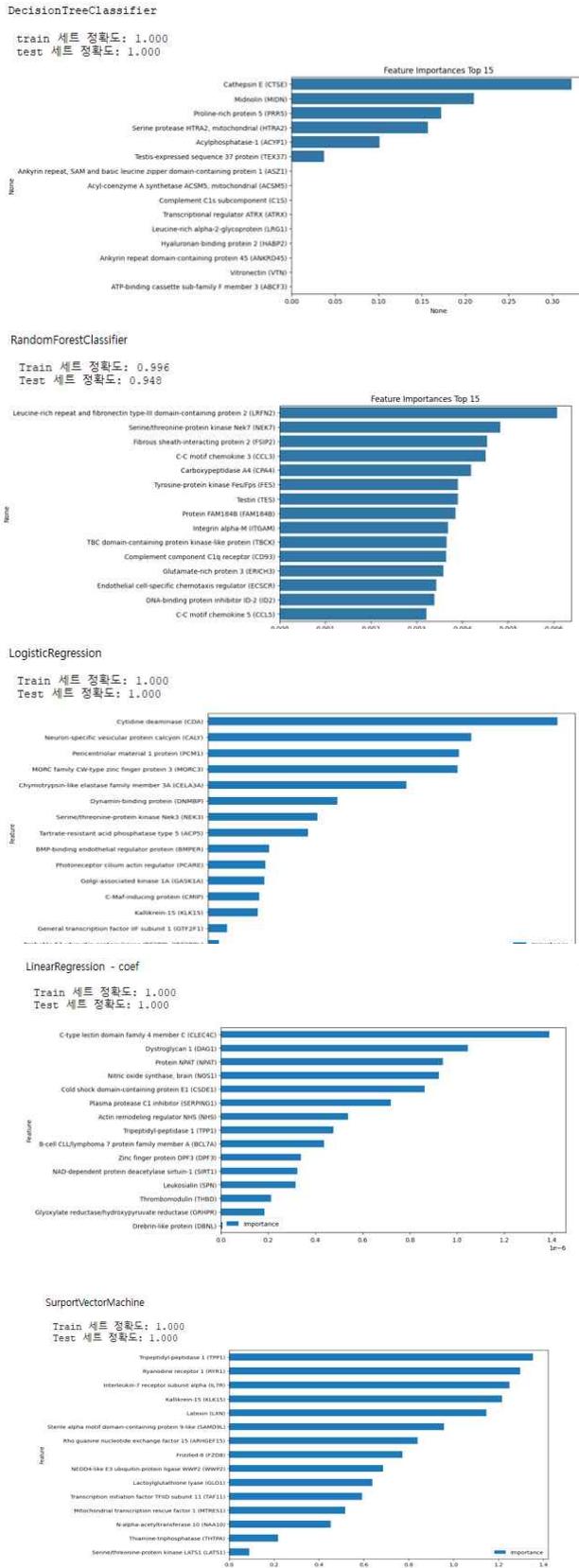
II. 인간혈청 프로테오믹스기반 근감소증 진단모델

2.1 데이터 수집 방법

데이터를 수집하기 위해서 원광대학교 병원 IRB (Institutional Review Board)의 승인을 받은 전향적 연구를 시행하였다(WKUH2023-08.31). 환자는 아시아 분류 척도(AWGS)를 사용하여 근육량, 악력 그리고 신체기능 평

2.4 주요 인자 추출

각 인공지능 모델을 수행한 결과의 주요 Feature를 추출한 결과를 다음 그림 6과 같다.



(그림 6) Feature 추출을 위한 Multi Model 학습 결과

과적합이 발생한 이유 중 5,420개가 넘는 input이 이유가 될 수 있다는 판단하에 학습데이터의 주요 Feature를 추출하기 위해 5가지 모델을 학습하고 그 결과 그림 6과 같은 Feature를 획득할 수 있었다. Decision Tree는 6개의 Feature에 대해서만 의미가 있다고 평가하였으며, Random Forest와 Logistic, Linear Regression에서는 앞서와는 다른 15가지의 Feature가 추출된 것을 확인할 수 있다.

(표 1) 추출된 Feature의 ID

Decision Tree	Random Forest	Linear Regression	Logistic Regression	Support Vector Machine
P14091	Q9ULH4	Q8WTT0	P32320	O14773
Q504T8	Q8TDX7	Q14118	Q9NYX4	P21817
P85299	Q5CZC0	Q14207	Q15154	P16871
O43464	P10147	P23352	Q14149	Q9H2R5
P07311	Q9UI42	O75534	P09093	Q9BS40
Q96LM6	P07332	P05155	Q6XZF7	Q8IVG5
	Q9UGI8	Q6T4R5	P51956	O94989
	Q9ULE4	O14773	P13686	Q9H461
	P11215	Q4VC05	Q8N8U9	O00308
	Q8TEA7	Q92784	A6NGG8	Q04760
	Q9NPY3	Q96EB6	Q9UFP1	Q15544
	Q5RHP9	P16150	Q8IY22	Q9P0P8
	Q19T08	P07204	Q9H2R5	P41227
	Q02363	Q9UBQ7	P35269	Q9BU02
	P13501	Q9UJU6	Q9H1B7	O95835

이 결과 각 모델에서 주요 특징으로 추출된 단백질을 표1에서 보이고 있다. 이들 중 특정 단백질(TPPI, KLK15)이 여러 모델에서 중요한 Feature로 확인되었다. 하지만 본 모델은 과적합된 결과여서 새로운 데이터에 대한 일반화 능력이 떨어지는 문제점을 갖고 있다.

따라서 과적합 방지를 위해서 더 많은 데이터를 수집하거나 K-fold 교차 검증과 같은 방법을 활용하거나 SMOTE (Synthetic Minority Over-sampling Technique)과 같은 기법을 통해 데이터 균형과 가중치 조절을 적용해야 할 계획이다.

III. 결론

우리는 근감소증 환자를 정상과 근감소증 그리고 심한 근감소증 환자로 분류하고 5,420개의 혈액 분석 결과를 통해 근감소증 진단을 위한 바이오마커를 생성하기 위해 인공지능 학습을 통해 모델을 생성하기 위한 연구를 진행하였다. 과적합과 데이터 불균형 문제 해결을 위한 연구를 진행할 계획이다.

ACKNOWLEDGMENT

본 연구는 과학정보통신부의 재원으로 한국연구재단(NRF-2021R1A5A8029876) 지원에 의하여 이루어진 것임.

참고 문헌

[1] Cruz-Jentoft, Alfonso J., and Avan A. Sayer. "Sarcopenia." The Lancet 393.10191 (2019): 2636-2646.

- [2] Lian, Rongna, et al. "Blood biomarkers for sarcopenia: A systematic review and meta-analysis of diagnostic test accuracy studies." *Ageing Research Reviews* (2023): 102148.
- [3] Silva-Fhon, Jack Roberto, et al. "Sarcopenia and blood albumin: A systematic review with meta-analysis." *Biomedica* 41.3 (2021): 590-603.
- [4] Picca, Anna, et al. "Biomarkers shared by frailty and sarcopenia in older adults: A systematic review and meta-analysis." *Ageing research reviews* 73 (2022): 101530.

디지털 디바이스 활용 능력에 따른 프라이버시 인식에 관한 연구

임석진
성결대학교
imseokjin@gmail.com

A Study on Privacy Perception by Digital Device Literacy

Im Seokjin*
Sungkyul Univ.

본 연구는 디지털 디바이스 활용 능력과 프라이버시 우려 간의 관계를 분석하고자 하였다. 디지털 디바이스 사용이 증가함에 따라 개인정보 침해와 무단 추적과 같은 프라이버시 문제가 더욱 중요해지고 있다. 이를 위해 본 연구는 2023년 한국미디어패널조사 데이터를 활용하여 디지털 디바이스 활용 능력이 높은 사용자가 프라이버시 침해를 어떻게 인식하는지 실증적으로 분석하였다. 분석 결과, 디지털 활용 능력이 높은 사용자는 프라이버시 침해에 대한 우려가 크고, 보호 조치를 더 잘 취하는 것으로 나타났다. 이는 디지털 활용 교육이 프라이버시 인식과 보안 실천 향상에 중요한 역할을 할 수 있음을 시사한다.

I. 서론

디지털 디바이스의 보급과 인터넷 사용의 증가로 인해 사용자들은 다양한 온라인 서비스에 접근할 수 있게 되었다. 이 과정에서 개인정보 보호 문제는 중요한 사회적 이슈로 부각되고 있으며, 특히 개인정보 도용, 해킹, 무단 추적 등 디지털 환경에서의 프라이버시 침해 우려가 심화되고 있다 [1].



그림 1. 디지털 디바이스와 프라이버시

그러나 디지털 디바이스 활용 능력이 높은 사용자는 이러한 프라이버시 우려에 대해 보다 명확한 인식을 가지고 적절히 대응할 수 있을 것이라는 가정이 존재한다. 본 연구는 디지털 디바이스 활용 능력이 높은 사용자들이 프라이버시 우려에 대해 어떻게 인식하고 있는지를 규명하고, 두 변수 간의 관계를 실증적으로 분석하는데 목적이 있다. 이를 위해 연구문제를 다음과 같이 설정하였다:

<연구문제 1> 디지털 디바이스 활용 능력은 프라이버시 인식에 어떤 영향을 미치는가?

<연구문제 2> 디지털 디바이스 활용 능력 수준에 따라 프

라이버시 우려는 어떤 차이를 보이는가?

II. 이론적 배경 및 가설설정

2.1 디지털 디바이스 활용 능력

디지털 디바이스 활용 능력은 사용자가 컴퓨터, 스마트폰 등 다양한 디지털 기기를 효과적으로 사용하고, 문제 해결 능력 및 정보처리 능력을 갖추고 있는 정도를 의미한다. 이전 연구에 따르면 디지털 디바이스 활용 능력이 높은 사용자는 디지털 환경에서 발생하는 보안 위험을 인지하고, 적절한 보호 조치를 취할 가능성이 높다[2].

2.2 프라이버시 인식

프라이버시 인식은 개인이 자신을 둘러싼 정보가 보호되고 있는지에 대해 느끼는 인식 정도를 의미한다[3]. 온라인 환경에서의 프라이버시 인식은 데이터 수집, 저장, 활용의 투명성과 이에 대한 개인의 통제권과 밀접한 관련이 있다. 선행연구에서는 온라인에서 개인정보가 수집되는 과정에 대한 인식이 높을수록 사용자들이 프라이버시에 대한 우려를 더 크게 가지는 것으로 나타났다[4][5].

2.3 가설설정

앞서 논의한 선행연구를 토대로 설정한 연구모형을 제시하면 아래와 같다.

가설 1: 디지털 디바이스 활용 능력이 높은 사용자는 낮은 사용자에 비해 프라이버시 침해에 대한 우려가 높을 것이다.

가설 2: 디지털 디바이스 활용 능력이 높은 사용자는 개인 정보 보호 조치를 더 잘 인식하고 적용할 것이다.

가설 3: 프라이버시 우려가 높은 사용자일수록 디지털 디

바이스 활용 능력이 높을 것이다.

III. 실증분석

3.1 연구 대상 및 자료

본 연구에서는 위와 같은 연구 질문을 확인하기 위해 정보통신정책연구원의 국가승인통계(승인번호 : 405001호)인 2023년 한국미디어패널조사의 원시데이터를 제공받아 실증 분석을 하고자한다. 해당 패널은 전국 4,077가구의 약 9,757명을 대상으로 하였으며, 미디어 기기 사용과 프라이버시 인식에 대한 종단 조사 데이터를 포함하고 있다. 이 중 디지털 디바이스 활용 능력 및 프라이버시 침해 우려에 대한 응답 데이터를 추출하여 분석에 사용하였다.

3.2 변수 측정

디지털 디바이스 활용 능력은 설문 응답 중 PC 및 스마트 기기 활용 능력 문항을 통해 측정하였다. 각 응답자는 기기 설정, 인터넷 검색, 금융 거래 등 다양한 활동에 대한 수행 능력을 리커트 5점 척도로 평가하였다. 프라이버시 인식은 프라이버시 침해 우려, 개인정보 수집에 대한 걱정, 온라인에서의 개인정보 도용 우려 등과 관련된 문항을 통해 측정하였다. 또한 연령, 성별, 소득 수준, 학력 등을 통제 변수로 설정하여 상관관계 분석 시 변수의 영향을 고려하였다.



그림 2. 분석을 위한 측정 변수

IV. 결론

본 연구는 디지털 디바이스 활용 능력이 프라이버시 인식에 중요한 영향을 미친다는 점을 실증적으로 밝혔다. 디지털 디바이스 활용 능력이 높을수록 프라이버시 침해에 대한 우려가 커지며, 이로 인해 사용자들은 더욱 철저한 보호 조치를 취하는 것으로 나타났다. 이러한 결과는 디지털 디바이스 활용 교육이 프라이버시 보호와 보안 인식 향상에 기여할 수 있음을 시사한다. 추후 연구에서는 세부 기기별 활용 능력 차이와 프라이버시 인식의 미시적인 차이를 더 깊이 분석할 필요가 있다. 따라서 본 연구는 디지털 디바이스 사용 교육을 강화하여 프라이버시 보호 인식을 높이고, 실제적인 보안 조치 사용을 권장할 필요가 있다. 또한 프라이버시 보호를 위한 소프트웨어 사용법 및 데이터 관리 방법에 대한 대중 교육이 요구되는 것을 함의한다.

참 고 문 헌

- [1] 류성진, 고흥석. “디지털 리터러시가 정보 프라이버시 염려에 미치는 상대적 영향: 미디어 기기 이용역량과 디지털 원주민-이주민 집단 간 비교를 중심으로”, 한국방송학보, 제35권, 제6호, pp. 149-186, 2021.
- [2] 김현정, 김범수. “스마트기기 활용역량과 프라이버시 우려: 온라인 사회참여 활동과 정보 사실성 판단 능력의 매개효과 및 프라이버시의 조절효과”, 지식경영연구, 제24권, 제1호, pp. 51-72, 2023.
- [3] Trepte, S., Teutsch, D., Masur, P. K., Eicher, C., Fischer, M., Hennhöfer, A., & Lind, F. “Do people know about privacy and data protection strategies? Towards the “Online Privacy Literacy Scale”(OPLIS)“, Reforming European data protection law, pp. 333-365, 2015.
- [4] 권혁진. “온라인 활동과 OTT 서비스 이용에 따른 프라이버시 우려의 인식에 관한 연구: 구매성향의 상호작용효과를 중심으로”, e-비즈니스연구, 제25권, 제4호, pp. 57-69, 2024.
- [5] 임문선, 정남호, “디지털 리터러시와 정보 프라이버시 우려가 정보의 비판적 이해에 미치는 영향: 보호동기 이론과 감각추구 성향의 적용”, 인터넷전자상거래연구, 제24권, 제4호, pp. 1-25.

분산 환경에서 메타데이터 관리를 위한 MapReduce 기반의 Fabricator

윤채연*, 문석재**
광운대학교 스마트융합대학원
{dbscodus, msj8086}@kw.ac.kr

Fabricator based on MapReduce for metadata management in distributed environments

Chae-Yeon Yun*, Seok-Jae Moon**
Graduate School of Smart Convergence, Kwangwoon Univ.

요 약

본 논문은 데이터 관리 및 접근을 일관되고 통합된 방식으로 제공하는 메타데이터 레지스트리 구축 기술인 Fabricator 시스템을 제안한다. Fabricator 시스템은 메타데이터 스키마의 추가 및 변경을 용이하게 관리하여, 분산 환경에서의 유연한 메타데이터 관리를 가능하게 한다. 이 시스템의 핵심 요소는 MapReduce를 활용해 설계된 매칭 도구인 CMR Lexi로, 스키마 매칭 과정을 효율적으로 지원한다. Fabricator는 다양한 데이터 소스를 실시간으로 통합하여 데이터의 유연성과 사용성을 극대화하는 것을 목표로 하며, 이를 통해 분산 환경 내에서의 원활한 접근과 운영 효율성을 높인다. 이러한 접근 방식은 이질적인 데이터를 관리하는 복잡성을 해결하고, 끊임없이 변화하는 데이터 환경에서의 확장성과 적응성을 지원한다.

I. 서 론

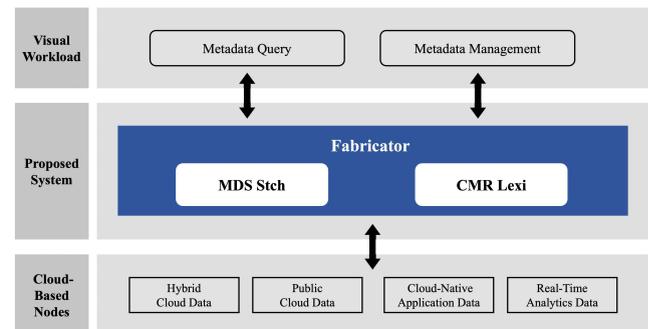
분산 환경에서 데이터 패브릭은 다양한 데이터 소스와 형식을 통합하여 일관된 데이터 관리 및 접근을 제공하는 혁신적인 기술이다[1]. 이러한 기술은 기업들이 분산된 데이터 환경에서 데이터를 효율적으로 관리하고 활용할 수 있도록 지원하여, 유의미한 인사이트를 도출하고 비즈니스 가치를 극대화하는 데 기여한다[2]. 데이터 패브릭을 구축하기 위해서는 지리적으로 분산된 기업이 보유한 데이터 소스들을 통합할 때 데이터 일관성을 유지하고 데이터 거버넌스 체계를 확립하는 것이 필수적이다. 전체 데이터 소스에 대한 메타데이터 기반의 스키마 설계는 상당히 복잡한 작업일 수 있으며, 정리 과정 또한 많은 어려움을 동반한다. 각 데이터 소스의 스키마를 분석하고 이를 통합된 스키마와 매칭하여 일관된 뷰를 제공해야 하며, 이 과정에서 메타데이터 스키마 관리 도구의 성능과 매칭 품질도 신중하게 고려해야 한다[3]. 따라서, 효과적인 메타데이터 관리 시스템이 필요하다.

본 논문에서는 메타데이터 레지스트리를 구축하여 각 데이터 소스의 메타데이터를 중앙에서 효과적으로 관리하고, MapReduce 기술을 활용하여 성능을 향상시키며, 매칭 전의 스키마 구조를 점검하여 새로운 메타데이터가 신속하게 추가되거나 업데이트될 수 있는 환경을 조성한다. CMR Lexi라는 매칭 도구를 설계하여 메타데이터 스키마 요소들을 저장하는 데 효과적인 매칭 로직을 적용한 구조를 제안하며, 이러한 방식은 데이터 패브릭의 효율성을 높여 기업이 다양한 데이터 소스를 통합하여 최대한의 가치를 창출할 수 있도록 지원할 것이다.

II. 본문

2.1 System Overview

본 논문에서는 분산 환경에서 메타데이터 관리를 위한 MapReduce 기반의 Fabricator 설계 시스템을 제안한다. Fabricator 설계 시스템은 크게 3계층으로 구성한다.



(그림 1) 제안 시스템 구성 요소

최상위 계층인 Visual Workload 계층은 사용자가 시스템에 요청하는 메타데이터 쿼리와 메타데이터 관리를 처리하는 역할을 하며, 여기서 사용자에게 의한 시각적 작업이 이루어진다. 중간 계층인 Proposed System 계층은 시스템의 핵심 기능을 담당하며, 메타데이터 스키마의 제어 및 조정을 수행하는 MDS-Stch(MetaData Schema-Stitch)와 메타데이터 매칭 및 관리 기능을 제공하는 CMR-Lexi(Core MapReduce-Lexus)를 통해 메타데이터의 추가, 변경, 관리 및 매칭을 처리한다. 이 계층은 또한 실시간 데이터 분석과 같은 다양한 작업을 지원하여 시스템

의 효율성을 높인다. 최하위 계층인 Cloud-Based Nodes 계층은 다양한 데이터 소스가 존재하는 클라우드 기반 노드를 나타내며, Hybrid Cloud Data, Public Cloud Data, Cloud-Native Application Data, 및 Real-Time Analytics Data를 포함한다. 이 계층은 데이터 패브릭의 근본적인 인프라를 구성하여, 다양한 환경에서 데이터의 통합 및 처리를 가능하게 한다.

2.2 System Algorithm

알고리즘 1은 메타데이터 관리 및 데이터 통합을 위한 Fabricator 시스템을 구현하며, MetadataRegistry 클래스가 메타데이터 스키마를 관리하고, DataIntegrator 클래스가 데이터 소스를 통합 및 분석하는 기능을 제공한다. Fabricator 클래스는 이 두 클래스를 조합하여 사용자에게 간단한 데이터 스키마 정의 및 분석 인터페이스를 지원한다. 이 알고리즘은 메타데이터 관리와 데이터 통합을 간단하게 수행하여 사용자가 효과적으로 데이터를 활용할 수 있도록 한다.

```

def inAdminMetaInfo(self, location, classification, structure):
    self.metadata_info = {"location": location, "classification":
classification, "structure": structure}
    def reqStdSchInfo(self):
        self.standard_schema_info = {"columns": ["col1", "col2"], "types":
["int", "varchar"]}
    def retStdSchInfo(self):
        return self.standard_schema_info

fabricator = Fabricator()
fabricator.process_schema("User", {"id": "int", "name": "string"})
fabricator.process_schema("Order", {"order_id": "int", "user_id": "int",
"amount": "float"})

user_schema = fabricator.process_query("User")
print("User Schema:", user_schema)

msii = MSIIClassification()
msii.inAdminMetaInfo("location1", "classification1", "structure1")
msii.reqStdSchInfo()
standard_schema_info = msii.retStdSchInfo()
print("Standard Schema Info:", standard_schema_info)
    
```

(알고리즘 1) 제안 시스템 알고리즘

```

class MetadataRegistry:
    def __init__(self):
        self.metadata_schemas = {}
    def add_or_update_schema(self, schema_name, schema):
        action = "Updating" if schema_name in self.metadata_schemas else
"Adding new"
        print(f"{action} schema: {schema_name}")
        self.metadata_schemas[schema_name] = schema
    def query_metadata(self, schema_name):
        return self.metadata_schemas.get(schema_name)

class DataIntegrator:
    def __init__(self):
        self.integrated_data = []
    def integrate_data(self, data_sources):
        for source in data_sources:
            self.integrated_data.extend(source)
        print("Data integration complete.")
    def analyze_data(self):
        print("Analyzing integrated data...")
        return len(self.integrated_data)

class Fabricator:
    def __init__(self):
        self.registry = MetadataRegistry()
        self.integrator = DataIntegrator()
    def process_schema(self, schema_name, schema):
        self.registry.add_or_update_schema(schema_name, schema)
    def process_query(self, schema_name):
        return self.registry.query_metadata(schema_name)
    def integrate_and_analyze(self, data_sources):
        self.integrator.integrate_data(data_sources)
        return self.integrator.analyze_data()

class MSIIClassification:
    def __init__(self):
        self.metadata_info = {}
        self.workflow_process_info = None
        self.standard_schema_info = None
    
```

III. 결론

본 논문에서는 분산 환경에서 메타데이터 관리를 위한 MapReduce 기반의 Fabricator 설계 시스템을 제안하며, 기업의 데이터 통합 과정에서 메타데이터 스키마 관리의 효과적인 매칭 및 성능 향상을 위해 MapReduce 인덱스 구조를 활용하여 문제를 해결하였다. Fabricator는 비구조화된 메타데이터에 중점을 두고 대용량 데이터의 빠른 처리 속도를 제공하며, 다양한 데이터 소스를 실시간으로 통합하여 데이터의 유연성과 사용성을 극대화할 수 있는 가능성을 지닌다. 향후에는 실시간 데이터 처리 및 분석 지원을 통해 다양한 데이터 워크로드를 처리할 수 있도록 높은 성능을 유지하고 최적화 문제를 해결하는 것이 필요하며, 데이터 패브릭 구축 및 유지보수에 따른 비용 또한 고려해야 한다.

ACKNOWLEDGMENT

본 논문은 2024년도 광운대학교 교내 학술 연구비 지원에 의해 연구되었습니다.

참고 문헌

[1] C.-Y. Yun and S.-J. Moon, "A Fabricator Design for Metadata CI/CD in Data Fabric," International Journal of Internet, Broadcasting and Communication, vol. 16, no. 2, pp. 193 - 202, May 2024.

[2] Li, M. Yang, X. Xia, K. Zhang, and K. Liu, "A Distributed Data Fabric Architecture based on Metadata Knowledge Graph," 2022 5th International Conference on Data Science and Information Technology (DSIT). IEEE, Jul. 22, 2022.

[3] V. Sharma, B. Balusamy, J. J. Thomas, and L. G. Atlas, Eds., "Data Fabric Architectures." De Gruyter, May 04, 2023.

처방전 자동 입력을 위한 복용 관리 앱 설계

유동환, 홍정민, 강민석, 한석준*

*신한대학교

{rdwhan, goid987,1212msk}@naver.com, *sukjun1124@gmail.com

Design of a Medication Management App for Prescription Automatic Keyin

Yoo Dong Whan, Hong Jeong Min, Gang Min Seok, Han Seok Jun*

*Shinhan Univ.

요 약

본 논문은 만성질환 고령자들의 복용 순응도를 향상시키고, 복용 누락으로 인한 건강 위험을 최소화할 수 있는 앱을 제공하는 것을 목적으로 한다. 기존의 복용 앱들은 노안으로 스마트폰 사용이 어려운 고령자들이 처방 관련 정보들을 입력하는 불편함이 존재하였으나, 제안 시스템은 이러한 문제점을 개선하여 사용자가 앱에서 처방전을 카메라로 촬영만 하면 처방 관련 정보들을 AI기술을 활용하여 자동 입력되어 환자의 처방 정보 관리 정확도와 사용자의 앱 사용 편의성을 향상시킬 수 있다.

II. 기존 관련 연구

AI를 활용한 알약 인식 서비스를 제공하여 사용자에게 자동 복용 알림을 제공하는 앱에 관한 연구가 이루어졌다 [4]. 이 연구의 특징은 알약 인식을 통한 복용 관리를 돕는 것이다. ‘약 먹을땐(Paprica Care)’ 앱은 복용 알림을 설정해 주며, 여러 약물의 상호작용을 확인할 수 있다. 또한 복용한 약물에 대한 기록을 남기고, 병원에 방문할 때 앱에 저장된 정보를 통해 편리하게 복용 중인 약을 관리할 수 있게 해 준다.

‘커넥트케어’ 앱은 처방전을 스캔하여 약품 이미지와 사용처 등의 정보를 자동 제공한다. 그리고 복용 알림과 상호작용 확인 기능을 제공하며, 약물의 부작용 정보를 제공하여 복용 안전성을 높일 수 있다[5].

제안 ‘복용 관리 앱’은 스마트폰 사용이 불편한 고령 사용자들의 편의성을 높이는 데 중점을 두었다. 특히, 고령 사용자들이 처방전을 사진 촬영함으로써 처방 정보가 자동 입력되어 처리될 수 있도록 지원함으로써 사용자 친화적인 경험을 제공하는 데 목표를 두었다.

III. 본문

3.1 처방전 자동 입력 앱

제안 시스템은 Google Cloud Document AI(GCD AI) 서비스를 통해 처방전 이미지를 분석하여 복용 관리에 필요한 텍스트들을 추출한다. 추출된 텍스트 중 복용 관리에 필요한 정보들은 제안 시스템의 데이터베이스에 자동으로 저장되어 사용자가 직접 입력하는 과정이 생략될 수 있다. 따라서 사용자가 직접 복용 정보를 입력하는 경우에 비해 입력 오류를 최소화할 수 있으며, 사용 편의성 또한 향상

I. 서 론

대한민국은 빠르게 고령화가 진행 중이며, 2024년 기준으로 65세 이상 인구가 전체의 약 19%를 차지하고 있다. 2050년에는 이 비율이 40%를 넘을 것으로 예상된다[1]. 이로 인해 의료 분야에서는 노년층의 건강 관리가 중요한 과제로 떠오르고 있다. 고령자들은 고혈압, 당뇨병, 심혈관 질환 등 만성 질환을 관리하기 위해 여러 약물을 장기 복용해야 하지만, 약물 복용을 제때 하지 못하는 경우가 많다. 설문조사에 따르면, 고령 환자들이 약 복용을 놓치는 주요 이유로 66%가 ‘단순히 잊어버려서’라고 답했다[2]. 이는 치료 효과 저하, 질병 악화, 응급 상황 등 심각한 문제로 이어질 수 있다. 특히 복용 순응도가 낮은 환자는 목표 혈압에 도달할 확률이 낮고, 당뇨병 환자의 경우 혈당 조절에도 직접적인 영향을 미친다.

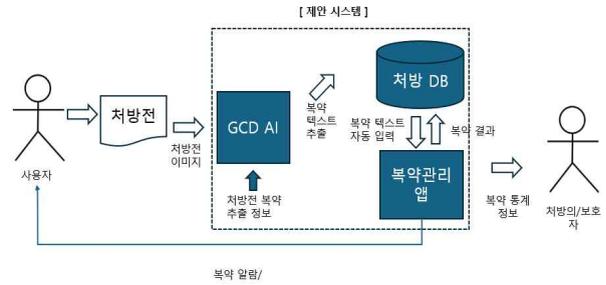
미국에서는 복용 순응도 부족으로 인한 비용이 연간 약 320조 원에 달한다고 추산하고 있다[3]. 이 문제는 환자 개인의 건강뿐만 아니라 가족과 보호자에게도 부담을 준다. 보호자들은 고령자의 복용 상태를 관리해야 하며, 이는 시간과 에너지를 소모하는 일이다.

환자들은 병원에서 진단 후 처방전을 발급받고, 그에 따라 복용을 수행해야 하지만, 고령으로 인한 신체적 불편함과 디지털 기기 사용에 대한 익숙지 않음이 문제로 작용한다. 특히, 노환이나 노안으로 인해 스마트폰을 사용하는 데 어려움을 겪는 환자들은 처방전에 기록된 정보를 정확하게 앱에 입력하는 데 많은 어려움을 느낄 수 있다. 이러한 문제를 해결하고자 본 논문에서는 ‘복용 관리 앱’을 설계하였다.

제안한 앱에서는 사용자가 처방전을 받은 후, 처방에 대한 데이터를 직접 입력하는 과정을 단순화시킬 수 있도록 하였다.

시킬 수 있다. 사용자가 촬영한 처방전은 GCD AI에 의해 분석·추출되어 설계된 DB에 자동 입력되므로 환자의 복약 관리 앱의 사용 편의성과 효율성이 향상된다.

제안 시스템은 일반 복약 관리 앱에서 제공하는 주요 기능들도 제공한다. 먼저 처방전 데이터를 기반으로 환자의 복용 일정을 관리할 수 있으며, 정해진 시간에 사용자가 설정한 알람 횟수에 따라 복약 알람 기능을 제공한다. 또한 의료 기관이나 사용자 복약 관리가 필요한 시스템에 환자의 복약 자료를 제공할 수 있다. 이러한 기능들은 환자의 질병 치료를 위한 복약 순응도를 높이고, 장기적인 건강 관리에 기여할 수 있을 것으로 판단된다.



(그림1) 시스템 구성도

3.2 시스템 설계 및 기능

제안한 시스템의 주요 기능은 카메라와 AI를 활용한 처방전 자동 입력 기능이다. 사용자는 앱 내에서 카메라를 통해 처방전을 비추기만 하면 AI가 처방전에 기록된 병명, 약물 이름, 처방 기록, 복용 시간 등의 주요 데이터를 자동으로 추출한다. 이를 통해 환자는 수동으로 데이터를 입력하는 과정을 생략할 수 있으며, 정보 입력 오류를 줄일 수 있다. 기존의 복약 관리 앱들은 사용자의 여러 가지 질병별 관리 기능을 제공하지 않고 있으나 제안 시스템은 사용자가 여러 질병들을 관리해야 할 경우, 질병 종류별 복약 관리할 수 있도록 기능을 제공한다.

제안 시스템의 상세 기능은 다음과 같다.

- 1) 처방전 입력
 - 사용자가 처방전을 촬영하면 AI 기술 (광학 문자 인식(OCR))을 통해 자동입력한다.
- 2) 복약 알람 설정
 - 복약 알람 설정 당시 횟수를 설정하고 설정된 시간 후에 알람을 다시 받을 수 있도록 설정한다.
- 3) 복약 알람
 - 사용자가 설정한 복약 시간이 되면 알람이 울리고, 복약 후 [확인]을 누르면 복약 정보가 DB에 자동 입력된다.
- 4) 복약 통계
 - 사용자의 복약 상태를 주별/월별/분기별/연도별로 관리하고 분석할 수 있도록 정보를 제공한다.
- 5) 질병 종류 관리
 - 사용자가 여러 가지 질병 관련 처방을 받고 있을 경우, 각 질병별 복약 및 통계 정보들을 관리할 수 있다.
- 6) 복약 정보 공유
 - 사용자의 복약 관리 정보를 담당의사나 보호자에게 제공한다.

제안 시스템의 구성도는 (그림1)과 같다.

3.3 GCD AI

Google Cloud Document AI(GCD AI)은 이미지 상의 텍스트를 자동으로 인식하여 디지털 데이터로 변환하는 기술로, 스캔된 문서, 사진, 또는 디지털 이미지에서 텍스트 데이터를 추출하는 데 널리 사용된다. 최근에는 AI 기술이 결합되어 정확성과 성능이 크게 향상되었으며, 다양한 언어와 문자에서도 신속하고 정확한 텍스트 추출이 가능해졌다.

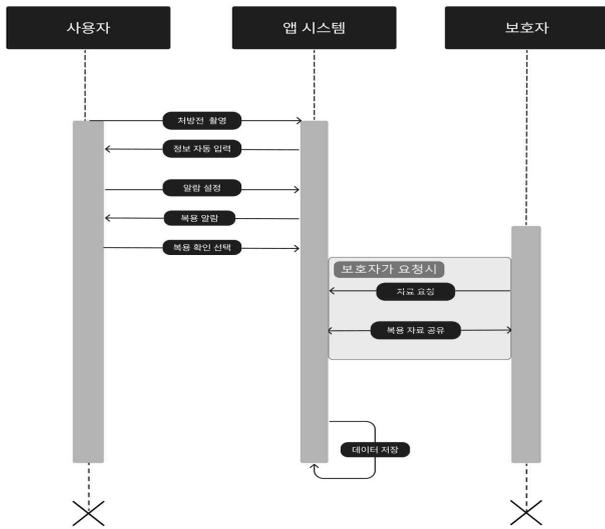
본 연구에서는 Google Cloud AI OCR 기능을 통해 환자의 처방전에서 텍스트 데이터를 자동으로 추출하고 이를 데이터베이스에 입력하는 시스템을 설계하였다. 사용자가 카메라로 처방전을 촬영하면, Document AI API가 이미지를 분석하여 텍스트를 인식하고 데이터를 분리한다. 시스템은 주요 정보를 필터링해 데이터베이스에 자동 저장한다. 이를 통해 환자들은 복잡한 처방 정보 입력 과정이 간소화되며, 노인 환자들도 손쉽게 사용할 수 있는 사용자 친화적인 환경을 제공한다.

3.4 처방전 입력 DB 설계

사용자 복약 정보와 관련된 처방 테이블을 공유받는다 제안 시스템의 복약 관리 DB는 'GCD AI'를 통해 처방전에서 추출된 문자들에서 제안 앱에서 필요한 복약 관련 사용자 정보들을 재추출하여 아래의 DB 테이블에 자동 입력되도록 설계하였다.

- 1) 질병 테이블
 - 사용자가 처방전을 촬영하면 GCD AI 기술을 통해 자동입력한다.
- 2) 처방의 테이블
 - 사용자의 복약 관련 정보를 공유받는다.
- 3) 알람 테이블
 - 복약 알람 설정 당시 횟수를 설정하고 설정된 시간 후에 알람을 다시 받을 수 있도록 설정한다.
- 4) 처방 테이블
 - 사용자의 복약 여부를 저장한다.
- 5) 처방전 테이블
 - 처방의가 발급한 처방전에서 추출해야 할 복약 관

련 필드들을 관리한다.



(그림2) 순차 다이어그램

IV. 결론

본 연구에서는 스마트 폰 사용이 어려운 고령 환자들의 복용 관리 어려움을 해결하기 위해 '처방전 자동 입력 복용 관리 앱'의 설계를 제안하였다. AI와 카메라 기술을 결합하여 처방전 데이터를 자동으로 추출하고 입력하는 과정을 단순화함으로써, 사용자들이 손쉽게 자신의 복용 정보를 관리할 수 있도록 설계하였다. 제안 연구에서는 앱 설계만 진행된 상태이며, 향후 이 설계를 바탕으로 앱의 개발을 진행할 예정이다.

※ 감사의 글 : “본 연구는 2023년도 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업 지원을 받아 수행되었음” (2023-0-00089)

참 고 문 헌

[1] https://kostat.go.kr/board.es?mid=a10502020100&bid=3207&act=view&list_no=161820&tag=&nPage=1&ref_bid=3203,3204,3205,3206,3207

[2] <http://www.mdon.co.kr/mobile/article.html?no=4794>

[3] <https://www.mkhealth.co.kr/news/articleView.html?idxno=8184>

[4] ‘처방전 관리 어플리케이션 : AI를 활용한 알약 인식 및 복용 관리 시스템 개발’, ASK 2024 학술발표대회 논문집, 31권 1호, 13-14p

[5] <https://connectcare.kr/>

암호화 기반 하이브리드 데이터 방송을 위한 프레임워크 설계

임석진
성결대학교
imseokjin@gmail.com

A Framework Design for Encryption-based Hybrid Data Broadcast

Im Seokjin*
Sungkyul Univ.

요 약

지능형 IoT 서비스는 자율주행 자동차와 드론 기반의 PAV와 같은 미래형 이동수단까지 확장되어 다양한 서비스를 제공할 수 있도록 확장되고 있으며, 대규모의 클라이언트를 지원할 수 있어야 하며 데이터의 보안성을 확보해야하는 특징이 있다. 본 논문은 암호화된 데이터와 평문 데이터를 같이 방송하여 보안성이 확보된 지능형 IoT 서비스를 가능하게 하기 위한 암호화 기반 하이브리드 데이터 방송을 위한 프레임워크의 설계를 제안한다. 제안된 프레임워크는 방송 서버와 클라이언트로 구성되고, 방송 서버는 방송컨트롤러, 암호화 모듈, 하이브리드 데이터 스케줄러와 인덱스 생성기 그리고 데이터 전송기로 구성된다. 제안된 프레임워크는 암호화 데이터와 평문 데이터의 효율적인 액세스와 검색 및 다운로드가 가능하게 하여 확장성과 데이터의 보안이 보장된 무선 데이터 방송 시스템을 구현할 수 있도록 한다.

I. 서 론

지능형 IoT 서비스는 모바일 기기를 이용한 서비스와 함께 자율주행 자동차와 드론 기반의 PAV와 같은 미래형 이동수단까지 확장되어 다양한 서비스를 제공할 수 있다 [1, 2]. 이러한 컴퓨팅 환경은 네트워크에 연결된 대규모의 클라이언트의 수와 함께 데이터의 보안성이 요구되고 있다. 무선 데이터 방송 기법은 대규모의 클라이언트에 데이터를 전달할 수 있는 확장성이 보장된 방법으로 지능형 IoT 서비스에 적절한 데이터 전달 기법이다 [3, 4].

무선 데이터 방송은 서버에서 무선 채널에 클라이언트가 필요로 하는 데이터를 데이터 방송시간에 대한 정보를 포함하는 인덱스 정보를 함께 방송하고 클라이언트는 무선 채널에 액세스하여 인덱스 정보를 통해서 필요로 하는 데이터의 방송시간을 확인하고 그 시간에 데이터를 다운로드하기 때문에 클라이언트가 급격히 증가하더라도 안정적인 정보 전달 서비스를 가능하게 한다 [5, 6]. 따라서 무선 데이터 방송에 암호화된 데이터를 방송하여 암호화된 데이터를 전달하여 확장성과 보안성이 확보된 지능형 IoT 서비스가 가능하게 된다.

본 논문에서는 암호화된 데이터와 평문 데이터를 같이 방송하여 보안성이 확보된 지능형 IoT 서비스를 가능하게 하기 위한 암호화 기반 하이브리드 데이터 방송을 위한 프레임워크를 설계한다. 제안되는 하이브리드 방송 기법은 암호화 데이터와 평문 데이터를 효과적으로 무선 채널에 배치하는 데이터 스케줄러와 클라이언트가 데이터의 방송시간을 확인할 수 있는 인덱스 생성기, 그리고 데이터와 인덱스를 무선 채널에 배치하는 채널 구성기로 구성되고,

클라이언트가 데이터를 검색하여 다운로드할 수 있도록 클라이언트의 구성요소를 설계한다.

II. 암호화 기반 하이브리드 데이터 방송 프레임워크 설계

2.1 하이브리드 데이터 방송서버 컴포넌트 설계

암호화 기반 하이브리드 데이터 방송을 위한 시스템의 컴포넌트는 다음과 같다.

- 방송 컨트롤러(Broadcast Controller) : 무선 데이터 방송 시스템의 다른 컴포넌트를 제어하여 암호화 기반의 하이브리드 데이터 방송이 가능하게 한다.
- 암호화 모듈(Encryption Module): 암호화가 필요한 보안 데이터에 대해 암호화를 진행한다.
- 하이브리드 데이터 스케줄러(Hybrid Data Scheduler): 암호화 데이터와 평문 데이터를 무선 채널에 방송하기 위해 인터리빙(interleaving) 방식으로 무선 채널에 배치한다.
- 하이브리드 인덱스 생성기(Hybrid Index Generator): 암호화 데이터와 평문 데이터의 방송시간을 보유하는 인덱스를 생성한다. 인덱스는 평문에 대한 인덱스와 암호화 데이터에 대한 인덱스를 생성하여 보유하며 인덱스 구조는 데이터 스케줄러의 스케줄링 알고리즘에 영향을 받는다.
- 데이터 전송기(Data Transmitter): 인덱스와 스케줄링 데이터를 무선 채널에 전송한다

2.2 암호화 기반 하이브리드 방송 절차

그림 1은 재난 경보 서비스를 위한 무선 방송 시스템의 방송 서버가 재난 경보 데이터를 방송하는 절차를 보여준다.

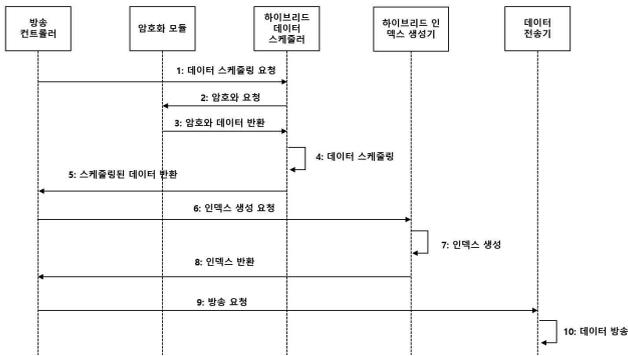


그림 1. 방송 절차를 위한 시퀀스 다이어그램

- 1: 방송 컨트롤러는 하이브리드 데이터 스케줄러에 데이터 스케줄링을 요청한다.
- 2, 3: 하이브리드 데이터 스케줄러는 암호화 모듈에 보안 데이터에 대한 암호화를 요청하고 암호화 모듈은 암호화된 데이터를 반환한다.
- 4, 5: 하이브리드 데이터 스케줄러는 암호화된 데이터와 평문 데이터를 스케줄링하여 방송 컨트롤러에 반환한다.
- 6: 방송 컨트롤러는 하이브리드 인덱스 생성기에 인덱스 생성을 요청한다.
- 7, 8: 하이브리드 인덱스 생성기는 암호화 데이터와 평문 데이터에 대한 인덱스를 생성하여 방송 컨트롤러에 반환한다.
- 9, 10: 방송 컨트롤러는 생성된 인덱스와 스케줄링된 데이터를 데이터 전송기에 전달하여 데이터 방송을 요청하고, 데이터 전송기는 데이터와 인덱스를 무선 채널에 방송하여 클라이언트들이 원하는 데이터를 다운로드할 수 있도록 한다.

2.3 하이브리드 무선 방송 시스템의 클라이언트 설계

클라이언트는 암호화 데이터와 평문 데이터가 방송되는 무선 채널에 액세스하여 각각이 필요로 하는 데이터를 검색하여 다운로드할 수 있다.

클라이언트를 구성하는 컴포넌트는 다음과 같다.

- 데이터 검색 매니저: 클라이언트에 포함된 각각의 컴포넌트를 관리하며 암호화 데이터와 평문 데이터 중에서 원하는 데이터를 다운로드할 수 있도록 한다.
- 데이터 선택기: 클라이언트가 검색하려는 데이터가 암호화 데이터인지 평문 데이터인지를 인덱스를 통해서 확인하여 원하는 데이터의 방송시간을 확인할 수 있도록 한다.
- 인덱스 검색기: 인덱스로부터 다운로드할 데이터의 방송시간을 확인하고 그 시간에 다운로드할 수 있도록 한다.

설계된 시뮬레이션 프레임워크를 이용하여 클라이언트가 원하는 지역에 대한 재난 경보 데이터를 수신하는 데까지 소요되는 시간인 액세스 시간과 액세스 하는 동안에 에너지 사용량을 대표하는 튜닝시간을 평가하여 무선 방송 시스템의 효율성을 평가할 수 있다.

III. 결론

본 논문은 지능형 IoT 서비스를 제공하기 위해 암호화가 필요한 보안 데이터와 평문 데이터를 하이브리드 방식으로 방송하는 무선 데이터 방송 시스템의 프레임워크를 제안한다. 제안된 프레임워크는 암호화 데이터와 평문 데이터의 효율적인 액세스와 검색 및 다운로드가 가능하게 하여 확장성과 데이터의 보안이 보장된 무선 데이터 방송 시스템을 구현할 수 있도록 한다. 설계된 프레임워크는 다양한 지능형 IoT 서비스를 구성하여 다양한 상황에서 효율적인 서비스를 제공할 수 있다.

참고 문헌

- [1] Fatima Alshehri; Ghulam Muhammad, "A Comprehensive Survey of the Internet of Things (IoT) and AI-Based Smart Healthcare", IEEE Access, vol. 9, pp 3360-3678, 2021
- [2] Maria Stoyanova, Yannis Nikoloudakis, Spyridon Panagiotakis, Evangelos Pallis, Evangelos K. Markakis, "A Survey on the Internet of Things (IoT) Forensics: Challenges, Approaches, and Open Issues", IEEE Communications Surveys & Tutorials, Vol. 22, No. 2, pp 1191-1221, 2020
- [3] A. B. Waluyo, F. Zhu, D. Taniar, and B. Srinivasan, "Design and Implementation of a Mobile Broadcast System", 2014 IEEE 28th International Conference on Advanced Information Networking and Applications, Vitoria, Canada, 2014
- [4]. S. Im, H. Youn, J. Choi, and J. Ouyang, "A novel air indexing scheme for window query in non-flat wireless spatial data broadcast", Journal of Comm. and Networks, Vol. 13, No. 4, pp. 400-407, 2011
- [5]. S. Im, and H. Hwang, "An Index based on Irregular Identifier Space Partition for Quick Multiple Data Access in Wireless Data Broadcasting," IEICE Transactions on Information and Systems, Vol. E99-D, No. 11, pp. 2809-2813, 2016
- [6] B.Zheng, W.C. Lee, D.L. Lee, "Spatial Queries in Wireless Broadcast Systems", Wireless Network, Vol. 10, No. 6, pp. 723-736, December, 2004.

레고 데이터 분석을 통한 트렌드 분석과 소비자 선호도 인사이트 도출

이가빈*, 문석재**
 광운대학교 스마트융합대학원
 {kabin0814, msj8086}@kw.ac.kr

Trend Analysis and Consumer Preference Insights Derived Through Lego Data Analysis

Ka-Bin Lee*, Seok-Jae Moon**
 Graduate School of Smart Convergence, Kwangwoon Univ.

요 약

본 연구는 1949년부터 2020년까지 레고 세트와 테마의 변화를 데이터 분석을 통해 조사하였다. 연구 결과, 레고 세트의 수는 꾸준히 증가하는 반면, 테마 수의 증가 속도는 상대적으로 느린 것으로 나타났다. 이는 제품의 다양성이 주로 세트 수의 증가에 의존하고 있음을 보여준다. 색상 다양성은 세트의 복잡도를 높이는 주요 요인이며, 이러한 복잡도는 2000년대 이후 급격히 상승하였다. 특히 Star Wars 테마가 가장 많은 세트를 보유하고 있어, 인기 영화 시리즈가 소비자 선호도에 긍정적인 영향을 미친다는 것을 시사한다. 이 연구 결과는 레고 제품의 발전과 소비자 선호도, 디자인 전략 수립에 기초 자료로 활용될 수 있으며, 향후 연구에서는 다양한 변수를 고려하여 소비자 선호도와 판매량 예측이 가능할 것으로 기대된다.

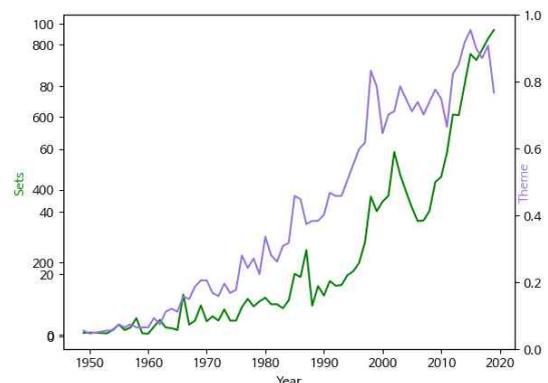
I. 서 론

레고와 같은 인기 있는 완구 브랜드는 매년 새로운 제품을 출시하여 소비자들의 관심을 끌고 있으며, 시대적 트렌드와 소비자 선호도의 변화를 반영하며 지속적으로 진화해왔다[1]. 본 연구는 레고 세트와 테마의 변화를 데이터 분석을 통해 탐구하고자 하며, 이를 위해 사용된 데이터는 세 가지 파일로 구성된다. themes.csv 파일은 레고 테마의 고유 식별 ID, 테마 이름, 상위 테마의 ID를 포함하여 레고 테마 구조와 계층적 관계를 파악하는 데 유용하다. sets.csv 파일에는 세트 번호, 이름, 출시 연도, 테마 ID, 부품 수가 포함되어 연도별 세트 수와 각 세트의 구성 복잡도를 분석하고, colors.csv 파일은 레고 부품에 사용된 색상의 고유 식별 ID, 이름, RGB 코드, 투명 여부를 제공하여 세트 구성에서 색상의 사용 경험을 분석하는 데 중요한 역할을 한다. 본 연구는 레고 데이터 분석을 통해 연도별 세트와 테마 수를 시각화하고 비교하여 시간에 따른 제품 다양성을 파악하고자 하며, 평균 부품 수를 통해 세트 복잡도의 증가 여부를 분석하여 레고 제품의 정교화 경향을 확인하고자 한다[2]. 특히 Star Wars 테마가 가장 많은 세트를 보유하고 Gear와 Friends 테마의 높은 세트 수는 소비자 선호도에 대한 중요한 인사이트를 제공한다. 이러한 결과는 레고 제품 발전에 대한 이해를 돕고, 머신러닝을 활용한 소비자 선호도 예측 및 디자인 전략 연구의 기초 자료로 활용될 수 있다[3]. 이 연구는 향후 다양한 변수를 고려하여 소비자 선호도와 판매량 예측을 가능하게 할 것으로 기대되며, 레고와 같은 완구 브랜드의 지속적인 발전과 변화에 대한 통찰력을 제공할 것이다.

II. 본문

2.1 연도별 세트 수와 테마 수

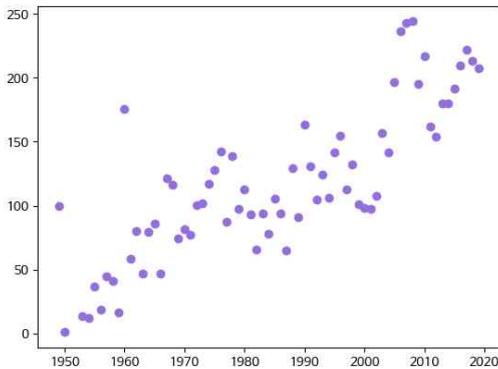
그림 1은 1949년부터 2020년까지의 세트 수와 테마 수의 변화를 비교하여 제품 다양성의 변화를 분석한 그래프이다. 세트 수는 지속적으로 증가하는 추세를 보이며, 새로운 제품 라인의 출현과 소비자 수요의 증가로 특히 2000년대 이후에는 눈에 띄게 상승했다. 반면, 테마 수는 세트 수에 비해 상대적으로 낮은 증가세를 보이고 있다. 테마 수의 점진적인 증가는 특정 주제나 콘셉트에 대한 관심이 증가하고 있음을 나타내는 반면, 그 증가폭은 세트 수에 비해 미미하다. 이러한 결과는 제품의 다양성이 세트 수의 증가에 크게 의존하고 있음을 시사하며, 이는 향후 제품 개발 및 마케팅 전략 수립에 중요한 시사점을 제공한다.



(그림 1) 연도별 세트 수와 테마 수

2.2 세트의 복잡도 분석

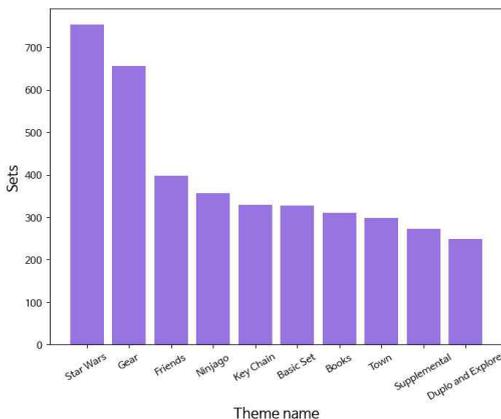
그림 2는 1949년부터 2020년까지의 세트 수에 따른 복잡도의 변화를 보여주는 그래프이다. 세트의 복잡도는 주로 구성 요소의 수와 그 조합의 다양성으로 정의되며, 시간이 지남에 따라 증가하는 경향을 보인다. 특히 2000년대 이후, 세트의 복잡도는 급격히 증가하였으며, 이는 세트의 복잡도를 더욱 증가시키는 요소로 작용한다. 레고 블록은 현재 135개의 고유한 색상을 보유하고 있으며, 다양한 색상 조합은 소비자에게 창의적인 조립 옵션을 제공하여 세트의 복잡도를 더욱 증가시키는 요소로 작용한다.



(그림 2) 세트의 복잡도

2.3 가장 많은 세트 수를 보유한 테마

그림 3은 레고 테마별 세트 수를 비교하여, 가장 많은 세트를 보유한 테마를 분석한 그래프이다. Star Wars 테마는 700개 이상의 세트를 보유하며, 가장 높은 수치로 나타났다. 이 테마는 인기 있는 영화 시리즈와 관련된 캐릭터와 우주선, 장면을 포함하고 있어 소비자들 사이에서 큰 인기를 끌고 있다. 그 다음으로 Gear 테마와 Friends 테마가 각각 500개와 400개 이상의 세트를 보유하고 있다. Gear 테마는 다양한 액세서리와 도구를 중심으로 구성되어 있으며, Friends 테마는 친구들 간의 일상적인 활동을 주제로 한 세트로, 지속적인 소비자 선호를 받고 있다. 이외에도 다양한 테마가 존재하나, 이들은 상대적으로 낮은 세트 수를 보유하고 있다. 이러한 결과는 특정 주제나 브랜드가 소비자에게 강한 작용을 하고 있음을 알 수 있다.



(그림 3) 레고 테마별 세트 수 현황

III. 결론

본 연구는 레고 세트와 테마의 변화를 데이터 분석을 통해 탐구하였다. 연구 결과, 연도별 세트 수는 지속적으로 증가하였으나, 테마 수의 증가는 상대적으로 더디다는 점이 확인되었다. 이는 제품의 다양성이 세트 수의 증가에 크게 의존하고 있음을 나타내며, 향후 제품 개발 및 마케팅 전략에 중요한 시사점을 제공한다. 색상의 다양성 또한 세트의 복잡도를 높이는 중요한 요소로 작용하고 있으며, 세트의 복잡도는 시간이 지남에 따라 증가하고, 특히 2000년대 이후 급격히 상승하는 경향을 보인다. Star Wars 테마가 가장 많은 세트를 보유하고 있으며, 이는 인기 영화 시리즈와의 연관성이 소비자 선호에 긍정적인 영향을 미치고 있음을 나타낸다. Gear와 Friends 테마 역시 높은 세트 수를 기록하였다. 이 연구 결과는 레고 제품 발전에 대한 이해를 돕고, 소비자 선호도 예측 및 디자인 전략 수립에 유용한 기초 자료로 활용될 수 있다. 향후 연구에서는 더 다양한 변수들을 고려하여 소비자 선호도와 판매량 예측을 가능하게 할 것으로 기대된다.

ACKNOWLEDGMENT

본 논문은 2024년도 광운대학교 교내 학술 연구비 지원에 의해 연구되었습니다.

참 고 문 헌

- [1] C. Hienerth, C. Lettl, and P. Keinz, "Synergies among Producer Firms, Lead Users, and User Communities: The Case of the <sc>LEGO</sc> Producer - User Ecosystem," *Journal of Product Innovation Management*, vol. 31, no. 4. Wiley, pp. 848 - 866, Dec. 06, 2013.
- [2] C. Bartneck and E. Moltchanova, "LEGO products have become more complex," *PLOS ONE*, vol. 13, no. 1. Public Library of Science (PLoS), p. e0190651, Jan. 02, 2018.
- [3] K. Lubiński, "The Effects of Nostalgia-based Advertising on a Toy Company's Brand Equity: The Example of the LEGO Group," *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie*, no. 6(990). Uniwersytet Ekonomiczny w Krakowie - Krakow University of Economics, pp. 29 - 48, Jul. 23, 2021.

빅데이터 기반의 공공인프라 공간정보 플랫폼 개발

서지훈*, 김유진*, 원서연*, 주길홍**

강남대학교, 경인교육대학교

{jihoon, cpyj126, 202303109}@kangnam.ac.kr, khjoo@ginue.ac.kr

Development of a Public Infrastructure Spatial Information Platform based on Big data

Seo Ji Hoon*, Kim Yu Jin*, Won Seo Yeon*, Joo Kil Hong**

Kangnam University*, Gyeongin National University of Education**

요 약

본 연구는 공간 데이터를 확보하여 빅데이터 환경을 구축하고 지도 기반의 API를 이용하여 공공인프라에 대한 평판 분석 플랫폼을 개발하였다. 주거지역의 긍정적인 시설과 부정적인 시설 정보를 제공할 뿐만 아니라 국민이 직접 참여하여 관공서 및 시설 인프라에 대한 객관적인 의견을 비정형 데이터로 제시할 수 있는 오피니언마이닝 기반의 지역 정보 공유 시스템으로 향후 참여형 플랫폼 분야에 기여할 것으로 예상된다.

I. 서 론

정보통신 기술의 급속한 발전으로 인하여 인터넷 보급률이 사회 전반에 걸쳐 인간의 영역으로 빠르게 확산되고 있으며, 유비쿼터스를 지향하는 모바일 서비스 기술과 클라우드 컴퓨팅 기술, 대용량 데이터를 관리하는 빅데이터 예측 기술, 사물인터넷(IoT) 기술이 점진적으로 진보하고 있다. 현재 국내는 IT 강국으로써 경제 성장률의 증가와 OECD 회원국으로써의 국가 브랜드의 가치와 위상을 높이고 있다. 또한 IT 기술력을 기반으로 다양한 융합기술 콘텐츠의 확보와 선진 인프라 시설의 구축을 통한 삶의 질적 향상을 높이고 있다. 이에 따라 국내의 교육과 S/W, 사회 기반 콘텐츠, 위치정보 서비스, 지역 환경 정보 시스템, 사회안전망 등의 기술이 보급됨에 있어서 인간중심 편익증진에 대한 기술이 주도하고 있는 실정이다. 이러한 기술에서 핵심 요소는 데이터의 활용이다. 전 세계적으로 공공데이터의 활용이 증가하는 추세에 국내 정부도 공공데이터를 민간업체와 연구기관에서 적극적으로 활용하도록 증진시키고 있다. 또한 웹을 기반으로 공간정보를 활용한 다양한 MAP API 기반의 문화 콘텐츠 시스템의 보급이 확대되고 있으며, 이와 같은 데이터를 활용하여 기술개발에 혁신을 불러일으키는 신생 스타트업 기업들이 증가하고 있다. 하지만 현대 사회는 창의적인 시스템이 경쟁력에 있어서 큰 성과를 거두고 이에 따라 신선한 아이디어 제품의 개발이 가속화 단계에 접어들었다. 현재 국내의 경우는 위치기반 서비스를 중심으로 부동산 서비스 및 매물 정보 서비스, 마케팅 등의 사회 환경에서 국민들이 필요한 시스템이 웹으로 출시되고 있는 실정이다. 하지만 국내의 지역적 특성상 작은 면적에 비해서 인구의 증가 현상에 따른 고령화는 높아지고, 땅값 시세의 폭등과 지역적 버블현상에 의한 주거 시설물의 높은 비용으로 인하여 서민들의 생활에 큰 어려움을 가지고 있으며, 주변 지역의 열악한 인프라 시설

환경임에도 부동산 가격은 상승하여 버블현상의 유지는 주기적으로 확산되고 있는 추세이다. 이러한 문제점을 개선하기 위해 본 연구는 정부에서 민간에게 오픈하고 있는 공공데이터를 수집하여 빅데이터로 저장 및 관리하고, 이를 활용하여 맵 API를 이용한 시각적인 오피니언 마이닝 분석 플랫폼을 개발하여 주거지역의 긍정적인 시설과 부정적인 시설 정보를 제공할뿐만 아니라 국민이 직접 참여하여 관공서 및 시설 인프라에 대한 객관적인 의견을 비정형 데이터로 제시할 수 있는 오피니언 지역 정보 공유 시스템을 개발하고자 한다.

II. 본문

2.1 플랫폼 개발의 시사점

본 연구 개발은 다음과 같은 목적을 기반으로 개발되었다. 첫 번째, 공공데이터를 활용한 지역 공간정보의 좌표 데이터를 통하여 각 지역 주민들의 생활 편의 시설 및 불편한 요소에 대한 정보를 제공하여 국민들의 주변 거주지역의 매매 및 시세를 판단하는데 중요한 척도를 제시할 수 있는 “국민의 알 권리” 보장제도 시스템이 필요하다. 두 번째, 현재의 상권 관련 부동산 정보 웹 서비스 등은 시스템 자체에서 제공하는 정보만을 국민들이 의지하여 판단하는 단점을 가지고 있다. 그러므로 단순히 시스템이 제공하는 정보에만 국한되지 않고 제약조건 없이 국민 스스로가 참여할 수 있어야 한다. 이에 따라 시설 정보에 대한 의견과 정보 공유를 통하여 지역의 인프라 시설에 따른 매매정보와의 연계를 기반으로 거주지역을 선택할 수 있는 콘텐츠가 필요하다. 세 번째, 공공분야의 빅데이터 오픈소스는 사회 시설망 형성과 예측 분석에 효과적인 정보를 제공하고 있지만 데이터의 활용이 정부의 기대에 미치지 못하는 흐름을 보이고 있다. 그러므로 본 연구에서 제시하는 기술은 개방된 과거의 공공데이터를 활용하여 위치기반 서비스

에 공간정보 시각화를 접목하여 포괄적인 지역 정보에 대한 긍정 인프라와 부정 인프라를 사용자에게 제공하는 동시에 국가가 개방한 폐기 데이터와 과거의 데이터를 재활용할 수 있는 순환적인 구조를 가지고 있으며 공공데이터를 활성화하고 활용하기 위한 중장기적 관점으로 보았을 때 효과적이다.

(표 1) 주변 지역 인프라 시스템의 문제 도출 기사 사례

기사 날짜	기사 제목
2009.06.17	아파트 창문만 열면 공동묘지가...주민들 대책 마련 요구
2019.10.29	창문열자 공동묘지, 숲세권 아닌 묘세권
2024.02.08	숲세권이러더니...창문열자 경악한 입주자들, 사기분양
2024.07.23	묘지뷰 아파트 갈등 6개월만에 일단락

부동산 시세의 증가로 인하여 내 집 마련을 위해 주변 상권에 대한 신중한 조사와 척도가 필요하지만 현재까지 주변 인프라 정보를 포괄적으로 제공해 주는 시스템이 부재이며, 높은 부동산 가격에 비해서 주변 시설물의 지리적 활용과 열악한 시설로 인하여 국민들이 주변 지역의 특성 및 매물 정보의 혼란이 야기되고 있다. 이에 따라 본 연구를 통하여 부동산 버블 현상의 최소화 및 쾌적한 환경의 인프라 정보 확산으로 경제적인 효율성을 해결할 수 있는 플랫폼을 구축하였다.

2.2 데이터 관리 환경 구축

본 논문은 공간정보 시각화 시스템의 개발을 위해 정부에서 지원하는 공공데이터 포털 및 각 부처의 데이터를 수집하였다. 수집한 정보의 긍정 인프라 요소와 부정 인프라 요소의 형태는 다음과 같다.

(표 2) 데이터 정보

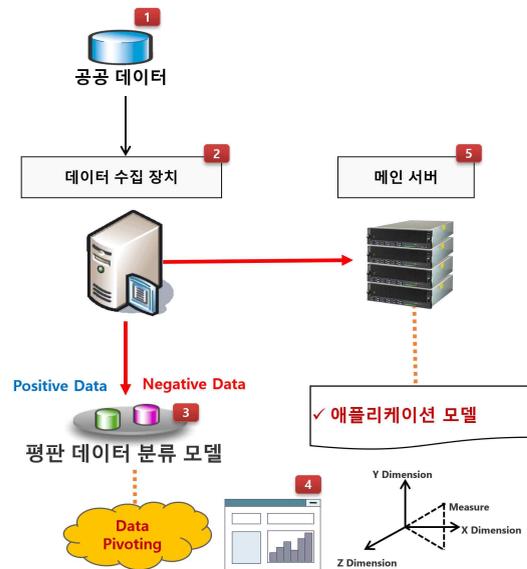
Positive Infra	Negative Infra
<ul style="list-style-type: none"> 초, 중, 고등학교(인문계열, 특성화 구분) 대학(국공립, 사립 구분) 및 교육기관 교육시설(어린이 집, 유치원, 학원가) 병원 및 응급실 파출소 및 경찰서, 어린이 보호지역 소방서, 우체국 및 그 밖의 관공서 은행, 레저시설, 숙박시설 편의시설 및 그 밖의 쇼핑물, 극장, 백화점 	<ul style="list-style-type: none"> 성범죄자 거주지 폐기물 시설 및 쓰레기 매립지역 사고 위험지역 범죄발생 지역 교도소 및 군부대 시설

오픈 데이터를 활용하여 사용자의 주거지역에 해당하는 긍정적인 인프라(Positive Infra) 시설과 부정적인 인프라(Negative Infra) 시설을 구분하기 위한 전략으로 데이터수집 서버를 별도로 구성하였으며, 수집 저장소와 각 긍정,

부정의 분야별 서버를 구축하여 분산 시스템을 병행하는 시스템으로 통제하였다. 이에 따라 데이터의 분류를 원활하게 할 수 있으며, 과부하의 장애를 최소화 시킬 수 있다. 오픈 맵(Open Map)에 위치 정보를 등록하기 위한 요구조건으로 각각의 데이터 좌표값을 n-차원 구조의 모델링으로 형성하였으며 피벗 테이블로 구성하였다. 공공데이터의 수집항목에는 일반적으로 관공서 및 편의시설을 위주로 수집을 수행하였고 긍정 인프라 업데이트를 위한 수집에는 관공서를 기준으로 경찰서, 소방서, 우체국, 은행 등의 데이터를 수집하였다. 그 밖에 교육기관 및 학교와 학원가를 중심으로 데이터를 확보하였으며 편의시설에는 극장 및 레저, 골프장, 쇼핑몰, 백화점 등의 데이터를 수집하여 긍정 인프라에 대한 종합적인 데이터를 저장소에 축적하였다. 부정 인프라는 성범죄자 거주지역에 대한 위치 정보와 폐기물 시설, 쓰레기 매립지역의 밀집 정보, 사고 및 빈번히 발생하는 범죄자 발생 위험지역을 기준으로 공공데이터를 수집하였다. 기본적으로 위치 좌표 값이 가공된 XML 문서를 수집하여 수집 서버 데이터베이스에 정보를 삽입하는 방식으로 진행되었다.

2.3 메인 서버 시스템 구축

본 연구에서 제시하는 메인 서버 시스템 구축 방식은 데이터 수집 장치 서버와 메인 서버의 두 가지 형태로 구성된다. 데이터 수집 장치 서버는 공공데이터를 축적하기 위한 목적으로 구동되며, 메인 서버의 기능은 공간정보 위치 기반 시각화 플랫폼을 구축하여, 공공데이터의 좌표 값을 오픈 맵에 정보를 삽입하는 용도로 사용된다. 또한 사용자로부터 인프라 정보에 대한 비정형 데이터가 입력되어 저장되면 이를 축적하고 관리하기 위한 목적이 있다. 수집 서버와 메인 서버는 서로 연동되며, 수집 서버에 축적된 공공데이터 긍정/부정 인프라 정보는 구축된 메인 서버에 이전된다.



(그림 1) 서버 시스템 구성도

서버 시스템의 시나리오는 다음과 같은 항목들로 구성된다. 첫 번째, 정부에서 지원하는 공공데이터 포털 및 관련

부처에 접근하여 공공데이터(public Data)를 수집하고 수집이 완료된 데이터는 수집 서버(collection server) 장치에 저장한다. 두 번째, 수집된 데이터는 수집 서버에 저장되며, 메타 데이터로 관리된다. 데이터의 종류로는 관공서에 대한 위치 좌표 정보 외에 교육시설 정보와 이 밖에 현황 정보들을 관리한다. 세 번째, 수집 서버에서는 데이터 셋(Data Set)이 완료된 정보를 기반으로 분류 작업을 수행하며, 분류에 대한 구분에는 긍정 데이터와 부정 데이터로 분류 및 정제 과정을 수행한다. 여기서 제시하는 정제 과정은 불필요한 데이터를 필터링하고, 콘텐츠 개발에 필요한 데이터만을 산출하는 작업을 수행한다. 네 번째, 정제 작업을 마친 데이터는 피벗테이블을 통하여 n-차원 모델링으로 생성한다. 관공서의 위치 좌표와 ID 값, 기관의 정보를 구분하여 데이터를 축소하는 작업을 수행한다. 마지막으로 수집 서버에서 가공된 데이터는 연동을 통하여 메인 서버에 이관을 수행하며, 메인 서버에는 웹 서버의 구축과 콘텐츠 개발을 위한 서버로써 데이터베이스 설계 및 테이블을 구성한다.

2.3 공간정보 플랫폼 개발

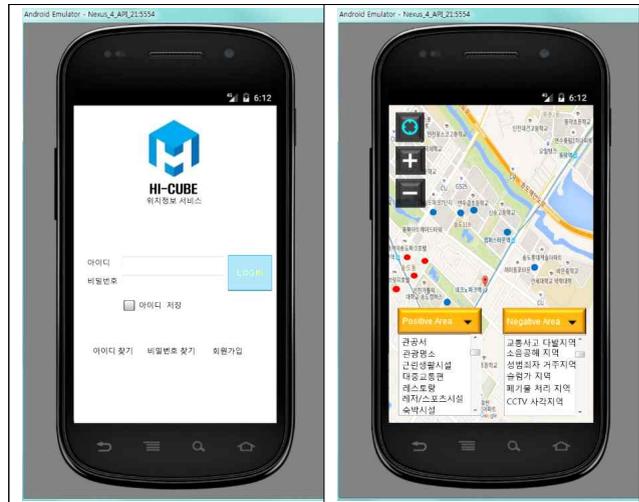
본 연구에서 제시하는 개발은 두 가지의 항목을 중요시한다. 첫 번째로 클라이언트의 웹 페이지의 접근성을 강화하고, 두 번째로 HTML5 환경을 통하여 모바일 웹에서도 안정적으로 접근할 수 있는 시스템을 구축한다. 공간정보 플랫폼은 오피니언마이닝 기반의 긍정/부정 인프라 검색 엔진 기능이다. 일종의 평판을 기준으로 사용자들에게 위치적으로 긍정적인 시설물과 부정적인 시설물의 위치 정보를 제공하는 기능을 의미하며, 본 개발의 핵심이 되는 요소이다.



(그림 2) 서버 시스템 구성도

현재 국내에 존재하는 부동산매매 관련 사이트는 주거지 매매 및 매입을 목적으로 사용자에게 원하는 방을 알선해 주고, 중개사무소와 연계하여, 클라이언트가 직접 하우스에 방문하지 않고도 거주지 내부의 정보를 확인할 수 있는 편리한 시스템이다. 하지만 이러한 시스템은 단순히 사용자에게 주거지 내부의 정보를 제공하는 시스템으로 되어 있으며, 주거지 주변 지역의 인프라 시설에 대한 정보를 자세하게 확인할 수는 없다. 일반 시민이 주거지를 선택하는 조건에는 주거지 내부 환경에 대한 요인도 중요하지만 최근 들어 주변 지역의 편의시설을 확인하고 입주하는 시민

들이 증가하고 있다. 이와 같은 예로 교육시설, 관공서, 위험지역 등의 정보가 주변 지역의 인프라 정보라고 할 수 있다. 본 기능은 주거지의 내부 환경이 아닌 주거지역의 외부 요인들의 인프라를 사용자가 파악하고, 주거지를 결정할 수 있는 시스템이다. 웹에 API를 연동하여 UI 및 레이아웃을 구축하고, 맵(Map)에 피벗이 완료된 공공데이터를 웹과 연동된 데이터베이스 서버를 통하여 좌표 값을 삽입하는 작업을 수행하였다. 여기서 제공될 내용은 긍정 인프라와 부정 인프라이며, 사용자 거주지를 중심으로 반경 거리 값을 설정하여 검색하는 기능을 부여하였다.



(그림 3) 시스템 인증 및 구성 화면 프로토타입

본 시스템은 반응형 웹으로 구축되어 모바일에서 웹 모드로 사용할 수 있지만 앱을 별도로 개발하여 접근성을 강화하였다. 위의 화면은 안드로이드 기반의 모바일 앱으로 개발한 화면이다.

오피니언 마이닝 기반의 긍정/부정 검색 기능 시스템의 시나리오 절차는 다음과 같다.

[Step. 1] 사용자는 웹에 접근하여, 맵을 활성화 한다. 모바일로 맵에 접근할 경우 클라이언트의 위치가 GPS 기능을 통하여 지도상에 현재위치로 출력되며, 데스크톱 클라이언트로 접근시에는 검색하고자하는 위치를 수동으로 지정할 수 있다.

[Step. 2] 클라이언트는 원하는 지역을 검색하고, 지정된 위치로부터 반경 값을 설정하여, 긍정/부정 인프라 정보를 확인할 수 있다. 반경 값은 클라이언트가 설정한 주거지역을 중심으로 반경 100M, 200M, 300M, 500M, 1KM의 선택 값을 통하여 긍정/부정에 대한 인프라 정보가 자동으로 생성된다. 이와 같은 서비스는 위치 정보에 대한 인프라 시설의 정확성을 높이고, 검색 시간을 최소화함으로써 사용자의 실용성을 향상시킨다.

[Step. 3] 클라이언트는 긍정 및 부정에 대한 정보를 키워드로 입력하여 원하는 정보만을 검색할 수 있는 기능이 부여된다. 즉, 중심 지역에서 반경 100M에 초등학교에 대한 정보만을 확인하고 싶은 경우에 검색창에 키워드를 입력하여 원하는 정보만을 출력할 수 있다. 긍정과 부정은 대부분류와 중분류로 구성되며 예를 들어 교육시설이 분류될 때

대분류의 경우 교육시설의 종류(초등학교, 중학교, 고등학교, 대학교, 학원 등)와 개수, 중분류는 교육시설의 명칭으로 구분된다.

[Step. 4] 클라이언트는 검색한 반경 지역에 긍정 인프라와 부정 인프라에 대한 각각의 평판데이터 통계 정보를 검색할 수 있으며, 퍼센트(%) 레벨도(level)에 따라서 청정지역, 안전지역, 위험지역을 확인할 수 있다. 오피니언마이닝 통계는 식별하기 편리한 맵의 시각화처리를 통하여 해당 사용자에게 서비스가 제공된다.

III. 결론

본 연구는 인프라 데이터를 수집하여 긍정 데이터와 부정 데이터를 분류하여 공간정보를 활용한 평판 분석 플랫폼을 구축하였다. 현재 다양한 종류의 공공데이터가 오픈되었으나 현재 이와 같은 데이터를 활용할 수 있는 여건과 기능이 미비한 상태이므로 본 논문을 통해서 폐기된 데이터도 쉽게 활용하여 민간분야에 빅데이터 분석 활용 시스템으로 사용할 수 있는 기술력을 확보하였으며 본 플랫폼을 사용하는 클라이언트들의 참여를 통하여 새로운 비정형 데이터를 생산할 수 있으며, 이와 같은 비정형 데이터는 평판 분석을 통하여 의미 있는 데이터를 추출하는데 유용한 서비스를 제공할 수 있으므로 서비스의 확대 및 시장수요가 폭발적인 규모로 증가할 수 있는 잠재력을 기대한다.

참 고 문 헌

- [1] 차영일, 최성규, 한경석, “민간의 공공데이터 활용을 위한 이용의도에 미치는 영향에 관한 실증적 연구”, 디지털융복합연구, 15(6), 9-17, 2017.
- [2] 박지현, 김성진, 백순재, “공공데이터 개방 및 활용의 문제점과 개선 방안”, 한국컴퓨터정보학회 학술발표논문집, 32(2), 631-632, 2024.
- [3] 이상호, “기초생활인프라가 공동주택가격에 미치는 영향 분석”, 한밭대학교 산업대학원, 2022.

근감소증 표준 데이터 관리 플랫폼 개발

임동욱¹, 노시형¹, 김지희², 정창원^{1,3}

¹원광대학교 근감소증토달솔루션 센터

원광대학교 의과대학 및 재활의학과

²원광대학교병원 정보관리실 스마트 사업팀

{nosij123, dwl316}@wku.ac.kr, gold82mouse@hanmail.net, mediblue@wku.ac.kr

Development of a standard data management platform for sarcopenia

SiHyeong Noh¹, Dong-Wook Lim¹, Ji Hee Kim², Chang-Won Jeong^{1,3}

¹Sarcopenia Total Solution Center, Wonkwang University

²Department of Rehabilitation Medicine, Wonkwang University School of Medicine and Hospital

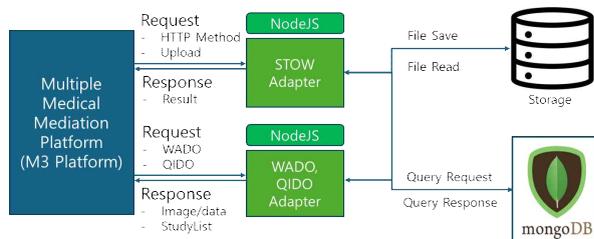
³Smart Business Team of Wonkwang University Hospital

요 약

본 연구는 근감소증(Sarcopenia) 진단을 위한 데이터 관리 플랫폼으로 임상현장에서 근감소증 진단 프로세스의 정립과 함께 신체기능평가를 통해 근감소증을 진단하고 이와 관련된 임상정보를 수집하기 위해서 웹기반 플랫폼을 개발하였다. 근감소증은 근육량과 신체기능의 저하를 특징으로 하며, 조기 발견이 중요하다. 이를 위해 플랫폼을 통해 체계적으로 데이터를 수집하여 관리하고 분류된 근감소증 환자에 대한 다양한 연구를 위한 활용 데이터로 제공할 수 있다.

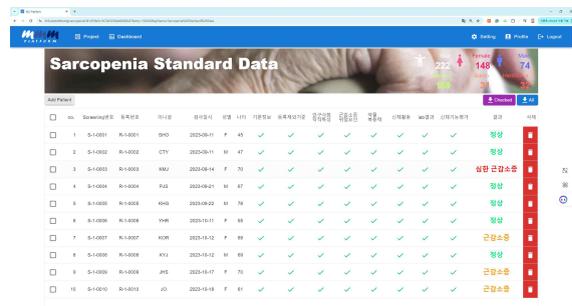
병원 재활의학과에서 수집하였다.

다음 그림 1은 우리가 개발한 웹기반 데이터관리 플랫폼의 소프트웨어 구조를 보이고 있다. 영상뿐만 아니라 데이터까지 포함하여 Node 서버에 의해 파일로 저장·관리하며 데이터베이스는 mongoDB를 사용하였다.



(그림 1) 웹기반 데이터 수집 관리를 S/W 구조

그림 2는 웹기반 근감소증 표준 데이터 수집한 결과 화면을 보이고 있다.



(그림 2) 웹기반 데이터 수집 관리를 위한 결과 화면

I. 서론

근감소증(Sarcopenia)은 고령화, 만성 질환, 신체 활동 감소 등 다양한 요인으로 인해 발생하는 근육량 감소와 관련된 질병으로, 신체 기능 저하와 삶의 질 하락을 초래하여 심각한 의료 문제로 대두되고 있다[1]. 근감소증 관리에 있어 진단과 치료의 일관성을 유지하려면 표준화된 데이터 관리와 통합된 플랫폼이 필수적이다[2]. 하지만 현재는 근감소증 진단과 관련된 기준과 데이터 관리 방식이 연구마다 상이하여, 임상 연구와 실무에서의 데이터 통합과 비교가 어렵다는 문제점이 존재한다.

본 논문에서는 근감소증 진단 기준과 임상 현장에서 환자로부터 얻어지는 데이터에 대한 정의를 기반으로 근감소증 데이터 표준화하고 이를 웹 기반 플랫폼으로 개발하였다. 이후, 제안하는 플랫폼 설계 및 구현 방안을 통해 표준화된 데이터 관리 체계를 구축하고, 향후 근감소증 관리와 임상 현장에 적용 가능성을 제안하고자 한다.

II. 웹기반 근감소증 표준 관리 플랫폼

2.1 데이터 수집 방법

우리가 수집하고자하는 데이터는 인구 사회학적 특성과 근감소증위험요인 그리고 약물 복용력, 신체 활동 그리고 Lab 결과와 신체기능 평가근감소증 환자는 아시아 분류 척도(AWGS)를 사용하여 근육량, 악력 그리고 신체기능 평가를 통해 근감소증 진단 결과를 정상, 근감소증 그리고 심한 근감소증으로 분류하여 수집하였다. 관련 데이터는

현재 수집된 총인원은 222명이고 이 중에 여자는 148명 그리고 남자는 74명이다. 그리고 정상은 159명이고 근감소증 환자는 31명 그리고 심한 근감소증 환자는 32명으로 63명이 근감소증 환자로 확인되었다.

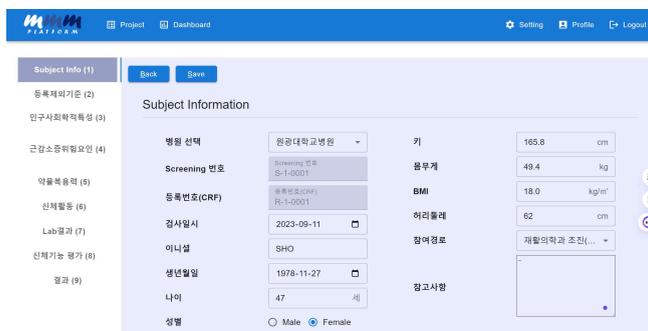
2.2 근감소증 진단 방법

임상현장에서 맞게 근감소증을 진단하는 방법으로는 근육량, 악력과 신체 기능을 평가한다. 이와 관련해서 ① 악력, ② 근육량 지수 및 ③ SPPB 점수가 모두 특정기준에 못미칠 경우 “심한 근감소증 상태”라고 진단하며, 이들 중에 악력 또는 SPPB점수가 미달인 경우는 근감소증 전단계 그리고 악력과 근육량지수 또는 SPPB 점수가 미달인 경우는 근감소증으로 진단한다. 이와 관련된 기준은 다음과 같다.

1. 악력: 양손 각 3회 측정된 값의 평균 값 (M<28kg, F<18kg)
2. SMI: 체성분 검사하여 나온 근육의 양 (M<7.0kg/m², F<5.7kg/m²)
3. SPPB(≤9점)
 - ▶ 정적균형(4점)
 - 일반자세 10초 유지 여부(1점)
 - 반일렬자세 10초 유지 여부(1점)
 - 일렬자세 10초 유지 여부 (2점)
 - ▶ 보행속도(4점)
 - 4m 보행 소요시간 및 속도(m/s)
 - ▶ 일어서기(4점)
 - 의자에서 5회 일어날 때 소요 시간

2.3 근감소증 연구대상자의 표준 데이터 수집 결과

우리는 연구대상자의 표준데이터 수집을 위해서 앞서 언급한 진단을 위한 측정 정보 외에 인구사회학적 특성, 근감소증 위험요인 그리고 약물 복용력과 신체 활동과 Lab 결과에 대한 데이터를 함께 수집하도록 구성하였다.



(그림 3) 근감소증 표준 데이터 수집 화면 중 연구대상자 정보

연구대상자의 근감소증 진단 결과는 다음 그림 4와 같이 임상현장에서 측정된 결과에 따라 확인할 수 있도록 하였

다. 결과화면에 출력된 연구대상자는 진단 조건에 해당하는 악력과 근육량 지수 그리고 신체기능 평가가 평가기준에 못미치는 결과로 “심한 근감소증” 판정을 받았음을 확인할 수 있다.



(그림 4) 근감소증 진단 결과 화면

III. 결론

우리는 근감소증 질환의 연구대상자에 대한 다양한 데이터를 수집하고자 근감소증 표준 데이터 관리 플랫폼을 제안하였다. 특히 임상 현장에서 측정하여 결과를 확인하고 이를 근거로 환자에게 근감소증에 대한 인식과 함께 운동 처방 및 식이요법에 대한 가이드라인을 제시하는데 활용한다. 추후 연구로는 운동처방과 식이요법에 대한 정보 및 지역로컬푸드 산업과 연계한 서비스를 개발하기 위한 연구를 수행할 계획이다.

ACKNOWLEDGMENT

본 연구는 과학정보통신부의 재원으로 한국연구재단(NRF-2021R1A5A8029876) 지원에 의하여 이루어진 것임.

참 고 문 헌

- [1] Han A, Bokshan SL, Marcaccio SE, DePasse JM, Daniels AH. Diagnostic Criteria and Clinical Outcomes in Sarcopenia Research: A Literature Review. Journal of Clinical Medicine. 2018; 7(4):70. <https://doi.org/10.3390/jcm7040070>
- [2] Tagliafico, A.S., Bignotti, B., Torri, L. et al. Sarcopenia: how to measure, when and why. Radiol med 127, 228-237 (2022). <https://doi.org/10.1007/s11547-022-01450-3>
- [3] 원장원, “Diagnosis of sarcopenia in primary health care” J Korea Med Assoc 2020.10, 63(10):633

사서의 인공지능 교육을 위한 학습 과정 정보 추천 시스템 설계

서지훈*, 배성민*, 정환웅*, 황태준**
강남대학교, 인천대학교

jihoon@kangnam.ac.kr, dkrak6779@gmail.com,
tree193nn@kangnam.ac.kr, tjhwang29@inu.ac.kr

Design of a learning process information recommendation system for artificial intelligence education of librarians

Seo Ji Hoon*, Bae Seong Min*, Jeong Hwan Woong*, Hwang Tae Jun**
Kangnam University*, Incheon National University**

요 약

본 논문은 도서관 사서를 대상으로 직무 기반의 인공지능교육을 수행할 때 원하는 교과과정을 추천받고 원하는 학습을 효과적으로 적용할 수 있는 추천 모델을 설계하였다. 본 모델은 학습 대상자인 사서들에 대한 개인 정보를 데이터를 기반으로 문제점을 진단함으로써 사서들의 직무와 역량에 따라 의사결정에 의해서 필요한 교육 프로그램을 추천받을 수 있는 시스템이다. 본 모델을 통하여 사서들이 디지털 교육을 효과적으로 학습할 수 있도록 제공하여 교육적 질을 높이고자 한다.

I. 서 론

4차산업혁명에서 중요한 주제로 인식되고 있는 인공지능 분야는 점차 크게 발전하고 있다. 국내에서도 2019년 12월 AI 국가 전략 발표를 시작으로 2020년에 AI교육에 관한 정책을 발표하였다. 이와 같은 정책의 의미는 교육 현장에 인공지능 교육이 의무적으로 도입되며 관련된 교재 개발, 교원 양성 등으로 인프라 구축의 확대가 이루어질 것으로 예상된다. 또한 인간의 일상과 직무에서도 인공지능의 비중이 높아질 수 있다는 가능성을 내포하고 있다. 이와 같은 이슈들은 다양한 인간의 삶과 직종에 큰 영향을 주고 있으며 하나의 특이점이 되어 인간과 인공지능이 공존하기 위한 방법을 모색할 시기이다. 이러한 환경 속에서 도서관 사서의 직종은 시대 흐름에 맞게 적응하기 위한 여러 가지 방안들을 제시하고 있다. 미래 일자리의 위협을 받을 수 있는 직종에 하나로써 문헌 정보학 분야에서는 도서관의 전문 사서의 양성을 디지털 분야로 융합시키고 있다. 이에 따라 직무의 활용이 높은 데이터를 중심으로 “데이터 전문 사서”의 역량으로 인재를 양성하기 위한 노력을 진행하고 있다. 본 논문은 전문 데이터 사서를 양성하기 위한 교육 과정으로 사서들에게 직무 중심의 효과적인 교과과정을 추천해 줄 수 있는 모델을 설계하였다. 본 연구를 통하여 사서들이 개인의 역량과 직무에 적합한 교육을 통하여 학습의 효과성을 높일 수 있기를 기대한다.

II. 본문

2.1 연구의 시사점

본 연구 개발은 다음과 같은 목적을 기반으로 연구되었다. 현재 국내의 인공지능교육 시스템의 경우 점진적인 도

입기를 지나고 있으며 해외의 선진화된 다양한 교육 모델 사례를 기반으로 학습자에게 효율적인 교육과정을 제공함과 동시에 다양한 변화를 시도하고 있다. 과거의 2차 산업 혁명은 밀집된 환경에서 산업 환경의 생산성을 높이기 위한 인재 양성 교육에 집중하였다. 하지만 지금의 사회는 정보화 시대를 지나서 과학적 진보와 발견을 통하여 다양한 융합기술이 파생되어 새로운 학문이 등장함에 따라 21세기 교육은 미래 교육에 적합한 변화를 제시하고 있다. 현재 종사하고 있는 전문사서의 경우 아날로그 시대에서 스마트 시대를 지나온 세대이다. 그러므로 새롭게 파생되어 등장한 인공지능 분야는 추후 도서관 시스템에 반드시 적용될 기술이므로 이에 대응할 수 있는 사서를 위한 직무 유형별 인공지능 추천 학습모델이 필요하다.

2.2 학습 과정 정보 추천 시스템 설계

본 논문에서 제시하는 사서들을 위한 학습 과정 정보 추천 시스템의 설계 절차도 빅데이터 분야의 절차를 응용하였다. 인공지능을 정책이나 교육 인프라에 적용한다는 것은 반드시 데이터와 실제로 문제 해결에 활용할 수 있는 서비스를 포함한다는 것을 의미한다. 이에 따라 본 논문에서는 사서에 대한 인공지능 교육 프로세스도 AI 프로젝트 사이클을 기반으로 1) 사서 교과과정에 대한 설정 단계, 2) 사서 개인 정보에 대한 데이터 수집 단계, 3) 사서 데이터 기반의 인공지능 교과과정 추천 단계, 4) 인공지능 교육과정 학습 단계, 5) 사서 기반의 직무 적용 평가 단계, 6) 활용 단계로 사이클 모형을 제시한다. 이와 같은 절차는 중장기적 관점으로 보았을 때 효과적이다. 또한 현재 이슈가 되고 있는 데이터 전문 사서의 영역에서는 데이터 리터러

시와 기초통계, 데이터 마이닝 기술 등의 교과 과정이 포함되었지만 본 논문에서는 직무에 필요한 순수 인공지능 교육만을 기반으로 교과 영역에 포함하고 그 밖에 교과 항목들을 점진적으로 확대하는 방향으로 연구를 진행하였다. 아래의 그림은 전체 개발 내용에 대한 설계 절차이다.



(그림 1) 설계 절차

본 시스템은 기계학습 알고리즘을 기반으로 축적된 데이터의 패턴을 추출하고 의사결정 시나리오에 따라 사서에게 개인 직무에 맞는 적합한 맞춤형 인공지능교육 교과과정을 제공함으로써 사서의 직무 역량에 도움을 줄 수 있는 효과를 발휘한다.

2.2.1 사서 정보수집을 위한 데이터 수집기

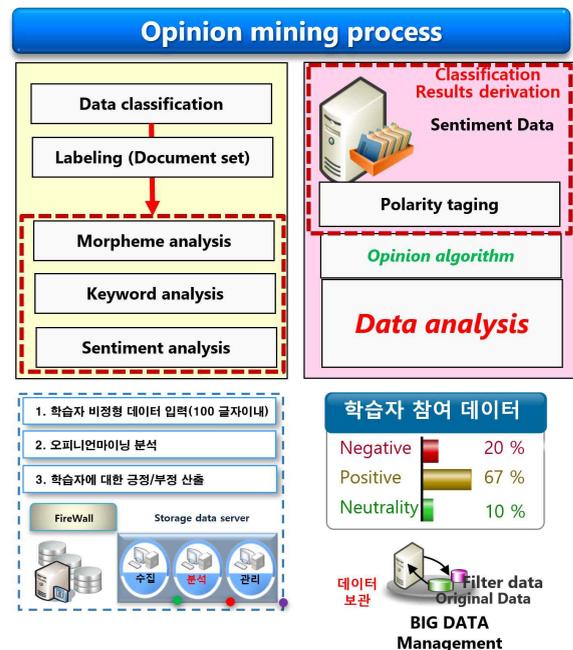
사서는 학습자의 유형으로 기존에 개인 정보를 웹 플랫폼을 이용하여 텍스트로 입력한다. 여기서 학습자의 정보의 필수 항목은 직무 유형과 개인의 역량 정보이며 해당하는 내용을 모두 기재하면 데이터는 서버에 저장이 된다. 여기서 단어의 태깅 분석을 통하여 사서의 역량에 맞는 난이도를 판별하고 직무에서 활용할 수 있는 맞춤형 인공지능 교육 교과과정을 추천한다. 학습자의 정보는 지능형 학습 플랫폼 저장소에 축적되어 대용량 비정형 데이터(빅데이터)로 관리되며 진단 텍스트(비정형 데이터) 정보를 실시간으로 분석하여 시스템에 자동 수집 및 가공처리 되는 수집 방법을 채택하였다. 학습자인 사서들의 분석 결과에 따라 맞춤형 교과과정을 제공함과 동시에 개인의 학습 현황과 설문 데이터, 직무와 역량 데이터, 난이도에 따른 문제해결 능력지수 등을 저장하고 관리한다.

2.2.2 학습모델 설계

기존에 제시된 소프트웨어 교육의 학습 절차는 인공지능 교육에 적용하기에는 효과적이지 않다는 학자들의 논의가 제시됨에 따라 인공지능교육에 적합한 시현중심모델, 구성중심 모델, 개발중심 모델, 디자인중심 모델, CT요소 중심 모델을 확장하여 한국형 인공지능교육 기반의 강인한 학습 모형을 설계하고 이를 플랫폼에 적용하였다. 또한 본 논문에서 중요하게 고려할 사항은 사서의 직무에 맞는 교육 프로그램을 제공하는 것이다. 이에 따라 현재 데이터 전문 사서 양성을 위해 다루고 있는 과정에서 인공지능 교육에 적합한 항목들을 분류하여 추천 과정 엔트리에 포함하였으며 전반적으로 데이터 분석에 활용되는 학습 내용으로 구성하였다.

2.2.3 추천 시스템의 원리

본 논문에서 제시하는 추천 시스템은 하나의 교과목을 추천하지는 않는다. 다만 사서 직무에 필요한 인공지능 교육 과정을 추천하기 때문에 예를 들어 학습자의 직무가 데이터 분석 분야이면 머신러닝 분야에서 사서직에 필요한 데이터 분석 기법 등의 챗터별 교육내용을 추천하여 사서에게 제공한다. 즉, 머신러닝이나 딥러닝과 같은 큰 틀의 학문을 추천하는 것이 아니고 머신러닝에서 필요한 데이터 분석 기법의 내용을 추천하여 제공한다. 추천 방식의 조건은 학습자인 사서가 제공한 비정형 데이터의 평판 분석에서 도출된 평가 항목과 설문지 문항의 척도를 적용하며 설문지의 경우 서버에 저장된 정합도 간의 유클리드 거리를 연산하여 교과과정을 추천한다.



(그림 2) 추천 시스템 학습에 필요한 평판데이터 생산과정

본 시스템에는 자가 진단, 설문 시스템, 학습 진도 및 현황 시스템, 학습에 따른 개인 효과성 분석 시스템 등을 포함한다. 또한 본 시스템에는 한국어 문법 기반의 159,000여 종의 감성 단어를 분석할 수 있는 단어사전이 구축되며,

이를 통해 학습자의 생각, 견해, 애로사항, 문제점, 심리 상태 등을 파악할 수 있는 장점이 있다.

III. 결론

본 논문에서는 인공지능의 영역이 빠르게 발전함에 따라 도서관 사서들의 직무도 디지털 기반의 직무로 변화할 가능성을 고려하여 사서를 대상으로 직무 기반의 인공지능교육을 수행할 때 원하는 교과과정을 추천받을 모델을 설계하였다. 본 논문에서 제시한 모델은 학습 대상자인 사서들에 대한 개인 정보를 데이터를 기반으로 문제점을 진단함으로써 사서들의 직무와 역량에 따라 의사결정에 의해서 필요한 교육 프로그램을 추천받을 수 있는 시스템이다. 본 논문에서 제시하는 시스템은 추후 자율전공 대학생들의 전공 선택이나 더 나아가서 다른 분야의 추천 시스템으로 활용이 가능할 것으로 예상된다.

참 고 문 헌

- [1] JI HOON SEO. (2022). BIG DATA-BASED SW EDUCATION CURRICULUM RECOMMENDATION PLATFORM DESIGN FOR LEARNERS. Journal of Theoretical and Applied Information Technology, 16(100), 4971-4978.
- [2] 서지훈, 이선아, 김백현, 황태준, 추교남. (2021). 학습자의 데이터를 활용한 커리큘럼 추천 시스템 설계. 한국정보기술학회 학술발표논문집, 109-112.
- [3] Seo J, Jo H, Choi J, "Design for opinion dictionary of emotion applying rules for antonym of the korean grammar", JKIIIT 13(2), pp.109 - 117, 2015.

딥러닝을 활용한 메타유전체 분류 도구의 정확도 향상

장세인, *문석재
광운대학교 스마트융합대학원, *광운대학교 스마트융합대학원
seine1025@kw.ac.kr, *msj8086@kw.ac.kr

Improving the Accuracy of Metagenomic Classification Tools Using Deep Learning

Jang Seine, *Moon Seok jae
Graduate School of Smart Convergence Kwangwoon Univ.,
*Graduate School of Smart Convergence Kwangwoon Univ.

요 약

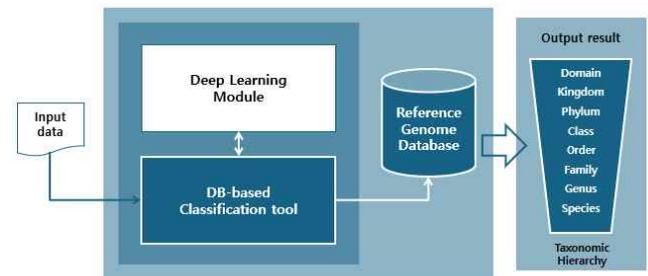
본 연구에서는 메타유전체 분류의 정확성과 속도를 개선하기 위한 딥러닝 기반 모듈 DeepOpt를 제안한다. 메타유전체는 특정 미생물 군집의 모든 유전체를 포함하는 이론적 집합으로, 고처리량 시퀀싱 기술의 발전에 따라 방대한 데이터의 효율적 분석이 필수적이다. 기존의 데이터베이스 정렬 기반 방법은 참조 유전체 데이터베이스의 품질에 따라 한계가 있으며, 이는 정확성에 영향을 미친다. DeepOpt 모듈은 입력 데이터를 사전 학습하여 특징을 추출하고, 이를 기반으로 데이터베이스 기반 분류 도구를 활용하여 최종 분류를 수행. 이 접근 방식은 특히 알려지지 않은 종의 분류에서 강점을 보이며, 대규모 시퀀싱 리드에 대해 신속하고 정확한 분석을 가능하게 한다.

I. 서 론

메타유전체는 특정 미생물 군집의 모든 구성원으로부터의 유전체의 이론적 집합을 의미한다 [1]. 메타유전체학을 통해 순수 배양으로 쉽게 얻을 수 없는 대부분의 미생물의 유전체를 밝혀낼 수 있다 [2]. 고처리량 시퀀싱 기술이 계속해서 발전하면서, 생성된 방대한 양의 데이터를 분류하기 위한 분류학적 분류는 생물정보학 분야에서 중요하게 자리잡았다 [3]. DNA의 정확한 분류학적 분류를 위한 대규모 시퀀싱 데이터를 효율적으로 분석하는 것은 여전히 어려운 과제로 남아있다 [4]. 기존의 분류학적 분류 기법에서 지배적인 패러다임은 데이터베이스 기반 및 기계 학습이다. 데이터베이스 접근법은 주로 알려진 참조 유전체 데이터베이스에 시퀀싱 리드를 정렬하여 분류학적 분류를 결정하는 데 중점을 둔다 [5]. 데이터베이스 기반 방법은 참조 유전체 데이터베이스가 광범위하고 완전할 경우 높은 정확도를 달성할 수 있지만, 이는 참조 유전체 데이터베이스의 품질에 의해 제한된다 [6]. 반면, 기계 학습 방법은 데이터베이스가 불완전하거나 훈련 세트에 포함되지 않은 종을 분류해야 하는 특정 상황에서 강점을 보이며, 훈련 데이터에서 알려지지 않은 종의 존재를 추론할 수 있다. 그러나 이러한 모델의 성능은 종종 훈련 데이터의 대표성과 양에 의해 제한된다 [7, 8]. 본 논문에서는 데이터베이스 정렬 기반 방법의 분류 도구 정확도를 향상시키기 위한 딥러닝 기반 모듈 DeepOpt를 제안한다. 이 모듈은 참조 유전체 데이터베이스에서 분류를 수행하기 전에 입력 데이터를 학습하여 1차적으로 분석을 실행한 후, 데이터베이스

기반 분류 도구를 활용해 최종 분류를 진행한다. 이를 통해 대규모 시퀀싱 리드에 대해 서열을 정렬하여 종을 식별하는 기존의 정렬 기반 방법의 분류 속도와 정확도를 개선하는 것을 목표로 한다.

II. 본문



(그림 1) 딥러닝 기반 모듈이 추가된 메타유전체 분류 시스템

그림 1은 딥러닝 기반 모듈이 추가된 메타유전체 분류 시스템의 구조를 보여준다. 이 시스템은 입력 데이터, 딥러닝 모듈, 데이터베이스 기반 분류 도구, 참조 유전체 데이터베이스, 그리고 출력 결과로 구성된다. 입력 데이터는 사용자나 외부 시스템에서 제공되는 원시 시퀀싱 데이터로, 시스템의 초기 입력으로 사용된다. 딥러닝 모듈은 이 입력 데이터를 처리하여 특징을 자동으로 특징 추출하고 학습을 수행한 후 결과를 데이터베이스 정렬 기반 도구로 전송한다. 데이터베이스 정렬 기반 분류 도구는 딥러닝 모듈에서

처리된 데이터를 바탕으로 참조 유전체 데이터베이스에서 쿼리를 실행하여 최종 분류 결과를 도출한다. 이 결과는 계통 분류의 계층 구조를 포함하여 도메인, 계급, 문 등의 분류 수준을 나타낸다.

표 1과 같이 본 연구에서는 TensorFlow를 사용하여 다양한 딥러닝 모델을 훈련하고 평가하기 위한 프레임워크를 제안한다. 이 프레임워크는 모델 구성, 훈련, 평가 및 예측 과정을 간소화하여 여러 아키텍처를 수용할 수 있는 모듈화된 접근 방식을 제공한다. initialize_model 함수는 모델 유형과 구성 매개변수를 입력으로 받아, 지정된 모델 유형에 따라 해당 모델 인스턴스를 반환한다. 이를 통해 사용자는 합성곱 신경망(CNN), 장단기 메모리 네트워크(LSTM), 주의 기반 모델 등 다양한 아키텍처 간에 쉽게 전환할 수 있다. process_data는 입력 특징, 목표 레이블, 운영 모드(훈련, 평가 또는 예측) 및 구성 매개변수를 입력으로 받아들인다. 모델을 초기화하고, 입력 특징을 모델에 통과시켜 로짓(logits)을 계산한 후, 예측 결과를 도출한다. 이 과정에서 손실 값을 계산하고, 훈련 모드에서는 최적화 작업을 정의하여 모델의 성능을 향상시킨다. 평가 모드에서는 정확도를 기록하고, 최종적으로 예측, 손실 및 메트릭을 포함한 EstimatorSpec을 반환한다.

(표 1) DeepOpt 모듈 알고리즘

```

알고리즘
BEGIN
  IMPORT libraries (tensorflow, numpy, logging)

  FUNCTION initialize_model(model_type, config):
    RETURN model instance based on model_type

  FUNCTION process_data(inputs, targets, mode, config):
    model = initialize_model(flags.FLAGS.model_type,
    config)
    outputs = model(inputs)
    results = {'classes': ARGMAX(outputs), 'probabilities':
    SOFTMAX(outputs)}

    IF mode == PREDICT:
      RETURN EstimatorSpec with results

    loss_value = sparse_softmax_cross_entropy(outputs,
    targets)
    LOG loss_value

    IF mode == TRAIN:
      RETURN EstimatorSpec with training operations

    accuracy_value = METRICS_ACCURACY(targets,
    results['classes'])
    LOG accuracy_value
    RETURN EstimatorSpec with results, loss_value, and
    metrics

  FUNCTION train_model(flags_obj, model_processor):
    CONFIGURE estimator
    TRAIN model

  FUNCTION evaluate_model(flags_obj, model_processor):
    EVALUATE model
    
```

```

  FUNCTION make_predictions(flags_obj,
  model_processor):
    RETURN predictions

  FUNCTION main_execution():
    IF running_mode == 'eval':
      CALL evaluate_model(flags.FLAGS,
      process_data)
    ELSE:
      CALL train_model(flags.FLAGS, process_data)
  IF script is main:
    CALL main_execution()
  END
    
```

III. 결론

본 연구에서는 메타유전체 분류의 정확성과 속도를 개선하기 위해 딥러닝 기반 모듈인 DeepOpt를 제안하였다. 메타유전체학의 발전과 함께 고처리량 시퀀싱 기술이 생성하는 방대한 데이터의 효율적인 분석이 필수적임을 인식하고, 기존의 데이터베이스 정렬 기반 방법의 한계를 극복하고자 하였다. DeepOpt 모듈은 입력 데이터를 사전 학습하여 특징을 추출하고, 이를 기반으로 데이터베이스 기반 분류 도구를 통해 최종 분류를 수행함으로써, 기존 방법보다 향상된 성능을 기대할 수 있다. 이 시스템은 특히 알려지지 않은 종의 분류가 필요한 상황에서도 강점을 발휘하며, 대규모 시퀀싱 리드에 대한 신속하고 정확한 분류를 가능하게 한다.

사 사 문 구

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2024년도 문화체육관광연구개발사업으로 수행되었음(과제명: 3D 홀로그래픽 인포테인먼트 시스템 디자인 R&D 전문 인력 양성, 과제번호: R-2024-00401213, 기여율: 100%)

참 고 문 헌

- [1] Luis M Rodriguez-R and Konstantinos T Konstantinidis. Estimating coverage in metagenomic data sets and why it matters, The ISME Journal, 8(11),. Oxford University Press (OUP), : 2349 - 2351, 2014.
- [2] Victor Kunin et al. A Bioinformatician's Guide to Metagenomics, Microbiology and Molecular Biology Reviews, 72(4),. American Society for Microbiology, : 557 - 578, 2008.
- [3] Gustavo Hassemer et al. The application of high-throughput sequencing for taxonomy: The case of Plantago subg. Plantago (Plantaginaceae), Molecular Phylogenetics and Evolution, 138. Elsevier BV, : 156 - 173, 2019.
- [4] Laura-Jayne Gardiner et al. Re-purposing software for functional characterization of the microbiome, Microbiome, 9(1),. Springer Science and Business Media LLC, 2021.
- [5] Nathalie Bonin et al. MEGARes and AMR++, v3.0:

an updated comprehensive database of antimicrobial resistance determinants and an improved software pipeline for classification using high-throughput sequencing, *Nucleic Acids Research*, 51(D1),. Oxford University Press (OUP), : D744 - D752, 2022.

- [6] Donovan H. Parks et al. A complete domain-to-species taxonomy for Bacteria and Archaea, *Nature Biotechnology*, 38(9),. Springer Science and Business Media LLC, : 1079 - 1086, 2020.
- [7] Florian Mock et al. Taxonomic classification of DNA sequences beyond sequence similarity using deep neural networks, *Proceedings of the National Academy of Sciences*, 119(35),. *Proceedings of the National Academy of Sciences*, 2022.
- [8] Marcel Martínez-Porchas, Enrique Villalpando-Canchola, and Francisco Vargas-Albores. Significant loss of sensitivity and specificity in the taxonomic classification occurs when short 16S rRNA gene sequences are used, *Heliyon*, 2(9),. Elsevier BV, : e00170, 2016.

소셜 네트워크 데이터 기반의 트렌드룩에 대한 오피니언마이닝 분석

송채빈*, 이서영*, 서지훈**
강남대학교

202204226@kangnam.ac.kr, pinklee831@naver.com, jihoon@kangnam.ac.kr

Opinion mining analysis of trend looks based on social network data

Song Chae Bean*, Lee Seo Young*, Seo Ji Hoon**
Kangnam University*, Kangnam University**

요 약

소셜 네트워크 서비스로 인하여 인류는 다양한 정보를 실시간으로 획득할 수 있는 시대를 개척하였다. 이러한 과급력은 지구 반대편에 있는 국가에서 일어나고 있는 정보도 쉽게 공유받을 수 있다. 소셜 네트워크 서비스가 탄생하고 가장 큰 혜택을 누린 분야가 패션이다. 본 논문은 패션에 민감한 영향이 있는 트렌드룩에 대한 소셜 네트워크 데이터를 수집하여 평판 분석을 수행하였다.

I. 서 론

소셜 네트워크 서비스의 등장으로 정보의 전달이 기하급수적으로 증가함에 따라 패션 분야는 가장 많은 수혜를 보게 되는 분야로 인식되었다. 그러므로 본 논문에서는 클라이언트의 소셜네트워크 데이터를 활용하여 평판 분석을 수행하였다. 자체 개발한 한국어 감성 사전을 구축하고, 소셜 네트워크 데이터를 수집하여 전처리를 진행하고 트렌드룩에 대한 평판을 분석하여 의류 마케팅에 활용할 수 있는 감성 트렌드 솔루션을 도출하였다. 본 연구는 다음과 같은 프로세스로 절차로 분류된다. 첫 번째, 클러스터링을 통한 데이터의 필터링 과정, 두 번째, 소셜 네트워크 데이터의 성향에 맞는 한국어 어휘의 필터링 과정, 세 번째, 어휘의 연관규칙을 적용한 정확성 구성 과정, 네 번째, 과거에 선행 연구를 기반으로 OSAR 알고리즘을 적용한 감성 사전 구축, 다섯 번째, 트렌드룩 분석 데이터 도출로 결론을 내린다.

II. 본문

2.1 연구 방법론

본 연구에서는 소셜 네트워크 데이터를 기반으로 트렌드룩을 분석 및 예측하기 위해 OSAR(Opinion Sentiment Association Rule) 알고리즘을 적용한 감성 사전을 구축하고, 이를 활용하여 트렌드룩에 대한 분석을 진행하였다. 연구 절차의 단계는 데이터 수집 및 전처리, 감성 사전 구축, OSAR알고리즘 적용, 트렌드룩 분석으로 진행된다.

2.2 데이터 수집 및 전처리

소셜 네트워크 서비스에서 패션에 관련된 키워드로 의류에 대한 비정형 데이터를 수집하였다. 각 데이터는 비정형 텍스트로 구성되어 있으며, 이를 분석 가능한 형태로 전처리하기 위해 다음 같은 과정을 거쳤다.

첫 번째, 토큰화 (Tokenization): 문장을 단어 단위로 분할한다. 두 번째, 불필요한 데이터 제거: 이모티콘, URL, 특수문자 등 불필요한 요소를 제거하였다. 세 번째, 형태소 분석: 한국어 특성에 맞게 명사, 동사 등의 형태소 분석을 실시하여 단어의 품사를 구분하였다.

2.3 감성 사전

감성 사전은 소셜 네트워크 서비스에서 수집된 패션 관련 텍스트데이터를 기반으로 구축되었으며 감성 사전 구축 과정은 다음과 같다. 첫 번째 수집한 텍스트 데이터를 유사한 의미를 가진 그룹으로 나누기 위해 K-means 클러스터링을 적용하였다. 클러스터링된 결과는 트렌드 분석을 위한 패션 카테고리별로 분류되었다. K-means 알고리즘의 목적 함수는 다음과 같습니다.

$$J = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (1)$$

여기서, k는 클러스터의 개수이며, C_i 는 i번째 클러스터를 의미하고, x는 클러스터 C_i 에 속한 데이터 포인트를 의미한다. 두 번째로 두 단어가 함께 등장할 확률을 분석하여, 연관성이 있는 단어들을 찾는 작업을 수행하였다.

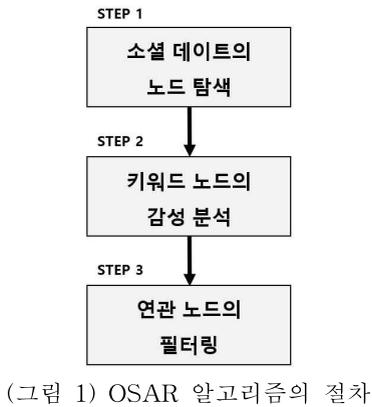
2.4 OSAR 알고리즘 적용

OSAR 알고리즘은 감성 사전을 기반으로 소셜 네트워크 데이터에서 도출된 내용을 분류하고, 트렌드 예측을 수행하는 데 사용되었다. OSAR 알고리즘은 다음과 같은 단계를 거친다.

- 1) 트윗 텍스트의 노드 탐색: 트윗에서 감성 단어를 포함하는 문장을 탐색하여 노드로 분류한다.
- 2) 키워드 노드의 감성 분석: 각 노드에서 중심이 되는 키

워드를 찾고, 해당 키워드가 긍정적 또는 부정적 감성을 갖는지 분석한다.

3) 연관 노드 필터링: 키워드와 연관된 단어를 찾아내어 트렌드에 중요한 영향을 미치는 단어를 필터링한다.



(그림 1) OSAR 알고리즘의 절차

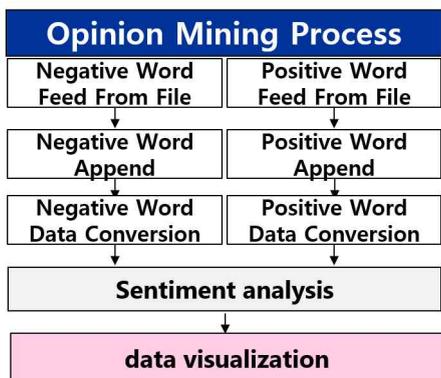
OSAR 알고리즘의 핵심 규칙은 다음과 같이 적용된다.

(표 1) OSAR 알고리즘 규칙

<p>*OSAR 규칙</p> <p>Rule 1: 트윗 문장에서 키워드 노드 탐색: $\{P1, P2, \dots, Pn\}$ 트위터 문장에서 긍정/부정 단어를 포함한 키워드 노드 P_i를 탐색한다.</p> <p>Rule 2: 연관된 단어 노드 탐색: $\{N1, N2, \dots, Nm\}$ 키워드와 연관된 단어들을 N_i로 구분하며, 이를 기반으로 트렌드류의 분석을 수행한다.</p> <p>Rule 3: 상위 4개의 노드에 대해 연관성 랭킹을 부여한다.</p>

2.5 데이터 분석 기법

다음은 오피니언 마이닝에 대한 분석기법에 대한 절차는 다음과 같다.



(그림 2) 오피니언 마이닝 분석 절차

본 논문에서는 자체적으로 구축한 감성 사전을 서버에 저장하고 소셜 네트워크 데이터에 대한 오피니언 마이닝 분

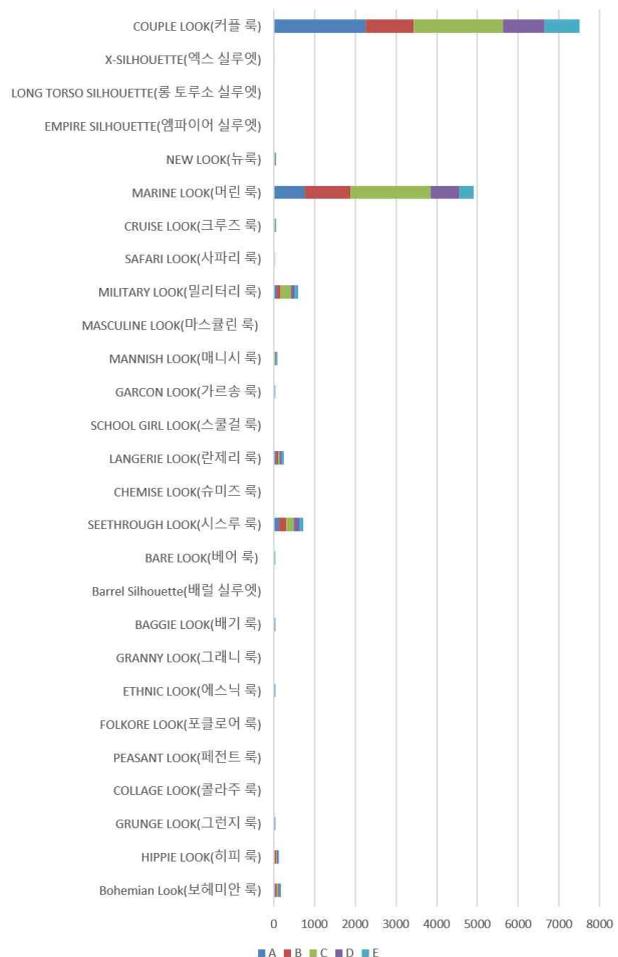
석을 수행한다. 감성 단어는 긍정과 부정으로 데이터 결과를 추출할 수 있으며 저장된 데이터 사전의 일부는 아래의 그림과 같다.

Row No.	word	states
1	조마	y
2	간지	y
3	간지나는줄	y
4	감각적이면서	y
5	강인함	y
6	갈증쯤	n
7	개성있는	y
8	개미배	y
9	개진리인	y
10	개전달	y
11	관찰다	y
12	거부감	n
13	거칠고	n
14	걱정됨	n
15	걱정했는데	n
16	결레 짝	n
17	견량아요	y
18	겸비한	y
19	고민	n
20	고민 고민	n
21	괴롭다는	n
22	편향인데	y
23	괴로워요	n
24	구려짐	n
25	구린	n
26	귀여운	y

(그림 3) 감성 데이터 샘플링

2.6 데이터 분석

본 논문의 분석 결과는 아래와 같다.



(그림 4) 언급도 결과

본 논문에서 수집한 소셜 네트워크 데이터의 수집 건수는 71만여 건의 데이터를 확보하였으며 여기서 분석의 빈도수가 가장 높은 “트렌드룩”을 기반으로 데이터를 분석하였다. 소셜 네트워크 데이터에서 언급도가 가장 많이 도출된 커플룩에 대한 결과이다. 커플룩의 경우 언급도가 높게 나왔지만 긍정도 부분에서도 높은 수치를 보였다. 두 번째로 높은 언급도를 보인 마린룩도 대부분 긍정의 의견이 많이 도출되었다. 전반적으로 계절에 유행을 많이 타지 않는 의류들이 소셜 네트워크 데이터에 대부분 언급되었다. 또한 한국인들의 정서에 맞는 항목들임을 확인할 수 있었다. 데이터 분석의 시기가 봄에서부터 가을까지인 것을 고려할 경우 여름 의류가 가장 인기도가 높은 것을 확인할 수 있었다. 반면, “밀리터리 룩”과 “시스루룩”은 상대적으로 부정적인 평가가 높았다. 이러한 감정 분석은 패션 브랜드가 마케팅 전략을 세우는 데 중요한 지표로 활용될 수 있다.

III. 결론

본 논문은 과거의 선행 연구를 기반으로 소셜 네트워크 데이터를 기반으로 트렌드룩을 분석하였다. IT와 패션의 융합 기술은 향후 더욱 발전할 가능성이 크며, 이를 통해 보다 정교한 트렌드 분석이 가능해질 것이다. 또한, 감정 사진의 정확성을 높이기 위해서는 다양한 SNS 데이터와 사용자 네트워크 분석이 필요하며, 향후 연구에서는 이를 고려한 더 구체적인 알고리즘 개발이 요구된다.

참 고 문 헌

- [1] James Manyuka, et al., 2011, "Big Data : The Next Frontier for Innovation, Competition, and Productivity", McKinsey Global Institute, America.
- [2] 이운주, 서지훈, 최진탁. (2016). SNS 텍스트 콘텐츠를 활용한 오피니언 마이닝 기반의 패션 트렌드 마케팅 예측 분석. 한국정보기술학회, 12(12), 163-170.
- [3] Ghose, P. G. Ipeirotis and A. Sundararajan, 2007, "Opinion Mining Using Econometrics : A Case Study on Reputation System" Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic, pp.416-423.
- [4] A. Mittermayer and G. F. Knolmayer, 2006, "Text Mining Systems for Market Response to News: A Survey", The Institute of Information Systems, University of Bern, Switzerland.
- [5] Manovich, 2011, "Trending : The Promises and the Challenges of Big Social Data", Debates in the Digital Humanities, America.

대형 생활 폐기물 분류 애플리케이션

박태영, 김도영, 홍기동, 김유희
 신한대학교 소프트웨어융합학과

typark0916@gmail.com, dajaba2462@naver.com, gidong010203@gmail.com,
 euhkim@shinhan.ac.kr

Large Household Waste Classification Application

Tae Young Park, Do Young Kim, Gi Dong Hong, Eu Hee Kim
 Shinhan Univ.

요약

환경 오염과 지구 온난화가 심각해지면서 폐기물 배출 처리는 중요한 사회적 문제로 대두되고 있다. 특히 대형 폐기물의 배출은 수수료 스티커를 통해 이루어지는데, 이 스티커는 폐기물의 유형, 크기, 형태에 따라 가격이 달라지며, 지역마다 다른 규정으로 인해 신규 거주자나 고령자에게는 불편을 초래할 수 있다. 본 과제의 목적은 이러한 문제를 해결하기 위해 의정부시를 기준으로 대형 폐기물 수수료를 산정할 수 있는 웹 애플리케이션을 개발하는 것이다. 이를 통해 대형 폐기물 배출 절차를 간소화하고, 사용자 친화적인 환경을 제공하여 불편함을 최소화하고자 한다. 본 프로젝트는 성능 비교를 통해 선정된 ResNet-50 모델 개발을 통해 이미지 분류 시스템 구축과 웹 서비스 개발을 통한 사용자 편의성 증대를 목표로 하며, 이를 통해 환경 보호와 도시 관리에 기여할 수 있는 방안을 제시한다.

학습할 수 있는 모델이 필요하여 선택하게 되었다.

I. 서론

한국환경공단만의 전국 폐기물 발생 및 처리 현황 조사에 따르면, 발생량은 2020년 총 폐기물 발생량은 19,546만 톤으로 전년 대비 7.7% 증가하였고, 2021년에는 19,738만 톤으로 전년 대비 1% 증가하면서 꾸준히 증가하고 있다. 환경 보호를 위해서 폐기물 배출 관리는 중요시되지만 현재 폐기물 처리 과정은 굉장히 불편하다. 폐기물의 크기, 형태뿐만 아니라 지역에 따라 가격과 종류가 달라지며, 그에 따라 신규 거주자나 고령자는 수수료 금액 산정에 어려움을 겪고 있다[5].

본 연구는 ResNet-50, EfficientNet-B7, ViT 세 가지 모델을 선정하여 성능을 비교한다. 그 중 가장 좋은 성능을 보이는 이미지 분류 모델을 적용하여 의정부시를 기준으로 폐기물 수수료를 출력하는 웹 애플리케이션을 개발한다.

II. 관련 연구

2.1 ResNet-50

잔차 연결(skip connections)을 도입한 구조로, 깊은 네트워크 학습에서 발생할 수 있는 기울기 소실 문제를 효과적으로 완화한다. 이를 통해 더 깊고 복잡한 모델을 안정적으로 학습할 수 있어, 다양한 복잡성을 가진 이미지에서 좋은 성능을 보인다. 또한 ResNet-50은 다양한 데이터셋과 컴퓨터 비전 과제에서 범용적으로 우수한 성능을 보여왔다[1]. 이는 기존의 많은 연구에서 입증된 안정성과 신뢰성 덕분에 기준 모델로 적합하다. 본 과제의 경우에도 대형 폐기물의 다양한 유형과 복잡한 특징을 안정적으로

2.2 EfficientNet-B7

Neural Architecture Search를 통해 설계된 모델로, 폭, 깊이, 해상도를 동시에 최적화하여 적은 매개변수로도 높은 정확도를 달성할 수 있다. 이 모델은 특히 연산 자원이 제한된 환경에서도 우수한 성능을 발휘하므로, 향후 서비스의 실시간 처리와 모바일 환경 확장 가능성을 고려할 때 매우 적합하다[2]. EfficientNet-B7은 최신 CNN 아키텍처 중에서도 적은 연산량으로 높은 정확도를 제공하는 모델로, 이는 실시간 응답성과 처리 속도가 중요한 본 과제에서 매우 중요한 요소이다. 특히, 대형 폐기물 이미지의 고해상도 처리가 요구되는 상황에서 뛰어난 성능과 효율성을 제공할 수 있어 선택하였다. 또한 EfficientNet-B7의 구조적 효율성은 모바일 및 저전력 환경에서도 높은 정확도를 유지하면서 빠르게 처리할 수 있도록 해준다.

2.3 ViT (Vision Transformer)

최근 연구에서 CNN 기반 모델들을 능가하는 성능을 보여주는 혁신적인 모델로, 대규모 데이터에서 강력한 성능을 발휘한다. ViT의 구조는 폐기물 이미지의 복잡성과 다양한 패턴을 처리하는 데 적합하여, 기존 CNN 모델들이 해결하기 어려운 문제를 보완할 수 있다[3]. Transformer 아키텍처를 활용하여 이미지를 패치 단위로 나누어 학습하며, 이를 통해 이미지의 전역적 특징을 학습한다. 이는 기존의 CNN이 국소적 특징에 의존하는 것과 달리, 이미지 전체의 맥락을 이해하고 복잡한 패턴을 처리할 수 있어, 다양한 폐기물 이미지를 정확하게 분류하는데 유리하다. ViT는 특히 대규모 데이터셋에서 탁월한 성

능을 발휘하기 때문에, 데이터셋이 확장되거나 다양한 환경에서 학습을 지속해야 하는 경우에도 높은 유연성과 적응력을 제공한다. 이는 본 과제의 이미지 분류 성능을 극대화하고, 향후 데이터 확장 시에도 높은 활용도를 보일 수 있어 ViT를 선택하였다.

세 모델 모두 각각의 강점과 특성을 통해 대형 폐기물 이미지 분류 문제를 효과적으로 해결할 수 있는 가능성을 보여준다. ResNet-50은 기존 모델로서의 안정성과 성능을 제공하고, EfficientNet-B7은 효율적인 연산과 정확도를 바탕으로 실시간 처리에 적합하며, ViT는 최신 기술로서 이미지의 전역적 이해를 통해 복잡한 패턴을 정확히 분류하는 강점을 지닌다. 이러한 이유로 각 모델의 특성과 본 과제의 목표를 종합적으로 고려하여 세 모델을 선정하였으며, 성능 평가 결과를 바탕으로 최적의 모델을 선택하여 웹 애플리케이션에 적용하게 되었다.

III. 데이터 수집

본 연구에서는 대형 폐기물 분류를 위한 학습 데이터셋을 구축하기 위해 2020년에 진행된 서울형 뉴딜 일자리 AI 학습 데이터 구축 사업에서 2,300개와 2020년 행정안전부 공공데이터 기업 매칭 지원 사업으로 수집된 김해시 AI 기반 대형 생활 폐기물 학습데이터 40,329개를 수집하였다[6][7]. 그 중 의정부시에서 다루지 않는 폐기물을 제외하거나, 모델 성능을 저해할 가능성이 있는 불필요한 데이터를 삭제하는 등 데이터 정제를 통해 최종 34,755개의 이미지 데이터셋을 구축하였다. 데이터 정제 과정에서 중복 데이터, 비정상 데이터 등을 필터링하여 데이터셋의 질을 높였으며, 이를 통해 모델 학습의 신뢰성과 정확도를 향상시켰다. 최종 데이터 데이터셋을 학습용 데이터와 검증용 데이터를 8:2 비율로 분할하여 모델을 훈련시켰다. 학습용 데이터는 모델의 가중치를 학습하는 데 사용되었으며, 검증용 데이터는 훈련 과정에서 모델의 성능을 평가하고, 과적합을 방지하기 위해 활용되었다.

IV. 시스템 구성

(표 1) 모델 구현 환경

OS	Microsoft Windows 11 Home
CPU	Intel(R) Core i7-12700F, 2100Mhz
RAM	32GB
GPU	NVIDIA Geforce RTX 4060
Frontend	HTML
Backend	Python (PyTorch, Flask), Apache

V. 과적합 해결

본 연구에서는 PyTorch의 torchvision.transforms 모듈을 활용하여 이미지 데이터 증강을 수행하였다. 데이터 증강은 원본 이미지를 다양한 방식으로 변형하여 데이터셋의 다양성을 증가시키고, 모델의 일반화 성능을 향상시키는 중요한 기법이다. 이를 통해 학습 중 발생할 수 있는 과적합을 방지하고, 모델이 실제 환경에서 다양한 상황에 잘 적응할 수 있도록 하였다. 과적합을 방지하기 위한 방법은

다음과 같다.

5.1 크기 조정

Resize를 통해 입력 이미지를 224x224 픽셀로 크기를 조정하여, ViT와 같은 특정 모델의 입력 형식에 맞춘다. 일관된 크기의 이미지를 사용함으로써 모델의 학습 과정에서 발생할 수 있는 크기 변화로 인한 혼란을 줄인다.

5.2 회전

데이터의 다양성을 높이기 위해 RandomHorizontalFlip과 RandomRotation을 이용했다. 수평 반전과 무작위 회전을 이용하여 데이터의 다양성을 증가시키고 동시에 모델이 다양한 시점에서 객체를 인식하는 능력을 향상시켰다. 이는 객체 탐지 및 분류 문제와 모델의 일반화 능력을 개선시킨다.

5.3 정규화

이미지를 PyTorch 텐서 형식으로 변환하여 정규화에 필요한 형식으로 데이터를 변환한다. 픽셀 값이 스케일링된 이미지를 적용한 후 Normalize 변환으로 정규화를 수행하여 데이터의 평균과 표준 편차를 맞춘다. 일반적으로 ImageNet 데이터셋의 통계치를 사용하여 정규화하므로, 모델의 수렴 속도를 개선하고, 학습 과정에서의 안정성을 높인다. 정규화는 또한 데이터의 스케일을 맞추어 특정 특징이 모델의 학습에 과도한 영향을 미치지 않도록 한다.

5.4 사용자 정의 데이터셋

데이터셋을 구성하는 과정에서 사용자 정의 데이터셋인 CustomDataset 클래스를 구현하여 특정 인덱스의 데이터를 반환할 때 transforms 모듈을 사용하여 변환을 적용한다. 이를 통해 매번 새롭게 변형된 이미지를 모델에 제공할 수 있으며, 이는 데이터 증강의 핵심적인 기전이 된다.

CustomDataset 클래스 내에서 변환이 적용된 이미지를 실시간으로 생성함으로써 모델은 매 에폭마다 다양한 변형된 이미지를 학습하게 된다. 이러한 접근 방식은 데이터의 다양성을 극대화하고, 특정 데이터에 대한 과적합을 방지하는 데 기여한다. 또한 모델은 다양한 변형에 노출됨으로써 실제 환경에서 발생할 수 있는 다양한 변형에 잘 적응할 수 있도록 훈련된다.

VI. 성능 평가 방법

F1 Score는 모델의 성능을 평가하는 데 있어 중요한 지표로, Precision(정밀도)과 Recall(재현율)의 조화 평균을 계산한다. 이는 모델이 얼마나 정확하게 예측했는지(Precision)와 실제 정답을 얼마나 잘 찾아냈는지(Recall)를 동시에 고려하여, 모델의 예측 성능을 종합적으로 평가할 수 있는 지표이다. F1 Score는 다음과 같이 정의된다.

F1 Score가 적합한 이유는 특히 클래스 간 데이터 불균형이 있을 때, Precision과 Recall 사이의 균형을 유지하면서 모델의 전반적인 예측 성능을 평가할 수 있기 때문이다. 본 과제에서는 대형 폐기물 이미지의 다양한 클래스 간 불균형이 존재하기 때문에 F1 Score를 통해 모델이

얼마나 효과적으로 다양한 클래스에 대한 예측을 수행하는지 측정하는 것이 중요하다. 이 지표는 단순한 정확도 (Accuracy)만으로는 확인할 수 없는 모델의 세부적인 성능과 잘못된 분류로 인한 영향을 정량적으로 평가하는 데 큰 도움이 된다.

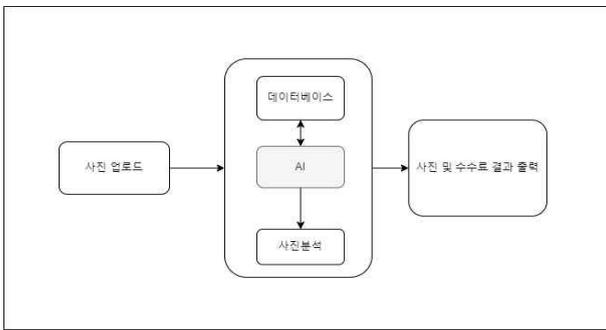
모델 성능을 비교한 결과, 각 모델의 F1 Score와 정확도를 기준으로 최적의 모델을 선정하였으며, 이로써 웹 애플리케이션에서의 실시간 분류 성능을 극대화할 수 있었다.

VII. 서비스 구성

본 연구에서 개발된 대형 폐기물 분류 웹 애플리케이션은 사용자 편의를 증대하기 위해 이미지 분류 모델과 웹 프레임워크 간의 원활한 연동을 목표로 한다. 웹 애플리케이션은 크게 사용자 인터페이스, 서버 측 모델 처리, 데이터베이스 관리, 그리고 최종 결과 출력 단계로 구성된다.

7.1. 사용자 인터페이스

사용자는 웹 애플리케이션을 통해 대형 폐기물 이미지를 업로드한다. 이 인터페이스는 사용자의 이미지를 받는 프론트엔드 역할을 하며, 이미지가 서버로 전달될 수 있도록 파일 업로드 기능을 지원한다.



(그림 1) 서비스 흐름도

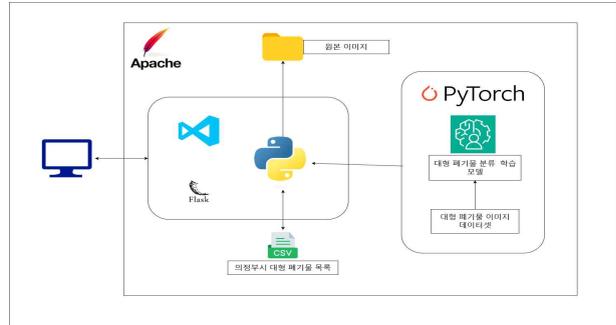
7.2. 서버 측 모델 처리

이미지가 업로드되면 서버 측에서 대기 중인 이미지 분류 모델이 실행된다. 본 연구에서는 성능 비교 후 선정된 ResNet-50 모델이 주요 분류 모델로 사용되며, 효율성 및 모바일 애플리케이션 확장을 위한 EfficientNet-B7 및 Vision Transformer(ViT) 모델도 구축되어 있다. 서버는 Python 기반의 Flask나 Django와 같은 웹 프레임워크를 통해 모델을 관리하며, 각 모델의 기능은 다음과 같다: ResNet-50: 기본 분류 모델로 작동하며, 복잡한 대형 폐기물 이미지에 대한 높은 안정성과 신뢰성을 제공한다. EfficientNet-B7: 연산 자원이 제한된 모바일 환경에서 사용할 수 있는 경량 모델로 설계되었다. 이 모델은 사용자 요구에 따라 서버에서 선택적으로 작동할 수 있다. ViT: 복잡하고 대규모 데이터를 다룰 때 효과적인 Transformer 기반 모델로, 향후 데이터셋 확장 및 고해상도 이미지 분류에 대비한 백업 역할을 수행한다.

각 모델은 서버의 모델 처리 모듈과 연동되어 있어, 사용자 요청 시 이미지 데이터를 입력받아 분류 결과를 산출한다. 이후 분류된 데이터는 일관된 포맷으로 변환되어 다음 단계로 전달된다.

7.3 데이터 관리

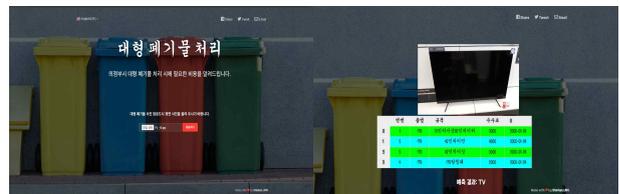
이미지 업로드 후 학습 모델이 분류한 결과는 애플리케이션 내 csv 파일을 통해 분류된 폐기물 유형, 수수료 결과를 출력한다. 원본 이미지를 자체적으로 저장하여 사용자는 향후 기록을 조회할 수 있다.



(그림 2) 시스템 구성도

7.4 결과 출력 및 사용자 피드백 수집

분류된 결과는 사용자 인터페이스로 다시 전송되어 사용자에게 폐기물의 분류 및 예상 수수료 정보를 제공한다. 결과 출력 화면에서 사용자 피드백을 수집할 수 있는 기능도 제공하여, 모델의 성능 개선에 필요한 데이터를 지속적으로 확보하고 있다.

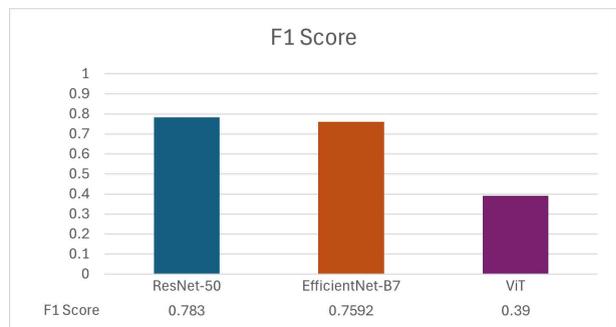


(그림 3) 초기 화면

(그림 4) 수수료 출력 화면

VIII. 모델 평가 및 분석

적합한 모델을 선정하기 위해 이전에 수집한 동일한 데이터셋으로 세 가지 모델의 학습을 진행하였다. 아래는 결과로 출력된 각 모델의 F1 Score를 비교한 그래프이다.



(그림 5) F1 Score 비교 그래프

8.1 ResNet-50

잔차 연결로 깊은 네트워크가 안정적으로 학습을 진행하며, 복잡한 이미지 특징을 잘 포착할 수 있었다. 수집된 데이터셋과 증강된 데이터로부터 다양한 패턴을 학습했기 때문에 상대적으로 더 나은 성능을 발휘할 수 있었다. 데이터 증강을 통해 다양한 변형된 이미지에 대응하는 일반화 능력을 향상시킬 수 있었으며, 이는 ResNet의 가장 높은 F1 Score로 이어졌다.

8.2 EfficientNet-B7

효율성을 최적화한 설계로 파라미터 수를 줄이면서도 높은 성능을 유지하지만, 매우 깊은 네트워크가 아니기 때문에 복잡한 이미지 특징을 학습하는 데 상대적으로 덜 유리했을 수 있다. 데이터 증강과 조기 종료는 과적합 방지에 도움을 주었으나, ResNet만큼 깊은 네트워크가 아니므로 더 복잡한 패턴 학습에서 제한이 있었다. 특히 EfficientNet의 효율성 설계가 복잡한 데이터셋에서 최고 성능을 발휘하는 데는 약간의 한계를 보였다.

8.3 ViT

대규모 데이터셋에서 강력한 성능을 발휘하지만, 현재 수집된 데이터셋은 효과적으로 학습하기에 상대적으로 작은 규모일 수 있다. 비록 데이터 증강이 적용되었지만 이러한 소규모 데이터셋에서 충분한 일반화 능력을 발휘하지 못했을 것으로 예상된다. ViT는 이미지를 작은 패치로 나누어 처리하는 방식이므로, 이미지의 세부 정보가 손실되거나, 중요한 지역적 정보가 충분히 반영되지 않을 수 있다. 수집된 데이터셋과 증강된 데이터로도 이 구조적 한계를 극복하기 어려웠던 것으로 보인다.

각 모델의 아키텍처 특성, 네트워크 깊이, 데이터셋 크기 및 처리 방식에 따라 성능 차이가 발생했다. ResNet과 EfficientNet은 비교적 안정적인 성능을 보였지만, ViT는 데이터셋 크기와 모델 특성 간의 불일치로 인해 낮은 성능을 기록했다. 위와 같은 분석에 따라 가장 높은 성능을 보인 ResNet을 서비스에 적용하였으며, EfficientNet은 성능 향상을 통해 향후 모바일 버전 확장에 적용할 가능성을 보였다.

IX. 결론 및 향후 과제

본 과제에서 구축한 대형 폐기물 분류 서비스는 다양한 활용으로 확장성과 지속성을 가지고 있다. 기술적인 발전과 함께 더 정확하고 빠른 식별 기술을 도입하여 사용자 편의성을 높일 수 있다. 현재 의정부시를 대상으로 제작하였지만 공공 및 민간 부문 간의 협력을 통해 대상 지역을 확장할 수 있다. 지자체와 협력을 통해 서비스 영역을 확장하면서 환경 보호 및 도시 관리에 더 큰 기여를 할 수 있을 것이다. 특히 모바일 애플리케이션을 추가하여 실시간 촬영을 통해 사용자는 폐기물 배출 수수료를 확인하고, 환경 보호를 위한 추가적인 시스템을 마련하는 발판으로 삼을 수 있다.

ACKNOWLEDGMENT

감사의 글 : “본 연구는 2023년도 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업 지원을 받아 수행되었음” (2023-0-00089)

참 고 문 헌

[1] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer

Vision and Pattern Recognition (CVPR), 770–778. DOI: 10.1109/CVPR.2016.90

- [2] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning (ICML), 6105–6114. PMLR. DOI: 10.48550/arXiv.1905.11946
- [3] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Proceedings of the International Conference on Learning Representations (ICLR). DOI: 10.48550/arXiv.2010.11929
- [4] Saha, S., & Naik, R. (2021). A Comprehensive Review of MobileNet Architecture and its Variants for Edge-Devices. Proceedings of the IEEE 18th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), 120–125.
- [5] 한국환경정책평가연구원. (2021). 한국의 폐기물 관리 정책 연차 보고서.
- [6] 서울열린데이터광장, 2020년 5월, 2024년 7월 26일 접속, <https://data.seoul.go.kr/etc/aiEduData.do>
- [7] 공공데이터포털, 2021년 1월 27일, 2024년 8월 2일 접속, <https://www.data.go.kr/data/15076741/fileData.do>

분산 클라우드 환경에서 딥러닝을 위한 파라미터 시스템에 관한 연구

김수연, 문석재*
 광운대학교 스마트융합대학원
 광운대학교 스마트융합대학원*
 {rlatndus0304, msj8086}@kw.ac.kr

A study on a Parameter System for Deep Learning in a Distributed Cloud Environment

Su-Yeon Kim, Seok-Jae Moon*
 Graduate School of Smart Convergence, KwangWoon University
 Graduate School of Smart Convergence, KwangWoon University*

요약

이 연구는 대규모 데이터를 처리하고 딥러닝 모델을 학습하는 과정에서 발생하는 자원 제약, 데이터 불균형, 학습 시간 지연 등의 문제를 해결하기 위해 수행되었다. 특히 로컬 시스템의 한계를 극복하고, 분산 클라우드 환경을 통해 자원 관리와 모델 성능을 개선하는 방법을 찾는 것을 목적으로 한다. 이를 위해 여러 서버에서 서로 다른 하이퍼파라미터 설정을 적용한 병렬 딥러닝 학습을 사용하였으며, 각 서버에서 생성된 학습 결과를 실시간으로 모니터링하여 최적의 하이퍼파라미터 값을 선택하고 조정하는 파라미터 시스템을 제안하였다. 제안된 시스템은 대규모 데이터 처리에서 자원을 효율적으로 활용하도록 하며, 딥러닝 모델의 학습 성능을 최적화하고 학습 시간을 단축할 수 있다.

I. 서론

빅데이터 시대에서의 인공지능과 딥러닝 기술은 제조, 금융, 의료 등 다양한 산업 분야에서 핵심적인 역할을 하고 있으며[1, 2], 대규모 데이터의 효율적인 처리가 필수적이다. 이러한 요구는 클라우드 컴퓨팅 환경에서 제공하는 분산 처리 기술을 통해 충족되고갯폼으로 자리 잡고 있다[3]. 그러나 로컬 있으며, 클라우드는 유연한 자원 관리와 높은 연산 능력을 바탕으로 딥러닝 모델의 학습과 배포에 적합한 플랫폼에서 대규모 데이터를 처리하는 데는 자원 제약과 데이터 불균형 문제가 여전히 일어나며 모델 성능 저하와 학습시간의 지연, 자원 관리의 어려움이 발생한다[4]. 이를 해결하기 위한 대안으로 분산 딥러닝이 주목받고 있으며, 여러 서버에서 병렬 처리를 통해 학습 속도를 향상시킬 수 있다. 특히 분산 파라미터 서버는 대규모 데이터 처리와 모델 학습에서 중요한 역할을 한다[5]. 하지만 분산 딥러닝 환경에서는 복잡한 설정, 하이퍼파라미터 최적화의 어려움, 다중 모델 학습의 한계 등 여러 도전 과제가 존재한다[6]. 특히 서버 간 통신 비용과 동기화 문제는 성능 최적화에 방해가 될 수 있으며, 데이터 불균형으로 인한 편향된 학습 결과가 발생할 수 있다. 따라서 이러한 문제를 해결하고 성능을 극대화하기 위해서는 자원의 효율적인 분배와 하이퍼파라미터 조정이 필수적이다. 이에 본 연구에서는 분산 클라우드 환경에서 딥러닝을 위한 효율적인 파라미터 시스템을 제안하고자 한다.

II. 본문

2.1 제안시스템

기존 연구는 분산 컴퓨팅 자원의 효율성을 극대화하고 통신 비용을 줄이며, 모델 병렬화를 통해 대규모 데이터를 효과적으로 처리하는 방법이 제시되었다. 이러한 방법들은 대규모 데이터 처리에 있어 중요한 성과를 거두었으나, 여전히 하이퍼파라미터를 설정과 학습 완료까지의 긴 시간적 제약이 존재한다. 특히, 대규모 데이터나 복잡한 딥러닝 모델을 학습할 때는 이러한 방식이 많은 시간이 소요되고, 자원의 비효율적인 사용을 초래할 수 있다. 본 연구에서는 그림 1에 제시된 것처럼, 분산 클라우드 환경에서 딥러닝을 위한 파라미터 시스템을 제안한다.

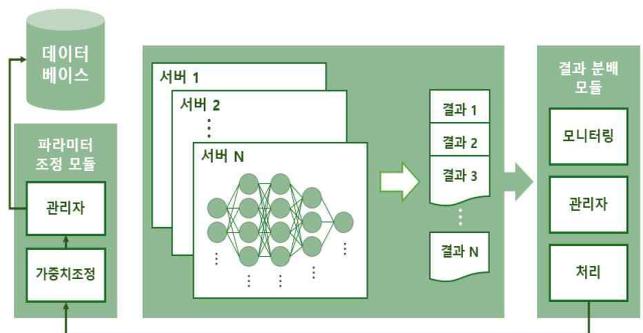


그림 1. 딥러닝을 위한 파라미터 시스템

제안된 시스템은 여러 서버에서 각기 다른 하이퍼파라미터 설정을 적용하여 병렬 학습을 수행한다. 동일한 훈련 데이터를 각 서버에 전송하고, 서버별로 배치 크기, 학습률, 에

포크 등의 하이퍼파라미터를 다르게 설정함으로써 여러 설정을 동시에 실험하여 다양한 학습 결과를 도출할 수 있다. 결과 분배 모듈은 각 서버에서 생성된 학습 결과를 실시간으로 모니터링하고, 서버에서 얻은 결과를 종합적으로 분석하여 최적의 성능을 보이는 하이퍼파라미터 설정을 파악한다. 이 설정은 파라미터 조정 모듈에 전달되며, 조정된 하이퍼파라미터 값은 다시 서버로 전송되어 추가 학습을 수행하게 된다. 이러한 과정은 반복적으로 이루어지며, 최종적으로 최적의 모델 성능을 도출할 수 있다.

표 1. 하이퍼파라미터 조정 알고리즘

Algorithm 1

```

1: Input: Server list  $S = \{S_1, S_2, S_3, S_4\}$ , MNIST dataset, Hyperparameters
   (batch sizes, learning rates, epochs)
2: Output: Final accuracy from each server
3: procedure INITIALIZEMODEL
4:   Initialize CNN model with two convolutional layers and two fully connected layers
5:   Return model parameters (weights, biases)
6: end procedure
7: procedure LOADDATA(batch_size)
8:   Load MNIST dataset with the specified batch size
9:   Return train_loader and test_loader
10: end procedure
11: procedure TRAINMODEL(server  $S_i$ , train_loader, hyperparameters)
12:   Initialize optimizer and loss function (SGD, CrossEntropyLoss)
13:   for epoch = 1 to hyperparameters.epochs do
14:     for each batch of data in train_loader do
15:       Forward pass through the CNN model
16:       Compute loss and backpropagate gradients
17:       Update model parameters using optimizer
18:     end for
19:   end for
20:   Return trained model weights
21: end procedure
22: procedure EVALUATEMODEL(model, test_loader)
23:   Set model to evaluation mode
24:   Initialize counters for correct predictions and total samples
25:   for each batch of data in test_loader do
26:     Forward pass through the CNN model
27:     Compare predicted results with actual labels
28:     Update accuracy based on correct predictions
29:   end for
30:   Return final accuracy
31: end procedure
32: procedure MAIN
33:   Initialize global model parameters and hyperparameters
34:   for each server  $S_i \in S$  do
35:     Initialize  $S_i$ 's model with global parameters
36:     train_loader, test_loader  $\leftarrow$  LOADDATA(server's batch size)
37:     Train server  $S_i$ 's model using TRAINMODEL( $S_i$ , train_loader, hyperparameters)
38:   end for
39:   for each server  $S_i \in S$  do
40:     Accuracy  $\leftarrow$  EVALUATEMODEL(trained model, test_loader)
41:     Output final accuracy for server  $S_i$ 
42:   end for
43: end procedure
44: Call Main

```

III. 결론

본 연구에서는 분산 클라우드 환경에서 딥러닝 성능을 극대화할 수 있는 효율적인 파라미터 시스템을 제안하였다. 제안된 시스템은 각 서버에서 독립적으로 학습을 진행하면서도 최적의 하이퍼파라미터 값을 공유하고 조정함으로써 학습 성능을 극대화하는 것을 목표로 한다. 이를 통해 기존에 발생하던 자원 관리의 어려움, 데이터 불균형으로 인한 성능 저하, 학습 시간 지연 등의 문제를 해결할 수 있다. 또한 하이퍼파라미터를 조정을 통해 딥러닝 모델 학습의 효율성을 향상시키고, 최적의 학습 환경을 유지할 수 있다. 향후 연구에서는 더욱 복잡한 모델에 적용하여 성능을 향상시킬 수 있는 방법을 연구할 예정이다.

사 사 문 구

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2024년도 문화체육관광연구개발사업으로 수행되었음(과제명: 3D 홀로그래픽 인포테인먼트 시스템 디자인 R&D 전문 인력 양성, 과제번호: R-2024-00401213, 기여율:100%)

참 고 문 헌

- [1] M. Chen, S. Mao, and Y. Liu, "Big Data: A Survey," *Mobile Networks and Applications*, vol. 19, no. 2. Springer Science and Business Media LLC, pp. 171 - 209, Jan. 22, 2014,
- [2] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245. American Association for the Advancement of Science (AAAS), pp. 255 - 260, Jul. 17, 2015,
- [3] Mu Li, David G Andersen, Jun Woo Park, Alexander J Smola, Amr Ahmed, Vanja Josifovski, James Long, Eugene J Shekita, and Bor-Yiing Su. 2014. Scaling distributed machine learning with the parameter server. In 11th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 14). 583 - 598
- [4] Y. Liu, W. Zhang, G. Qin, and J. Zhao, "A comparative study on the effect of data imbalance on software defect prediction," *Procedia Computer Science*, vol. 214. Elsevier BV, pp. 1603 - 1616, 2022,
- [5] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., *Tensorflow: a system for largescale machine learning.*, in: *Osdi*, Vol. 16, Savannah, GA, USA, 2016, pp. 265 - 283.
- [6] W. Zhang, S. Gupta, X. Lian, and J. Liu, "Staleness-aware Async-SGD for Distributed Deep Learning." *arXiv*, 2015,

삼킴장애 환자의 흡인과 침투 검출 알고리즘 개발

임동욱¹, 노시형¹, 문희경²

¹원광대학교 근감소증도발솔루션 센터

²원광대학교 교육혁신원

{dw1316, nosij123, ybnjcw@wku.ac.kr}

Development of an algorithm for detecting aspiration and penetration in patients with dysphagia

Dong-Wook Lim¹, Si-Hyeong Noh¹, Hee-Kyung Moon²

¹Sarcopenia Total Solution Center, Wonkwang University,

²Institute for Educational Innovation, Wonkwang University

요 약

본 논문은 삼킴장애 환자의 흡인 및 침투 감지를 위한 딥러닝 기반의 검출 알고리즘을 제안한다. 삼킴장애는 다양한 음식물이 식도로 안전하게 이동하지 못하고 기도으로 유입될 위험이 있는 상태로, 특히 흡인은 폐렴 등의 심각한 합병증을 유발할 수 있어 조기 감지 및 관리가 중요하다. 본 연구에서는 비디오투시검사(VFSS) 데이터를 기반으로 삼킴 과정 중 흡인 및 침투를 자동으로 탐지할 수 있는 딥러닝 모델을 개발하고자 하였습니다. 제안된 알고리즘은 YOLO 모델을 활용하여 VFSS 영상 내에서 기도와 식도 영역을 자동으로 구분하고, 식별된 영역에서 음식물의 위치 및 이동 경로를 추적하여 침투와 흡인 여부를 분류한다. 또한, 진단지표로 PAS(Penetration-Aspiration Scale)를 사용하고자 한다. 실험 결과, 본 알고리즘은 기존의 수작업 분석 대비 높은 정확도와 신속성을 보여주었으며, 다양한 환경에서도 안정적인 성능을 유지했습니다. 이러한 결과는 삼킴장애 환자의 객관적이고 효율적인 진단과 치료 계획 수립에 도움을 줄 수 있으며, 임상 현장에서의 적용 가능성을 확인하였다.

I. 서 론

삼킴장애는 음식물이 정상적으로 식도로 이동하지 못하고 기도으로 유입되는 현상으로, 특히 흡인으로 이어질 경우 폐렴 등 심각한 합병증을 유발할 수 있어 조기 발견과 관리가 매우 중요하다[1, 2]. 이러한 삼킴장애의 진단을 위해 주로 사용되는 비디오투시검사(VFSS)는 삼킴 과정의 생리적 문제를 분석하는 데 효과적이거나, 영상의 수작업 분석이 필수적이기 때문에 시간과 비용이 많이 소요되는 한계가 있다[3]. 이에 따라, VFSS 영상의 자동 분석을 통해 삼킴장애 환자의 흡인 및 침투를 신속하고 정확하게 탐지할 수 있는 알고리즘 개발이 필요하다.

최근 딥러닝 기술의 발전으로 영상 데이터를 기반으로 한 진단 보조 시스템이 활발히 연구되고 있으며, 특히 YOLO(You Only Look Once)와 같은 객체 탐지 모델은 빠르고 정확하게 다양한 객체를 인식하는 능력을 가지고 있어 의료 영상 분석 분야에서 주목받고 있다. 본 연구에서는 YOLO 모델을 삼킴장애 환자의 VFSS 영상 분석에 적용하여 흡인과 침투 여부를 자동으로 탐지하고자 하였다. 더불어, 삼킴 장애의 심각성을 평가하는 표준 지표인 PAS(Penetration-Aspiration Scale)를 진단 기준으로 활용하여 알고리즘의 성능을 객관적으로 평가하고자 한다.

본 연구의 목표는 YOLO 기반의 자동 탐지 알고리즘을 통

해 삼킴장애 환자의 흡인 및 침투를 실시간으로 분석함으로써 기존 수작업 분석 대비 진단의 효율성과 정확성을 향상시키는 것이다. 이를 통해 삼킴장애 진단 및 치료 계획 수립의 표준화를 촉진하고, 환자 관리에 있어 보다 신속한 대응을 가능하게 할 것으로 기대된다.

II. 삼킴장애 자동검출 알고리즘

본 논문에서 제안하는 삼킴장애 자동검출 알고리즘으로 YOLO를 채택하였다. 이 모델은 이미지 전체를 단일 패스로 분석하고, 다수의 객체를 실시간으로 탐지하는 데 효과적이어서, VFSS 영상에서 구강, 기도와 식도의 구조를 식별하고, 음식물의 위치와 이동 경로를 추적하는 데 유리하다. 아래는 YOLO 기반 알고리즘을 삼킴장애 진단에 활용한 과정과 특징이다.

2.1 VFSS 영상에서 구강, 기도와 식도 구역의 자동 구분

YOLO는 입력된 VFSS 영상을 그리드 형태로 분할하고, 각 그리드 셀마다 구강, 기도와 식도 영역을 구분하여 검출한다. 이때 각 영역은 경계 상자(bounding box)로 표시되며, 각 경계 상자에 대한 객체 확률을 계산하여 음식물이 이동하는 경로를 파악할 수 있다. 이를 통해 삼킴 과정

중 음식물이 식도로 향하지 않고 기도 쪽으로 이동하는 패턴을 감지할 수 있다.

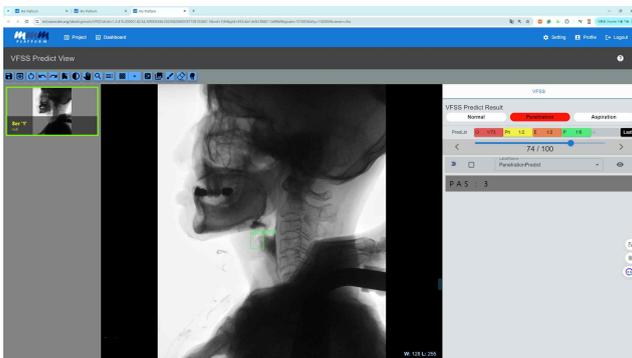
2.2 침투 및 흡인 여부 실시간 검출

본 연구의 핵심은 삼킴 과정 중 발생하는 침투 (penetration)와 흡인(aspiration)을 자동 검출해야 한다. YOLO 모델은 VFSS 영상에서 프레임 단위로 음식물의 위치와 이동을 분석하여, 음식물이 기도 쪽으로 유입되었는지 여부를 감지하고 이를 시각화하여 출력할 수 있다. 또한, Penetration-Aspiration Scale (PAS) 지표를 기반으로, 검출된 객체의 위치와 깊이를 PAS 점수와 비교하여 삼킴장애의 중증도를 평가가 가능하다. 이는 자동 검출 알고리즘이 흡인과 침투를 구체적으로 감지하고 중증도를 판별하는 데 도움을 줄 것으로 기대하고 있다.

2.3 실험 결과 및 성능 평가

YOLO 기반 자동 검출 알고리즘은 PAS 점수와 상관성을 바탕으로 흡인 및 침투를 성공적으로 탐지할 수 있음을 입증했다. 특히 수작업 분석과 비교하여 정확도가 침투, 0.92 그리고 흡인 0.96 높은 정확도와 40초이내의 빠른 분석 속도를 보였으며, 다양한 삼킴 상황에서도 안정적인 성능을 유지하였다. 실험 결과는 삼킴장애 환자 관리에 있어 신속하고 일관된 진단 지원 시스템으로서의 가능성을 확인시켜 주었으며, 임상 현장에서 효율적으로 활용될 수 있는 가능성을 제시하였다[4].

본 연구의 YOLO 기반 자동 검출 알고리즘은 삼킴장애 환자의 흡인 및 침투를 실시간으로 감지하고, 신속한 대응을 가능하게 하여 진단의 표준화를 촉진할 수 있을 것으로 기대한다.



(그림 1) 삼킴장애 검출 및 PAS Score

그림 1은 특정 환자의 동영상파일을 업로드할 경우 개발된 삼킴장애 진단 알고리즘에 의해 검출된 식도부분에 음식물이기도내로 약간 유입되거나 기침 등의 반응을 통해 기도 밖으로 배출되는 상태로 현재 “3”점의 결과로 예측된 결과를 보이고 있다.

III. 결론

본 연구는 삼킴장애 환자의 흡인 및 침투를 실시간으로 감지하기 위해 YOLO 기반 자동 검출 알고리즘을 적용하

였다. VFSS 영상에서 구강, 기도, 식도 구역을 자동으로 구분하고, PAS 점수를 활용해 중증도를 평가함으로써 높은 정확도와 신속성을 확인했다. 이 알고리즘은 삼킴장애 환자의 조기 진단과 관리 효율성 증진에 기여할 수 있으며, 향후 다양한 사례 데이터 확장을 통해 임상에서의 활용 가능성을 높이고자 한다.

현재 임상현장에서는 VFSS 이외에 별도의 셋톱박스를 통해 동영상 파일로 저장하는 방식을 사용하고 있어 이와 관련된 연구가 필요하다. 또한 우리가 개발한 삼킴장애 진단 알고리즘의 고도화를 통해 다기관 검증 데이터를 구축하여 성능평가를 진행할 계획이다.

ACKNOWLEDGMENT

본 연구는 보건복지부의 재원으로 한국보건산업진흥원의 보건의료기술 연구개발사업(RS-2024-00439936) 지원에 의하여 이루어진 것임.

참 고 문 헌

- [1] Baijens LW, Clavé P, Cras P, Ekberg O, Forster A, Kolb GF, Leners JC, Masiero S, Mateos-Nozal J, Ortega O, Smithard DG, Speyer R, Walshe M. European Society for Swallowing Disorders – European Union Geriatric Medicine Society white paper: oropharyngeal dysphagia as a geriatric syndrome. *Clin Interv Aging*. 2016 Oct 7;11:1403–1428. doi: 10.2147/CIA.S107750. PMID: 27785002; PMCID: PMC5063605..
- [2] Cichero, J. A., Lam, P., Steele, C. M., Hanson, B., Chen, J., & Dantas, R. O. (2017). "Development of international terminology and definitions for texture-modified foods and thickened liquids used in dysphagia management: The IDDSI framework." *Dysphagia*, 32(2), 293–314.
- [3] Kim JK, Choo YJ, Choi GS, Shin H, Chang MC, Park D. Deep Learning Analysis to Automatically Detect the Presence of Penetration or Aspiration in Videofluoroscopic Swallowing Study. *J Korean Med Sci*. 2022 Feb 14;37(6):e42. doi: 10.3346/jkms.2022.37.e42. PMID: 35166079; PMCID: PMC8845107.
- [4] Jeong, C.-W.; Lee, C.-S.; Lim, D.-W.; Noh, S.-H.; Moon, H.-K.; Park, C.; Kim, M.-S. The Development of an Artificial Intelligence Video Analysis-Based Web Application to Diagnose Oropharyngeal Dysphagia: A Pilot Study. *Brain Sci*. 2024, 14, 546. <https://doi.org/10.3390/brainsci14060546>

데이터 패브릭 아키텍처를 통한 빅데이터 관리 및 분석 전략 개선

이사랑*, 윤채연**, 문석재***
 광운대학교 스마트융합대학원
 {slmr3027, dbscodus, msj8086}@kw.ac.kr

Efficient Management and Analysis Strategies for Big Data through Data Fabric Architecture

Sa-rang Lee*, Chae-Yeon Yun**, Seok-Jae Moon***
 Graduate School of Smart Convergence, Kwangwoon Univ.

요 약

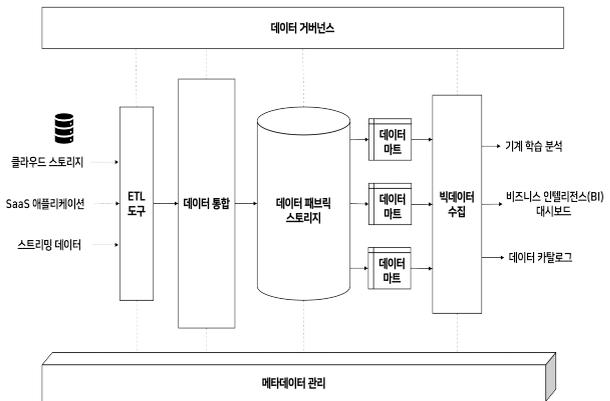
본 논문은 빅데이터 분석의 중요성에 따라 다양한 소스의 데이터를 통합하고 관리할 수 있는 데이터 패브릭의 필요성을 다루었다. 데이터 패브릭 아키텍처는 ETL 도구를 통해 데이터를 수집하고 통합하며, 중앙 스토리지를 통해 일관된 데이터 접근 환경을 제공한다. 데이터 마트와 메타데이터 관리는 비즈니스 요구에 맞춘 데이터 활용을 가능하게 하며, 데이터 거버넌스를 통해 데이터 품질과 보안을 보장한다. 기계학습 분석과 BI 대시보드를 통해 실시간 인사이트를 제공하는 데이터 패브릭은 기업의 데이터 관리와 분석 효율성을 높이고, 혁신적인 데이터 전략 수립을 지원할 가능성을 보여준다. 이러한 요소들은 데이터 중심의 의사 결정과 비즈니스 가치를 극대화하는 데 필수적이다.

I. 서 론

빅데이터 분석은 오늘날 디지털 시대에 가장 중요한 개념 중 하나로, 대규모 데이터의 숨겨진 패턴과 상관관계를 발견하여 의미 있는 정보를 도출하는 과정이다. 빅데이터는 비즈니스 거래, 소셜 미디어, 비디오 분석 등 다양한 분야에서 생성되며, 이를 효과적으로 통합하여 분석하는 것이 점점 더 중요해지고 있다. 효과적인 빅데이터 분석을 위해서는 다양한 데이터 소스를 통합하고 분석하는 것이 필수적이며, 이러한 통합을 가능하게 하는 혁신적인 기술 프레임워크로 데이터 패브릭이 주목받고 있다[1]. 데이터 패브릭은 분산된 데이터 환경에서 실시간으로 다양한 애플리케이션과 플랫폼 간에 데이터를 연결하고 관리할 수 있도록 하여, 클라우드 환경에서 심각한 데이터 사일로 문제를 해결하는 데 기여한다[2]. 빅데이터 저장소는 기업의 요구에 따라 다양하게 존재하며, 데이터를 일관된 방식으로 처리하면서도 분석을 위한 접근성을 극대화한다는 점에서 기존의 데이터 웨어하우스와 데이터 레이크와는 다른 방식으로 데이터 간의 연결과 통합을 제공한다[3]. 이러한 접근 방식은 데이터를 통합하는 동시에 보안과 규정을 준수하며, 기업의 데이터 관리 요구 사항을 충족할 수 있는 유연한 아키텍처를 구현할 수 있다. 데이터 패브릭은 비즈니스의 민첩성을 높이고, 데이터 기반 의사결정을 가능하게 하여 경쟁력을 강화하는 데 기여하며, AI와 머신러닝의 통합을 통해 자동화된 데이터 처리와 실시간 분석을 지원하여 효율성을 더욱 높일 수 있다. 따라서 본 논문에서는 데이터 패브릭 아키텍처 설계와 그에 필요한 도구 및 알고리즘을 분석하고, 이를 기반으로 현대 기업의 데이터 관리 전략에 적용할 수 있는 가능성을 검토한다.

II. 본문

2.1 데이터 패브릭 개요



(그림 1) 데이터 패브릭 아키텍처

그림 1은 데이터 패브릭 아키텍처의 구성 요소와 데이터 흐름을 설명한다. 왼쪽에는 클라우드 스토리지, SaaS 애플리케이션 데이터, 스트리밍 데이터 등 다양한 데이터 소스가 위치하며, 이들 데이터는 ETL(Extract, Transform, Load) 도구를 통해 표준화 및 변환 과정을 거쳐 데이터 패브릭으로 전달된다. 이렇게 통합된 데이터는 데이터 패브릭 스토리지에 저장되어 이후 분석과 처리에 활용된다. 저장된 데이터는 데이터 액세스와 데이터 관리 기능을 통해 다양한 용도로 접근 및 관리되며, 데이터 카탈로그를 통해 효율적인 검색과 접근이 가능하도록 메타데이터가 인덱싱된다. BI(Business Intelligence) 대시보드와 기계학습 분석 단계를 통해 시각화된 인사이트와 예측 모델이 제공되며,

상단의 데이터 거버넌스는 데이터의 일관성, 품질, 보안을 보장하고, 하단의 메타데이터 관리는 데이터 출처 및 처리 방법 등을 기록하여 관리 효율성을 높인다. 이와 같이 데이터 패브릭 아키텍처는 다양한 데이터 소스 간의 통합을 지원하며, 중앙에서 전체 데이터 환경을 관리할 수 있는 구조를 제공한다.

2.2 데이터 패브릭 아키텍처 기반 알고리즘

알고리즘 1은 그림 1을 기반으로 데이터 패브릭 아키텍처를 구현한 알고리즘이다.

(알고리즘 1) 데이터 패브릭 아키텍처 기반 알고리즘

```
def collect_data():
    data_sources = load_data_sources()
    collected_data = []
    for source in data_sources:
        data = extract_data_from_source(source)
        collected_data.append(data)
    return collected_data

def transform_data(collected_data):
    transformed_data = []
    for data in collected_data:
        transformed = perform_etl(data)
        transformed_data.append(transformed)
    return transformed_data

def store_data(transformed_data):
    storage = DataFabricStorage()
    for data in transformed_data:
        storage.store(data)
    return storage

def create_data_marts(storage):
    data_marts = []
    for purpose in business_purposes:
        mart = storage.create_data_mart(purpose)
        data_marts.append(mart)
    return data_marts

def analyze_data(data_marts):
    processed_data = []
    for mart in data_marts:
        filtered_data = filter_data(mart)
        processed_data.append(process_data(filtered_data))
    ml_results = perform_machine_learning(processed_data)
    return ml_results

def manage_and_utilize_data(ml_results):
    update_metadata_catalog(ml_results)
    dashboard = generate_bi_dashboard(ml_results)
    return dashboard

def main():
    collected_data = collect_data()
    transformed_data = transform_data(collected_data)
    storage = store_data(transformed_data)
    data_marts = create_data_marts(storage)
    ml_results = analyze_data(data_marts)
    dashboard = manage_and_utilize_data(ml_results)
    display_dashboard(dashboard)

main()
```

이 알고리즘은 데이터 파이프라인의 전체 과정을 다룬다. 'collect_data' 함수를 사용해 데이터 소스를 로드하고 각 소스에서 데이터를 추출하여 리스트에 저장한다. 이후 'transform_data' 함수는 수집된 데이터를 ETL 과정을 통해 변환한 후, 'store_data' 함수를 호출하여 변환된 데이터를 'DataFabricStorage' 객체에 저장한다. 'create_data_marts' 함수는 비즈니스 목적에 맞춰 데이터 마트를 생성하고, 'analyze_data' 함수는 이 데이터 마트에서 필터링과 처리를 통해 머신 러닝 결과를 도출한다. 마지막으로 'manage_and_utilize_data' 함수는 머신 러닝 결과를 바탕으로 메타데이터 카탈로그를 업데이트하고 비즈니스 인사이트를 제공하는 대시보드를 생성한다. 모든 과정은 'main' 함수에서 실행되며, 최종적으로 생성된 대시보드가 표시된다. 이 알고리즘은 데이터 수집, 변환, 저장, 분석 및 활용의 흐름을 보여준다.

III. 결론

본 논문에서는 데이터 패브릭 아키텍처가 현대 기업의 분산된 데이터를 효율적으로 통합하고 관리하는 강력한 솔루션임을 보여준다. 데이터 패브릭은 ETL 도구를 통해 다양한 소스에서 데이터를 수집하고 표준화하여, 통합된 데이터 환경을 제공한다. 이를 통해 기업은 데이터 사일로 문제를 해결하고 실시간 데이터 접근성을 확보하여 비즈니스 인사이트를 강화할 수 있다. 또한 데이터 마트, 데이터 카탈로그, 데이터 거버넌스를 통해 데이터 품질과 보안을 유지하면서 기계학습 및 BI 대시보드를 활용해 고급 분석과 신속한 인사이트 도출이 가능하다. 이러한 데이터 패브릭 아키텍처는 데이터 중심의 의사 결정을 지원하며, 기업의 비즈니스 가치를 창출하는 데 중요한 역할을 할 것이다.

ACKNOWLEDGMENT

본 논문은 2024년도 광운대학교 교내 학술 연구비 지원에 의해 연구되었습니다.

참 고 문 헌

- [1] X. Li, M. Yang, X. Xia, K. Zhang, and K. Liu, "A Distributed Data Fabric Architecture based on Metadata Knowledge Graph," 2022 5th International Conference on Data Science and Information Technology (DSIT). IEEE, Jul. 22, 2022
- [2] V. Sharma, B. Balusamy, J. J. Thomas, and L. G. Atlas, Eds., "Data Fabric Architectures." De Gruyter, May 04, 2023.
- [3] R. Hai, C. Koutras, C. Quix, and M. Jarke, "Data Lakes: A Survey of Functions and Systems," IEEE Transactions on Knowledge and Data Engineering. Institute of Electrical and Electronics Engineers (IEEE), pp. 1 - 20, 2023

원달러 환율 변동성 예측모형 추정 연구

안창호
서경대학교
choan@skuniv.ac.kr

Estimation study on the prediction model for the won-dollar exchange rate volatility

An Chang Ho
SeoKyeong Univ.

요 약

본 연구에서는 AR-GARCH 모형으로 원달러 환율 변동성 예측모형을 제시하였다. 원달러 환율 자료는 변동성 집중 현상이 존재하고 ARCH 효과가 존재하는 것으로 나타나 이분산모형을 적용하였다. 먼저 AR(1)-GARCH(1,5) 모형을 적용하여 모형을 추정한 후 Ljung-Box 검정으로 모형의 적합성을 확인한 결과 모형은 적절하지 않은 것으로 나타나 유의하지 않은 모수를 제외하고 재추정하고 재추정 결과를 반복하여 얻은 최종모형은 AR(1)-GARCH(1,(1,1)) 모형으로 나타났다. 최종 선택된 모형의 모수 추정치는 모두 유의하고, 추정모형으로부터 얻은 표준화 잔차와 표준화 잔차제곱에 대한 잔차분석 결과 모두 백색잡음 과정을 따르는 적합한 모형으로 나타났다.

I. 서 론

최근 원달러 환율은 긴축기조 장기화 전망과 글로벌 자산 시장 위험회피 경향, 무역수지 연속 적자, 지정학적 리스크 등에 기인하여 당분간 높은 변동성이 지속될 것으로 판단된다. 환율의 변동성은 각 국가의 경제활동과 지역 경제에 영향을 미친다. 환율이 상승하면 수출 자체는 유리하지만 영세한 지역 제조업은 계속 상승하는 유가와 맞물려 비용 부담이 커지게 된다. 원달러 환율 변동성 추정 선행연구는 다음과 같다. Edwards and Rigobon(2009)은 ARCH와 GARCH 모형을 사용하여 달러/칠레 페소 환율 변동성을 추정하였고[1], 김보미·김재희(2013)는 원-달러 환율을 이용해 ARIMA(1,1,0)+IGARCH(1,1) 모형이 실제 환율의 변동성을 잘 반영하는 모델이라고 하였다[2]. Seung Hee Lee, Hee Joon Han (2016)은 GARCH 모형과 변동성의 장기적 반모수 단일지표 변동성 모형을 이용하여 변동성을 분석하였다[3]. Do Kyun Chun (2017)은 GARCH(1,1)모형을 이용하여 원-달러, 원-엔, 원-유로, 원-파운드의 변동성을 추정하여 비교분석 하였고[4], Conrad and Kleen, (2020)은 혼합빈도 GARCH 모형은 장기 기억 프로세스와 유사한 긴 자기상관의 유형에 대해 상당한 설명력을 가진다고 주장하였다[5].

II. 본문

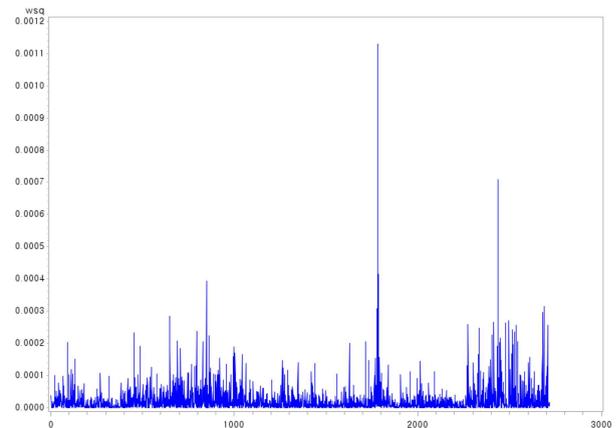
2.1 연구모형

본 연구에서 사용한 이분산모형은 오차가 m 차 자기회귀오차모형을 따르며 동시에 오차의 분산이 GARCH(p,q)모형을 따르는 AR(m)-GARCH(p,q) 모형으로 다음과 같다.

$$\begin{aligned} Z_t &= x_t' \beta + \varepsilon_t \\ \varepsilon_t &= -\phi_1 \varepsilon_{t-1} - \dots - \phi_m \varepsilon_{t-m} + \nu_t \\ \nu_t &= \sigma_t u_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^q \alpha_i \nu_{t-i} + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \end{aligned} \quad (\text{식 1})$$

2.2 원달러 환율 변동성과 ARCH 효과 검증

원달러 환율 변동성을 확인하기 위하여 원자료를 로그변환 하였다. 로그변환 자료를 1차 차분하고 1차 차분 자료의 제곱에 대한 시계열 그림으로 변동성을 확인한 결과 (그림 1)과 같이 변동성 집중(volatility clustering) 현상이 나타나고 있음을 알 수 있다.



(그림 1) 원달러 제곱 시도표

또한, AR(1) 모형을 적합한 후 오차들 사이에 ARCH 효과

의 존재 여부를 Ljung-Box의 Q-검정과 LM(Lagrange Multiplier)-검정으로 검정한 결과는 (표 1)과 같이 잔차의 제곱들이 12시차까지 유의적인 자기상관관계가 존재하고 있음을 알 수 있다.

(표 1) ARCH 효과 검정

Test for ARCH Disturbances Based on OLS Residuals				
Order	Q	Pr > Q	LM	Pr > LM
1	97.8521	<.0001	97.8262	<.0001
2	221.225	<.0001	186.2008	<.0001
3	299.2995	<.0001	216.372	<.0001
4	391.8673	<.0001	249.2486	<.0001
5	408.7251	<.0001	249.5690	<.0001
6	435.7233	<.0001	251.2869	<.0001
7	455.0718	<.0001	251.2869	<.0001
8	472.5965	<.0001	251.2869	<.0001
9	504.5055	<.0001	265.0865	<.0001
10	511.9999	<.0001	265.3306	<.0001
11	518.6988	<.0001	265.4004	<.0001
12	524.1213	<.0001	265.4026	0.0001

2.3 변동성 예측모형

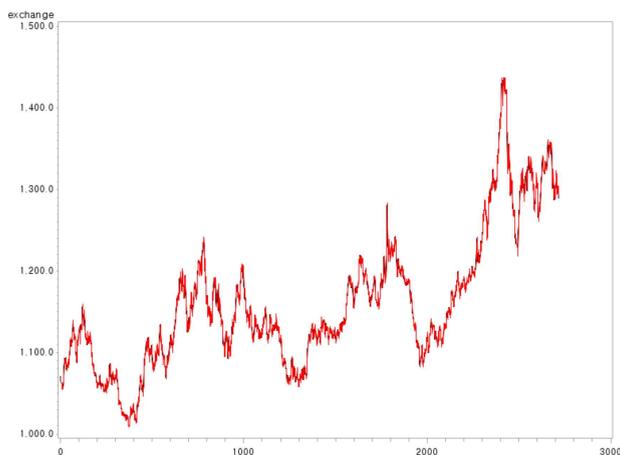
예측모형을 추정하기 위하여 유의하지 않은 모수를 제거하고 재추정 과정을 반복한 결과 AR(1)-GARCH(1,{1,1}) 모형이 최종모형으로 선택되었고, 모수를 추정한 결과 (표 2)와 같이 추정 모수들은 모두 유의한 것으로 나타났다.

(표 2) AR(1)-GARCH(1,{1,1}) 모형 모수추정

AR(1)-GARCH(1,{1,1}) Model Parameter Estimates					
Variable	DF	Estimate	Standard Error	t value	Approx Pr > t
w1	1	0.0834	0.0199	4.20	<.0001
ARCH0	1	2.7567E-7	7.435E-8	3.71	0.0002
ARCH1	1	0.0534	0.005805	9.21	<.0001
GARCH1	1	0.9350	0.007503	124.62	<.0001

2.4 원자료와 예측 시도표

변환된 자료를 원자료로 다시 변환하여 최종모형으로 예측한 시계열 그림은 (그림 2)와 같다.



(그림 2) 원자료와 예측 시도표

본 연구에서는 한국은행 경제통계시스템에서 제공한 2013.1.2 ~2023.12.29까지의 원달러 환율 자료를 사용하여 원달러 환율 변동성 예측모형을 제시하였다. 환율 변동성이 높아지면 경제의 불확실성이 확대되어 투자, 수출, 생산성 등에 부정적인 영향을 미친다. 환율 불확실성이 높아지면 수출입 물가가 상승하고 수출입물량은 감소하며, 인플레이션 불확실성도 높아져 금리를 상승시키고 주가를 하락시키는 등 경제 전반에 부정적 영향을 미친다. 정부와 금융기관 등은 불확실성을 축소하기 위한 정책적인 노력을 기울일 필요가 있으며 또한, 기업들도 환위험 관리의 중요성을 인식하고 결제통화의 다변화, 환위험 관리능력의 배양 등을 통해 환율 불확실성 증대에 따른 부작용을 최소화하려는 노력을 기울여야 한다. 따라서 본 연구에서 제시한 원달러 환율 변동성 예측모형은 최근의 자료를 이용한 실증분석으로써 정부 당국은 물론이고 기업들의 위험관리 근거 자료로 활용될 것으로 판단된다.

참 고 문 헌

- [1] Edwards, S., and Rigobon, R., “Capital controls on inflows, exchange rate volatility and external vulnerability”, *Journal of International Economics*, 78(2), 256-267, 2009.
- [2] 김보미, 김재희., “일별 환율데이터에 대한 시계열 모형 적합 및 비교 분석”, <응용통계연구>, 26(1), 14-27, 2013.
- [3] Seung Hee Lee, Hee Joon Han., Adopting Economic and Financial Variables to Explain Stock Market Volatility in Korea, “*The Korean Journal of Economic Studies*”, Vol. 64(2), pp. 67-95, 2016.
- [4] Do Kyun Chun., Volatility Estimation with Stochastic Volatility Model: Application in Korean Exchange Market, Korea Advanced Institute of Science and Technology, Theses Master, 2017.
- [5] Conrad, C. and O. Kleen., “Two are better than one: volatility forecasting using multiplicative component GARCH MIDAS models”, *Journal of Applied Econometrics*, Vol. 35(1), pp 19-45, 2020.

III. 결론

뉴욕시 택시 데이터를 통한 승객 수요 및 요금 특성 탐색

신지은*, 문석재**
 광운대학교 스마트융합대학원
 {woodlandhill, msj8086}@kw.ac.kr

Exploring Passenger Demand and Fare Trends in New York City Taxi Data

Ji-Eun Shin*, Seok-Jae Moon**
 Graduate School of Smart Convergence, Kwangwoon Univ.

요약

도시화와 교통량 증가로 효율적인 교통 관리 및 배차 시스템이 필수적이다. 택시는 도시 교통에서 중요한 역할을 하며, 택시 요금 예측은 교통 패턴 분석과 승객 수요가 높은 지역 및 시간대 파악에 유용한 정보를 제공한다. 본 연구는 2016년 3월 1일부터 3일까지의 뉴욕시 노란 택시 요금 데이터를 기반으로, 소규모 데이터셋에 머신러닝을 적용하고 지리 데이터를 시각화하여 택시 요금을 예측하는 신경망 모델을 구축하는 것을 목표로 한다. 특히 공항 등 주요 지점의 정액 요금 패턴을 반영하여 예측 모델의 실효성을 높였으며, 이를 통해 도심 교통 흐름 개선과 승객 수요 예측의 정밀도를 높이고자 한다. 또한 본 연구는 예측 모델이 도시 교통 관리에 실질적으로 기여할 가능성을 제시하며, 다양한 도시 데이터에 적용할 수 있도록 모델의 일반화 가능성에도 중점을 두었다.

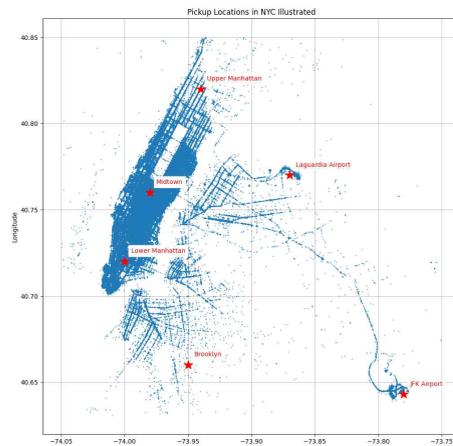
I. 서론

도시화와 교통량 증가로 효율적인 교통 관리 및 배차 시스템이 필수적이다[1]. 특히 택시는 도시 교통에서 중요한 역할을 하며, 택시 요금 예측을 통해 교통 패턴을 분석하고 승객 수요가 높은 지역과 시간대를 파악하는 것은 서비스 품질 향상에 중요한 정보를 제공한다[2]. 이에 따라 대규모 택시 승하차 데이터를 분석하여 요금을 예측하는 것은 도시 교통 문제 해결에 기여할 수 있는 중요한 요소이다. 본 연구는 뉴욕시의 2016년 3월 1일부터 3일까지의 노란 택시 요금 데이터를 활용하여, 소규모 데이터셋에 머신러닝을 적용하고 지리 데이터를 시각화하여 택시 요금을 예측하는 신경망 모델을 구축하는 것을 목표로 한다[3,4]. 특히 공항 등 주요 지점의 정액 요금 패턴을 반영하여 예측 모델의 실효성을 높였으며, 이를 통해 도심 교통 흐름 개선과 승객 수요 예측의 정밀도를 높이고자 한다[5]. 본 연구는 예측 모델이 도시 교통 관리에 기여할 가능성을 제시하며, 다양한 도시 환경에서도 활용할 수 있도록 모델의 일반화 가능성에 초점을 맞추었다.

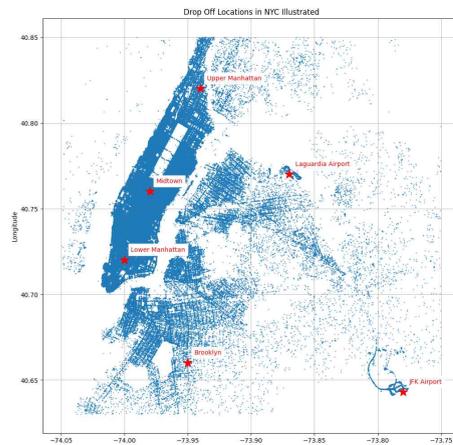
II. 연구 방법

2.1 위치 데이터 시각화

그림 1과 그림 2는 각각 위도와 경도를 점차트의 축으로 설정해 데이터셋에 포함된 승·하차 위치를 시각화하여, 각 점이 특정 위치를 나타낸다. 분석 결과 Upper Manhattan과 Brooklyn 등 거주 지역에서 하차가 빈번하다. 이 시각화는 지역별 승객 수요를 이해하는 중요한 기초 자료를 제공한다.



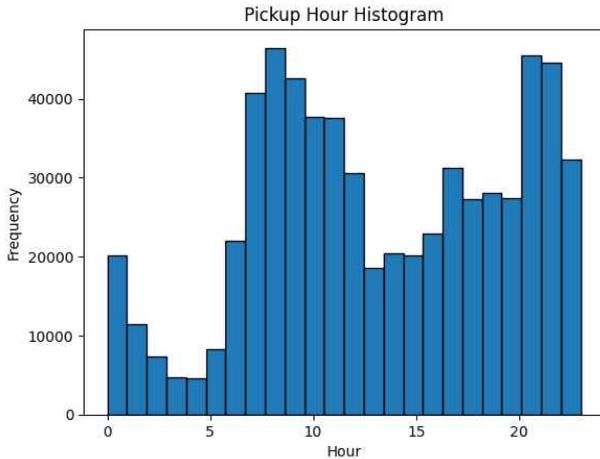
(그림 1) 승차 위치



(그림 2) 하차 위치

2.2 시간별 승차 통계 시각화

그림 3은 시간에 따른 승차 수를 집계하여 히스토그램으로 나타낸다. 분석 결과, 출근 시간대에 승차 수가 증가하는 경향이 있으며, 4시부터 꾸준히 상승하여 8시에 정점을 찍고 이후 점차 감소하는 패턴이 관찰된다. 또한, 21시를 기준으로 퇴근 시간에 급증하는 경향이 나타나, 이러한 패턴은 교통 관리 전략 수립에 유용한 정보를 제공한다.



(그림 3) 시간 별 승차 히스토그램

2.3 데이터 전처리

결측값과 이상치를 제거하는 데이터 전처리 작업은 데이터 분석의 중요한 첫 단계다. 본 연구에서는 총 6만 개의 로우 중 결측값이 있는 로우를 확인하고, 이를 제거하여 데이터의 무결성을 유지한다. 승객 수의 분포를 히스토그램으로 시각화하여 승객 수의 전반적인 패턴을 이해하고, 이상치의 영향을 분석한다. 특히 승객 수가 0인 데이터를 최빈값으로 대체함으로써 데이터의 왜곡을 방지하고, 분석의 신뢰성을 높이는 데 기여한다. 또한, 승차 및 하차 위치의 이상치를 점차트를 통해 시각화하여 가장자리에 위치한 이상치들을 식별하고 제거한다. 이러한 이상치 제거 작업은 데이터셋의 대표성을 높이며, 예측 모델의 학습 과정에서 발생할 수 있는 부정확성을 줄이는 데 중요한 역할을 한다. 이 모든 전처리 과정은 데이터 품질을 향상시켜 예측 모델의 신뢰성을 높이며, 보다 정확하고 신뢰할 수 있는 택시 요금 예측을 가능하게 한다. 따라서 전처리 단계에서의 철저한 작업은 연구의 성과를 좌우하는 중요한 요소로 작용한다.

2.4 새로운 특징 변수 생성

승차 위치 변수를 바탕으로 새로운 특징 변수를 생성한다. 유클리드 거리로 계산한 운행 거리와 요금 간의 관계를 점차트로 시각화하여, 운행 거리가 증가할수록 요금도 증가하는 경향이 있음을 확인한다. 그러나 일부 경우에는 운행 거리와 관계없이 일정 요금이 발생하는 현상이 관찰된다. 공항을 오가는 택시는 고정 요금인 52달러와 통행료를 추가하여 요금을 부과하기 때문에 이러한 현상이 발생하는 것으로 판단된다. 따라서, 뉴욕시 주요 공항과 승차 및 하차 위치 사이의 거리를 새로운 특징 변수로 추가할 필요가 있다.

III. 결론

본 연구에서는 뉴욕시 택시 요금 예측을 위한 효율적인 예측 모델을 제안한다. 위치 정보, 시간대별 데이터, 승객 수 등 다양한 변수를 포함하여 요금 예측의 정확도를 높이고자 하였으며, 이 모델은 향후 도시 교통의 효율적 관리를 지원하고 승객 수요가 높은 시간대와 지역에 대한 정보를 제공함으로써 서비스 품질을 향상시킬 수 있다. 특히, 공항과 같은 주요 지점의 정액 요금 패턴을 반영하여 예측 모델의 실효성을 높였다. 그러나 본 연구는 특정 기간의 데이터만을 활용하였고, 일시적인 교통 상황이나 날씨와 같은 외부 요인을 반영하지 못한 한계가 있다. 이러한 제한 사항은 향후 연구에서 개선할 수 있으며, 날씨 변화와 대규모 행사 등 외부 요인을 추가 고려하여 택시 수요 및 요금 예측의 정밀도를 더욱 높일 수 있을 것이다. 또한, 라쏘와 릿지 회귀 모델과 같은 다양한 회귀 기법을 활용하여 변수의 중요도를 평가하고 모델의 해석 가능성을 개선하는 방향도 중요한 연구 과제로 고려될 수 있다. 이러한 기법들은 데이터의 복잡성을 효과적으로 다룰 수 있도록 도와주며, 예측 정확성을 높이는 데 기여할 것으로 기대된다. 마지막으로, 본 연구에서 개발한 예측 모델을 실시간 데이터에 적용하여 요금 예측 및 배차 관리 시스템으로 발전시킬 가능성을 제시한다. 이를 통해 실제 택시 운행에서 요금 예측 결과를 활용하고, 궁극적으로 도시 교통 서비스의 효율성을 더욱 향상시키는 데 기여할 수 있을 것이다.

ACKNOWLEDGMENT

본 논문은 2024년도 광운대학교 교내 학술 연구비 지원에 의해 연구되었습니다.

참 고 문 헌

- [1] S. Glaister and D. J. Graham, "Proper Pricing for Transport Infrastructure and the Case of Urban Road Congestion," *Urban Studies*, vol. 43, no. 8. SAGE Publications, pp. 1395 - 1418, Jul. 2006
- [2] C. Zhang, F. Zhu, X. Wang, L. Sun, H. Tang, and Y. Lv, "Taxi Demand Prediction Using Parallel Multi-Task Learning Model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2. Institute of Electrical and Electronics Engineers (IEEE), pp. 794 - 803, Feb. 2022
- [3] F. Xie and D. M. Levinson, *Evolving Transportation Networks*. Springer New York, 2011.
- [4] M. (Muki) Haklay, Ed., "Interacting with Geospatial Technologies." Wiley, Mar. 19, 2010
- [5] S.-C. Lee and H.-K. Kwon, "Computational Algorithms and Modeling for the Traffic Flow," *Indian Journal of Science and Technology*, vol. 9, no. 44. Indian Society for Education and Environment, Nov. 22, 2016

체험형 힐링프로그램이 대학생 마음건강에 미치는 영향

이혜정¹, 한덕관²

¹원광대학교 교육혁신원

²원광대학교 도덕교육원

{redrose, kar0505}@wku.ac.kr

The Impact of Experiential Healing Programs on University Students' Mental Health

¹Hye-Jung Lee, ²Deok-Gwan Han

¹Office of Educational Innovation, Wonkwang University

²Moral Education Center, Wonkwang University

요 약

현재 우리나라의 고용 불안정성 상승, 무한 경쟁 사회로의 급격한 변화 및 코로나 19로 인한 장기적인 사회적 거리두기와 비대면 등의 요인들이 결합되어 감염병에 대한 불안감과 정신적·심리적 스트레스가 심화되면서 최근 사회적 고립률이 높은 청년층의 우울, 불안, 대인관계의 어려움, 사회적 소통의 부족 등이 급격히 증가되고 있다. 이에 따라 자해나 자살, 부정적 사고, 사회부적응에 따른 청년층의 심리적 불안, 무력감, 우울증 관리는 중요한 사회적인 이슈로 인식되고 있다.

본 논문에서는 대학생의 불안, 스트레스 및 우울감 등의 해소를 위한 체험형 힐링프로그램이 마음건강에 미치는 영향을 알아보고자 한다. 마음건강 증진을 위한 체험형 힐링프로그램은 다양한 소재와 방법을 다루는 경험을 통해 흥미와 집중, 자신감, 유희성 등을 체험하고, 이를 통한 정서적 안정과 소통, 창의적 표현을 향상한다. 또한 체험활동으로 집단에서 서로 소통하는 방법을 습득하고 자기조절 능력 및 상호작용을 통해 새로운 환경에 대한 적응력을 높일 수 있다. 이를 확인하기 위하여 프로그램에 참여한 대학생을 대상으로 사전-사후 조사를 실시하였다.

량을 강화하고 스트레스를 예방한다. 또한 집단에서 서로 소통하는 방법을 습득하고 자기조절 능력 및 상호작용을 통해 새로운 환경에 대한 적응력을 높일 수 있다. 이를 확인하기 위하여 체험형 힐링프로그램에 참여한 대학생을 대상으로 사전-사후 조사를 실시하고 분석하였다.

I. 서론

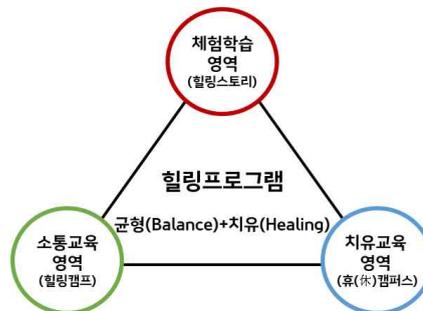
현재 우리나라의 고용 불안정성 상승, 무한 경쟁 사회로의 급격한 변화 및 코로나 19로 인한 장기적인 사회적 거리두기와 비대면 등의 요인들이 결합되어 감염병에 대한 불안감과 정신적·심리적 스트레스가 심화되면서 최근 사회적 고립률이 높은 청년층의 우울, 불안, 대인관계의 어려움, 사회적 소통의 부족 등이 급격히 증가되고 있다[1][2]. 이에 따라 자해나 자살, 부정적 사고, 사회부적응에 따른 청년층의 심리적 불안, 무력감, 우울증 관리는 중요한 사회적인 이슈로 인식되고 있다. 청년층에 포함되는 대학생은 사회적·심리적으로 주어진 새로운 발달 과업을 달성해야 하는 시기이며 경쟁과 책임감을 극복 및 적응하며 성인으로 발전하는 시기이다. 또한 미래의 진로나 직업에 대한 불확실함과 역량을 발전시키려는 심리적 압박에 의한 스트레스가 심화되고 있다[3][4].

본 논문에서는 대학생의 불안, 스트레스 및 우울감 등을 해소하기 위한 체험형 힐링프로그램이 마음건강 증진에 미치는 영향을 알아보고자 한다. 마음건강 증진을 위한 체험형 힐링프로그램은 다양한 소재와 방법을 다루는 경험을 통해 흥미와 집중, 자신감, 유희성 등을 체험한다. 이를 통해 스트레스와 지친 마음을 치유하고 정서적·심리적 안정을 도모하며, 정서회복을 위한 체험활동으로 자기관리 역

II. 본론

2.1 체험형 힐링프로그램의 구성

마음의 균형(Balance)과 치유(Healing)의 자연친화적 체험형 힐링프로그램을 정서안정 및 자존감 향상, 스트레스 해소와 대인관계 개선을 위해 (그림1)와 같이 체험학습 영역, 소통교육영역, 치유교육영역으로 분류하였다.



(그림 1) 체험형 힐링프로그램의 구성

체험학습영역은 공예테라피 과정으로 감성역량을 위한 힐링 체험 교육을 통해 집중력을 높이고 정서안정을 통한 자신감을 획득함으로써 긍정적 사고를 형성하고 성취감 및

몰입경험을 위한 신체활동 중심의 자기개발 프로그램으로 구성된다.

소통교육영역은 마음치유 과정으로 심신안정을 위한 힐링 체험 교육을 통해 감정을 표현하고 긴장을 억제함으로써 심리적 이완, 상호교류, 긍정적 사고를 위한 협동·공감·배려활동 등 정서안정 프로그램으로 구성된다.

치유교육영역은 건강생활 실천 과정으로 건강과 환경 실천 힐링 체험 교육을 통해 환경보호 및 건강한 (식)생활의 습관을 개선함으로써 불균형 해소, 친환경 실천, 감정조정을 위한 건강지킴 프로그램으로 구성된다.

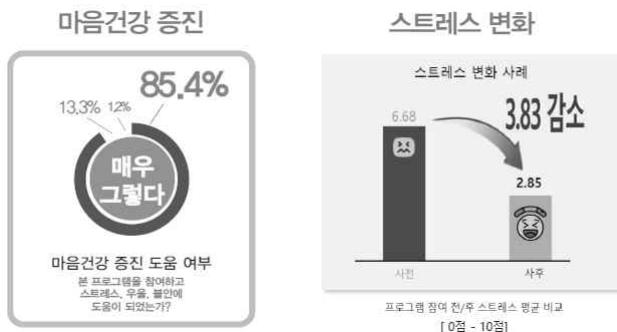
2.2. 체험형 힐링프로그램의 성과

체험형 힐링프로그램의 연도별, 학년별, 성별 참여 현황은 (그림2)와 같다. 2023년에 12개 체험형 힐링프로그램을 28회 운영하였고, 총 563명이 참여하였다. 특히 학년별 참여 현황에서 4학년 고학년의 참여율이 가장 높게 나타난 점은 취업과 학업에 대한 스트레스 극복욕구가 강하게 표출된 것으로 분석된다.



(그림 2) 연도별, 학년별, 성별 참여 현황

체험형 힐링프로그램을 참여한 후 마음건강 증진 도움 여부와 스트레스 변화에 대한 결과는 (그림3)과 같다. 체험형 힐링프로그램에 참여하고 스트레스, 우울, 불안에 도움이 되었다는 답변이 85.4%로 나타났으며, 스트레스 수치는 3.83 감소한 것으로 분석되었다.



(그림 3) 마음건강 증진 도움 여부 및 스트레스 변화

III. 결론 및 향후 연구

본 논문에서는 대학생의 불안, 스트레스 및 우울감 등의 해소를 위한 체험형 힐링프로그램이 마음건강 증진에 미치는 영향에 대해 알아보았다. 체험형 힐링프로그램을 통해 흥미와 집중, 경험, 유희성 등을 체험함으로써 정서적 안정과 소통, 창의적 표현이 향상됨을 알 수 있었다. 또한, 집단에서 서로 소통하는 방법을 습득하고 자기조절 능력 및 상호작용을 통해 공감능력의 향상과 새로운 환경에 대한 적응력이 높아짐을 파악할 수 있었다. 이를 확인하기 위하여 프로그램에 참여한 대학생들을 대상으로 프로그램 참여 사전-사후 조사를 실시한 결과, 다음과 같이 대학생 마음건강에 영향을 미친 것으로 분석되었다. 자연친화적 재료를 활용한 체험활동 및 소근육 신체활동을 통해 정서가 안정되었으며, 함께 활동하고 개인과업 수행으로 자신감·성취감을 회복하여 자아존중감이 향상되었으며, 적극적인 태도로 프로그램에 참여하여 자계개발에 동기부여가 강화되어 더불어 학교생활 적응력이 향상된 것으로 나타났다. 이러한 결과를 바탕으로 향후 연구 방향은 다음과 같다. 본 연구는 학교 현장에서 DIY활동 중심의 체험을 통해 학생들이 직접 참여하는 실제적 활동 중심의 측면에서 살펴본 기초 연구 자료로 향후 추가 연구를 수행할 계획이다. 또한 대학생들의 일상생활 적응력 향상과 마음건강을 위한 다양한 체험형 프로그램 개발과 다각적인 효과성 검증을 수행하고자 한다.

참고 문헌

- [1] Yoo, M. S., Shin, D. H., & Lee, M. J. "A Study on the Social Isolation of Young People and their Support Plan", National Youth Policy Institute, pp. 261-267, 2021
- [2] Baek, S. K. "Analysis of Education Policy Conditions: with a focus on social change", Korean Educational Development Institute, 11(15), pp. 5-7, 2014
- [3] 차남희, 김희수, "대학생 생활스트레스 조절을 위한 비대면 마음챙김 집단상담 프로그램 개발 및 효과 검증", 상담학연구:사례 및 실제, Vol.5, No.2, pp. 41-61, 2020
- [4] 조원희, 정유미, "대학생의 미술치료 기반 스트레스 관리 프로그램 개발 및 적용에 대한 연구", 기초조형학연구, 25권 3호 pp. 361-374, 2024

류마티스 관절염 조기진단을 위한 관절손상 평가를 위한 인공지능 모델 개발

정중혁¹, 이창훈¹, 이명수¹, 임동욱², 노시형², 문희경³

¹원광대학교병원 류마티스내과

²원광대학교 근감소증도발솔루션 센터

³원광대학교 교육혁신원

taylorchung@hanmail.net, {lch110, ckhlms, dwl316, nosij123, ybnjcw}@wku.ac.kr

Development of an AI Model for Joint Damage Assessment in Early Diagnosis of Rheumatoid Arthritis

Chong Hyuk Chung¹, Chang Hoon Lee¹, Myung soo Lee¹, Dong-Wook Lim², SiHyeong Noh², Hee-Kyung Moon³

Division of Rheumatology, Department of Internal Medicine, Wonkwang University Hospital

²Sarcopenia Total Solution Center, Wonkwang University

³Institute for Educational Innovation, Wonkwang University

요 약

류마티스 관절염은 자가면역 질환으로, 조기 진단과 신속한 치료가 관절 손상을 최소화하고 환자의 삶의 질을 향상시키는 것이 중요하다. 그러나 현재의 관절손상 평가는 방사선 이미지를 기반으로 한 의료진의 주관적 평가에 크게 의존하고 있어 평가의 일관성과 조기 진단의 정확성에 한계가 있다.

본 연구는 X-Ray 영상을 이용해 관절손상을 자동 평가하는 인공지능(AI) 모델을 제안한다. Erosion 및 Joint Space Narrowing(JSN) 스코어를 학습 데이터로 활용하여 CNN(Convolutional Neural Network) 기반 AI 모델을 구축하였으며, Grad-CAM 기법을 통해 관절 손상 부위를 시각적으로 확인할 수 있도록 하였다. 실험 결과, 제안된 AI 모델은 기존의 평가 방식보다 일관되고 신뢰성 높은 결과를 제공하며, 조기 진단을 지원할 수 있는 가능성을 보였다. 이 연구는 류마티스 관절염 진단에서 객관적인 평가 도구를 제공하여 조기 진단과 적절한 치료 방안을 제시하는 데 기여할 것으로 기대된다.

I. 서 론

류마티스 관절염은 자가면역 질환으로 관절과 주변 조직에 염증을 일으켜 통증, 부종, 운동 제한 등을 초래하는 만성적 질환이다[1,2]. 특히 조기 진단과 치료가 중요하며, 조기 치료가 이루어질수록 관절 손상을 최소화할 수 있다. 현행 진단 방식은 방사선 이미지를 통한 관절손상 평가(Erosion score 및 JSN score)로 의료진의 경험에 따라 평가가 주관적으로 이루어진다[3, 4].

현재 관절손상 평가는 주로 방사선 이미지를 기반으로 하며, 환자의 관절 상태에 대한 평가가 의료진의 주관적 해석에 크게 의존한다. 이로 인해 평가의 일관성과 객관성이 떨어지고, 조기 진단의 정확도 역시 낮아질 수 있다. 또한, Erosion 및 JSN 스코어는 관절 손상이 이미 진행된 후에야 변화를 감지할 수 있어 초기 단계에서의 미세 변화를 포착하기 어렵다는 한계가 있다[5].

본 연구는 인공지능 모델을 활용하여 관절손상을 자동으로 평가하는 시스템을 제안함으로써 진단 일관성을 높이고, 조기 진단을 통해 임상현장에 적용되길 기대한다.

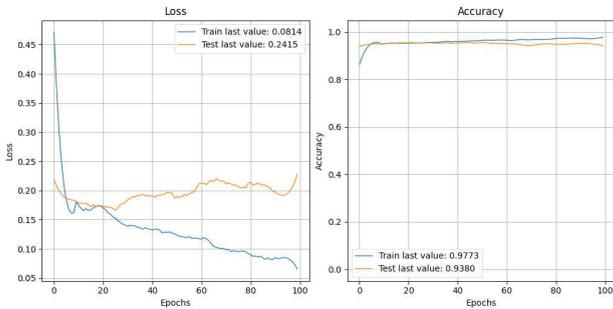
II. 관절손상 평가를 위한 인공지능 모델

본 연구에서는 기존 X-Ray 영상 데이터를 활용하여 관절손상 평가 스코어(Erosion 및 JSN)를 자동화하는 AI 모델을 개발하였다. 이를 위해 다수의 X-Ray 영상을 기반으로 CNN(Convolutional Neural Network) 모델을 학습시키고, 관절손상을 자동으로 스크리닝하여 결과를 정량화하는 시스템을 설계하였다. 관절손상 자동 진단 모델을 개발하기 위해서 우리는 X-Ray 수부 영상에서 관절별 레이블링과 스코어 값을 토대로 학습과 테스트를 진행하였다. 표1은 학습을 위한 데이터 분류와 개수를 보이고 있다.

(표 1) 데이터 구성과 분류

클래스	학습데이터	테스트 데이터
0	3,628	908
1	87	22
2	41	11
3	130	33
total	3,886	1,944

참 고 문 헌



[그림 1] 관절손상 학습과 테스트 결과

해당 모델의 학습과 테스트한 결과 97%의 정확도를 보였다. 하지만 검증을 위해 Grad-CAM으로 시각화하여 결과를 확인하니 스코어 0인 경우는 정확하지만 데이터가 부족한 1, 2, 3에 대한 오답을 보이는 경우가 많았다.

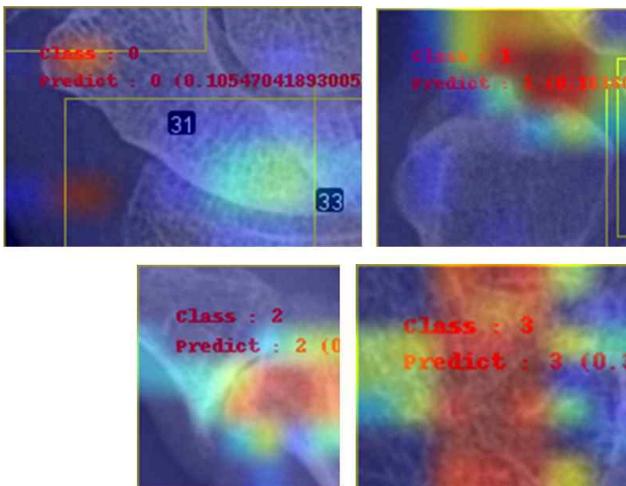
[1] Tan YK, Li H, Allen JC Jr, Thumboo J. Joint damage in rheumatoid arthritis: Lessons learned from an extended 36-joint ultrasonography and 28-joint Disease Activity Score assessments. *Int J Rheum Dis.* 2021 Jan;24(1):106-111. doi: 10.1111/1756-185X.14013. Epub 2020 Nov 7. PMID: 33159378.

[2] Tan YK, Li H, Allen JC Jr, Thumboo J. Joint damage in rheumatoid arthritis: Lessons learned from an extended 36-joint ultrasonography and 28-joint Disease Activity Score assessments. *Int J Rheum Dis.* 2021 Jan;24(1):106-111. doi: 10.1111/1756-185X.14013. Epub 2020 Nov 7. PMID: 33159378.

[3] van der Heide A, Remme CA, Hofman DM, Jacobs JW, Bijlsma JW. Prediction of progression of radiologic damage in newly diagnosed rheumatoid arthritis. *Arthritis Rheum.* 1995 Oct;38(10):1466-74. doi: 10.1002/art.1780381013. PMID: 7575696.

[4] Pincus T, Schmukler J, Block JA, Goodson N, Yazici Y. Should quantitative assessment of rheumatoid arthritis include measures of joint damage and patient distress, in addition to measures of apparent inflammatory activity? *ACR Open Rheumatol.* 2023 Jan;5(1):49-50. doi: 10.1002/acr2.11514. Epub 2022 Dec 20. PMID: 36540953; PMCID: PMC9837390.

[5] Bird, A., Oakden-Rayner, L., McMaster, C. et al. Artificial intelligence and the future of radiographic scoring in rheumatoid arthritis: a viewpoint. *Arthritis Res Ther* 24, 268 (2022). <https://doi.org/10.1186/s13075-022-02972-x>



[그림 2] 관절손상 스코어 2에 대한 검출 결과

클래스별 학습데이터에 부족으로 인하여 데이터의 균형을 위해 증강하고 세부적인 데이터의 분류 작업을 수행하였다. 손가락과 손목부분을 재분류하여 학습을 진행하였지만 82%의 정확도를 보였다.

III. 결론

우리가 제안한 AI 모델은 학습 데이터에서 Erosion 및 JSN 스코어를 기준으로 관절손상 정도를 평가하였다. 특히, Grad-CAM을 통해 손상 영역의 특징을 시각적으로 보였다. 실험 결과 제안된 AI 모델은 기존 방식 대비 일관된 평가 결과를 제공하며, 조기 진단 및 치료 계획 수립에 도움을 줄 수 있음을 확인하였다. 하지만 학습데이터가 부족한 스코어에 대한 예측은 아직 부족한 상황임을 알 수 있었고, 학습데이터를 확보하는 방안과 영상 전처리 기술을 적용하여 제안한 모델의 성능 향상연구를 수행할 계획이다.

의미적 분석과 연구자 관심 분야를 이용한 자동 추천 시스템

홍광희
여주대학교 소프트웨어융합과
khhong@yit.ac.kr

An Automatic Recommendation System using Semantic Analysis and Researcher Interesting Areas

Hong Kwang Hee
Yeojoo Institute Technology

요 약

본 논문에서는 시맨틱 분석과 사용자 관심 분야를 이용하여 APRPRS를 제안하였으며, 의미적 분석을 통하여 향상된 알고리즘을 제시하고 구현하였다. 사용자 프로파일의 갱신은 주제별로 수집된 연구논문들이 선택되면 도메인, 키워드, 토픽의 빈도수를 계산하고 빈도수를 증가시켜 각각의 발생률을 재계산하여 사용자 프로파일에 재반영한다. 시맨틱 키워드는 동일 도메인 내에 속한 연구논문 중 유사 논문들에서 사용된 의미적으로 유사한 키워드들을 추출하여 사용자 프로파일에 추가함으로써 이루어진다. 본 논문에서는 의미적 분석과 연구자 관심 분야에서 키워드의 의미적 확장방법과 시스템의 성능을 평가하고 검증하기 위해 시스템에서 추천하는 각 논문에 대한 만족도와 정확도를 측정하였다. 의미적 키워드 분석과 연구자 관심분야가 갱신시 약 12%의 성능향상이 있었으며, 실험 결과 제안 시스템은 높은 수준의 만족도와 정확도를 나타내었다.

이미 사용자의 관심분야의 관련 키워드들로 작성된 프로파일을 참조하여 질의어를 확장한다. 가중치 있는 키워드, 개념 및 연관규칙의 집합으로 사용자 프로파일을 표현하며 이는 정보검색에서 매우 활성화된 연구분야로 기계학습이나 정보검색에 기반한 다양한 구축기술을 사용하여 정보소스로부터 구축된다.[3] 프로파일에 기반하여 질의어를 확장하는 방법은 사용자가 질의를 입력하면 이미 사용자 관심분야의 관련 키워드들로 작성된 프로파일을 참조하여 질의어를 확장하며, 사용자 피드백에 의하여 확장하는 방법은 사용자가 검색된 결과를 본 후 질문을 수정하여 반복적인 탐색을 한다.[4] 연구자가 이전에 읽은 논문에 대한 평가를 기반으로 연구자의 선호 프로파일을 작성한다.[5] 단어의 의미적 관계를 적용한 단어의 의미적 확장을 적용하여 연구자 프로파일과 추천할 후보 논문 간 유사도를 측정하여 사용자에게 추천한다.

2.2 자동 추천 시스템

추천시스템은 정보필터링 기술을 활용하여 사용자가 원하는 내용을 찾아주며, 추출규칙을 통하여 기술되며, 가공을 위해서 미리 정의된 템플릿에 추출된 정보를 저장한다. 연구자가 이전에 읽은 논문에 대한 평가를 기반으로 연구자의 선호 프로파일을 작성한다. 단어의 시맨틱 관계를 적용한 단어의 의미 확장을 구축하며, 의미 확장을 적용한 연구자 프로파일과 추천할 후보 논문 간 유사도를 측정하여 사용자에게 추천한다.

I. 서 론

웹 기술의 발달로 많은 연구자들은 관심분야 주제의 연구 논문들을 다양한 검색엔진을 이용하여 수집하며, 메타데이터를 추가한 웹 문서의 내용검색모델이 연구되고 있다. 질의어 대신 프로파일 정보를 이용하여 유사한 연구논문들을 비교, 분석하여 우수한 연구논문을 작성하고자 한다. 프로파일 정보를 이용하여 사용자가 선호하는 카탈로그 문서를 제공하는 서비스가 있으며, 추천시스템에서의 최근 관심은 추천의 정확성을 향상시키는데 있다. 의미적 키워드 확장을 적용한 사용자 프로파일 기반 연구논문 추천시스템인 APRPRS를 제안한다. 각 저널에서 수집되는 분야별 신규 연구논문들을 사용자 선호정보 프로파일에 기반하여 필터링하고 이를 제공할 수 있게 개발하였다.[1]

사용자 프로파일은 추천된 연구논문들을 사용자가 클릭하였을 때 해당 연구논문의 키워드를 저장하고 있으며, 워드넷을 활용하여 의미적으로 유사한 주제어를 동일 도메인 내에서 확장하였다.[2] 또한 시맨틱 분석을 통한 사용자의 선호정보를 적절히 반영하여 사용자에게 분야별 신규 연구 논문을 제공할 수 있음을 확인하였으며, 추천되는 연구 논문에 대한 높은 수준의 정확도와 만족도를 향상시킬 수 있었다.

II. 관련연구

2.1 시맨틱 분석을 통한 키워드 확장

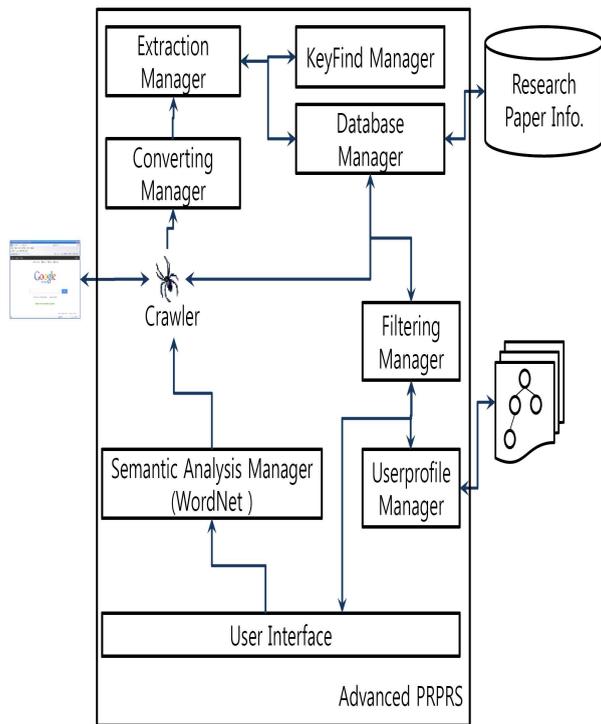
질의어 확장에 대한 연구는 개인, 그룹의 프로파일을 이용하는 기법과 사용자 피드백에 의한 방법이 있다. 프로파일을 작성하여 확장하는 방법은 사용자가 질의를 입력하면

III. 시스템 구성

3.1 APRPRS 시스템 구조

그림 1은 본 논문에서 제안하는 APRPRS 시스템의 구조이다. APRPRS는 (1)Extraction Manager, (2) KeyFind Manager, (3) Database Manager, (4) Filtering Manager, (5) Userprofile Manager, (6) User Interface (7)Semantic Analysis Manager, (8) Crawler, (9)Converting Manager 로 구성되며 각각의 기능은 다음과 같다.

(1) Extraction Manager는 제목, 저자, 초록, 키워드, 본문, 참고문헌 등 부분별 추출과 식별을 하며, 해당 논문에 대한 색인정보를 자동으로 생성한다. (2) KeyFind Manager는 Userprofile에 적용될 키워드를 추출하며, 키워드를 추출하지 못하는 경우 논문제목에서 유추한 키워드 및 유사논문에서 사용된 키워드들을 활용한다.



(그림 1) APRPRS 시스템 구조

(3) Database Manager는 수집된 논문들에 대한 색인정보의 관리기능과 수집된 각 연구논문의 정보를 저장하고 있는 데이터베이스에 접근하는 방법을 제공 및 데이터베이스에 데이터를 추가, 삭제, 변경 등의 기능을 담당한다. 데이터베이스에는 Crawler에 의해 수집된 신규 연구논문들에 대한 정보와 논문제목, 저자, 초록, 키워드 등의 정보를 저장하고 있다. (4) Filtering Manager는 사용자가 선택한 연구주제에 대해 Userprofile을 활용하고 필터링을 수행하며 저장된 연구논문들 중 사용자에게 가장 적절한 논문을 제공한다. 또한 정제된 논문 리스트를 UserInterface를 통하여 사용자에게 제공한다. (5) Userprofile Manager는 사용자의 연구논문 주제에 대한 기호정보를 저장하고 있는 사용자 프로파일을 관리하며, 사용자 프로파일은 검색결과로 제시된 각 연구논문을 사용자가 클릭할 때 갱신된다.

(6) User Interface는 사용자로부터 주제를 입력받아 여구문분석과 워드넷을 활용한 의미분석을 하며, 이 과정을 통하여 유사주제의 탐색 및 후보 키워드 추출에 활용한다. (7) Semantic Analysis Manager(WordNet)는 추출된 유사주제들을 Google scholar의 질의어로 활용함으로써 연구논문들을 수집한다. (8) Crawler는 수집된 연구논문들은 Converting Manager로 전달되며, 각 수집된 PDF 문서는 "PDF To TXT" 모듈을 통하여 Text로 변환하여 시스템에서 처리할 수 있도록 한다.

3.2 키워드 추출 알고리즘

질의어 확장은 사용자가 제시한 질의어와 관련된 단어를 추가하여 검색함으로써 연관성이 높은 문서들을 검색한다.. 최근 질의어 확장에 대한 연구는 개인, 그룹의 프로파일을 이용하는 방법과 사용자 피드백을 이용하는 방법이 있다. 프로파일을 작성하여 확장하는 방법은 사용자가 질의를 입력하면 이미 사용자의 관심분야와 관련된 키워드들로 작성된 프로파일을 참조하여 질의어를 확장한다. 사용자 피드백에 의하여 확장하는 방법은 사용자가 검색된 결과를 본 후 질문을 수정하여 반복적인 탐색을 한다.

$$Sim(d_i, d_j) = \frac{d_i \cdot d_j}{|d_i| |d_j|} = \frac{1}{W_{d_i} W_{d_j}} \sum_t w_{d_i} w_{d_j} \quad (1)$$

$$= \frac{1}{W_{d_i} W_{d_j} \sum_{t \in d_j} tf_{d_i}} \cdot \left(\log \frac{N}{df_t} \right)^2$$

$$KeySim(D_i, D_j) = \frac{\sum_m \max(WordMatch(d_i k_m, d_j k_n))}{DN_i} \quad (2)$$

* the weight of a term T_j in document $D_i : w_{ij}$

$$idf_j = \log(N/df_j)$$

$$w_{ij} = tf_{ij} \cdot idf_j$$

연구자가 이전에 읽은 논문에 대한 평가를 기반으로 연구자의 선호 프로파일을 작성한다. 단어의 의미적 관계를 적용한 단어의 의미 확장을 구축한다. 구축한 의미 확장을 적용한 연구자 프로파일과 추천할 후보 논문 간 유사도를 측정하여 사용자에게 논문을 추천한다.

해당 논문에 Keywords 섹션이 있으면 관련 연구에서 사전 정의된 규칙과 추출 알고리즘을 통하여 추출하여 사용하며, Keywords 섹션이 없으면 논문제목에 사용된 단어들 중 적절한 단어를 선택하여 키워드로 사용한다. 논문에 사용된 각 단어들 중 본문 내 TF(Term Frequency)를 통하여 2개의 단어를 키워드로 사용한다.

구글 스칼라(<http://scholar.google.com/>)로부터 연구논문들의 리스트를 얻고, 각 논문의 URL을 통하여 로컬에 저장하며, DB Manager를 통하여 중복저장을 방지한다.

키워드 확장은 해당 논문의 키워드 리스트를 인자로 받아 시작하며, 해당 논문과 동일 도메인 내 논문들을 대상으로 코사인 유사도를 측정하여 일정한 임계값(0.7) 이상의 유사도를 나타내는 논문들을 추출한다(Line 2). 유사 논문들로부터 키워드를 추출하여(Line 3-4) 인자로 전달된 각 키워드와의 유사성을 체크하여(Line 5-6) 임계값 이상이면 확장 키워드 리스트에 추가한다(Line 7-8). 인자로 전달된 키워드를 각 유사논문에 추가함으로써 해당 논문 뿐만 아니라 각 유사논문의 키워드 확장도 할 수 있도록 하였다(Line 10).

Algorithm 1 Keyword Expansion

```

1: function KeywordExpansion(keywords[])
2:   SimPapers[] ← getSimilarPapers(); // extract
   similar research papers with the given research
   paper
3:   while  $i < \text{length of keywords}$  do // repeat
   until  $i$  is greater than or equal to length of
   keywords
4:     while  $j < \text{length of SimPapers}$  do // repeat
   until  $j$  is greater than or equal to length of
   SimPapers
5:       sim_keywords[] ← SimPapers[j].getKeywords();
   // extract keywords from a similar research
   paper
6:       while  $k < \text{length of sim\_keywords}$  do //
   repeat until  $k$  is greater than or equal to length
   of sim_keywords
7:         word_match_value ← getWMV
   (keywords[i], sim_keywords[k]); // calculate
   keyword similarity
8:         if word_match_value > 0.68 then
9:           ExKeywords[m] ← sim_keywords[j]; //
   add sim_keywords[k] to ExKeywords at m
10:          m++; // increase index m
11:          SimPapers[j].putKeyword(keywords[i]);
   // add keywords[i] to SimPapers[j]
12:         end if
13:         k++; // increase index k
14:       end while
15:     j++; // increase index j
16:   end while
17:   i++; // increase index i
18: return ExKeywords;
19: end function

```

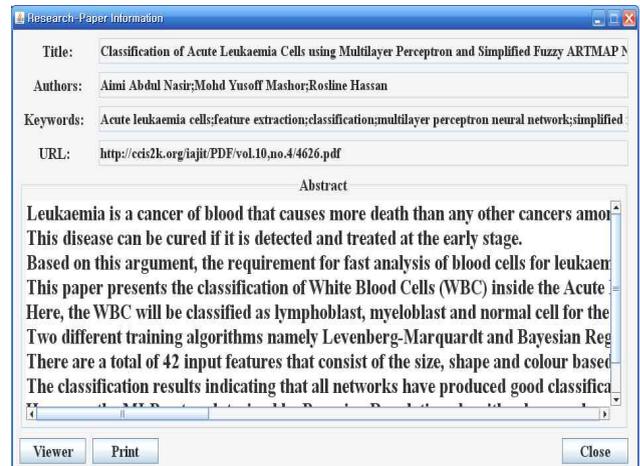
해당 링크가 연구논문의 PDF 파일에 대한 URL이거나, 링크를 찾아갔을 때 해당 연구논문의 PDF 파일에 대한 URL을 포함하는 경우이면 해당 PDF 파일을 다운로드 받고, 시스템에서 다룰 수 있는 TEXT 형태로 변환한다. 변환된 TEXT에서 제목, 저자명, 초록, 키워드 등의 필요한 정보를 추출(Extract) 한다[Line 6-8]. 반면 링크를 가지고 있고, 링크를 찾아갔을 때, 해당 페이지에서 연구논문에 대한 정보를 얻을 수 있으면 해당 HTML 소스를 분석하여 필요한 정보를 추출한다[Line 9-10]. 링크를 갖지 않는 검색

결과 결과는 해당 제목을 이용하여 구글에서 해당 연구논문에 대한 PDF 파일을 검색, 다운로드하며, 변환한 후 추출한다[Line 15-19]. 사용자 인터페이스를 통하여 각 주제별 수집된 연구논문을 클릭할 때 사용자 프로파일은 갱신된다. 논문에 사용된 단어들의 TF(Term Frequency)를 통하여 단어 2개를 주요 키워드로 사용하며, 단어는 본문 내에서 발생하는 단어에 따라 복합명사를 포함하여 제목에서 선택한다. 연구자가 작성한 연구논문에서 추출한 키워드와 기존 분야별 구축된 키워드를 함께 Userprofile에 등록하며, 등록된 키워드를 구글 스콜라에서 검색하여 필터링을 수행한 후, 연구자에게 신규 연구논문을 추천한다.

IV. 실험

4.1 환경

본 논문에서 제안한 시맨틱 분석과 사용자 관심분야를 이용한 시스템의 성능을 평가하고 검증하기 위하여 실험을 수행하였다. 연구자가 이전에 읽은 논문에 대한 평가를 기반으로 연구자의 선호 프로파일을 작성한다. 제안한 방법의 성능을 확인하기 위하여 이전 연구와 비교를 반복적으로 수행하였으며, 만족도와 정확도를 측정하였다.



(그림 2) 연구논문에 대한 사용자 인터페이스

각 사용자는 연구 관심분야와 관련된 도메인을 선택하고, 찾고자 하는 주제를 생성하였다. 그림 2는 연구논문에 대한 사용자 인터페이스이며, 연구논문의 상세정보를 나타내었다. 각 연구논문을 더블클릭하면 팝업 윈도우창을 통하여 해당 연구논문에 대한 상세정보를 제공한다.

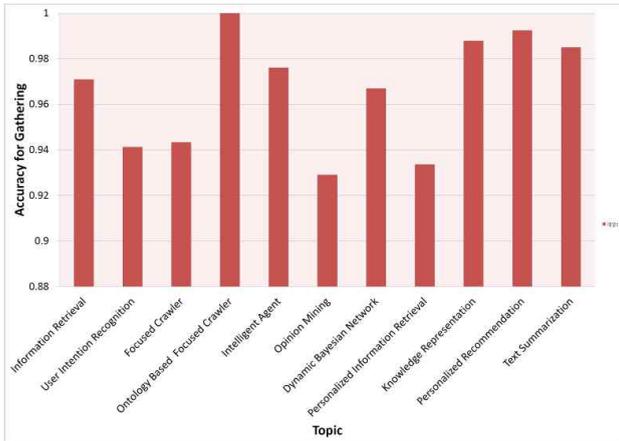
4.2 평가

만족도는 각 분야별 전체 추천된 논문 중 사용자가 만족하는 논문의 비율을 의미하며, 정확도는 각 분야별 전체 수집된 논문 중 사용자가 만족하는 추천된 논문의 수와 사용자가 만족하지 않는 추천하지 않은 논문의 수의 비율을 의미한다. 만족도를 측정하기 위해 한 번에 추천되는 연구논문의 수를 최대 100개로 제한했으며, 이미 추천된 연구논문은 비교하여 재 추천되지 않는다.

$$SAT = \frac{\#Satisfied\ Research\ Paper}{\#Recommended\ Research\ Paper} \quad (3)$$

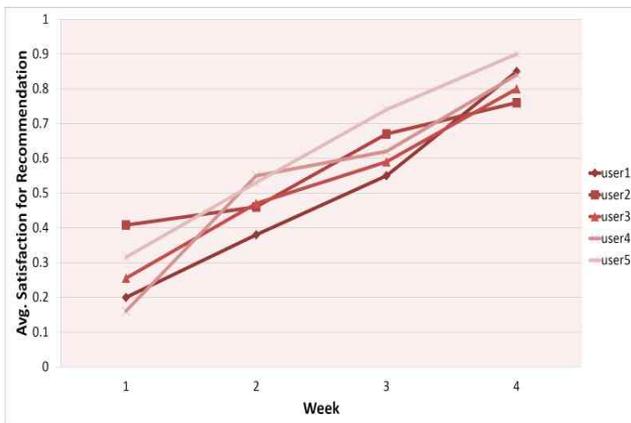
$$ACC = \frac{\#Recommended\ Sat + \#Not\ Recommended\ UnSat}{Total\ Saved\ Research\ Paper\ for\ a\ given\ Topic} \quad (4)$$

그림 2는 주제별 수집 정확도를 측정한 것이며 전체평균은 0.965이다. Ontology based Focused Crawler topic은 높은 정확도를 나타냈고, 추천 정확도를 측정하기 위해 각 사용자에게 추천되는 수를 제한하지 않았으며, 로컬에 저장된 전체 연구논문이 추천대상이다.



(그림 2) 주제별 수집 정확도

그림 3은 각 사용자에게 대한 평균 추천 만족도를 측정한 것이며, 전체 사용자에게 대해 시간이 지남에 따라 만족도가 빠르게 향상되는 것을 알 수 있었으며, 1주차의 평균 만족도에 비해 마지막 주의 평균만족도(0.83)는 약 3배가 향상되었다.



(그림 3) 각 사용자에게 대한 평균 추천 만족도

연구자 자신이 작성한 연구논문을 기반으로 하므로 연구 분야에 적합할 확률이 매우 높음을 나타낸다. 만족도는 비교적 높은 수준에서 시작하며, 높은 수준의 정확도로 인하여 전체적으로 만족도가 빠르게 상승하였다. 그림 3은 실험기간 동안 각 사용자의 측정된 만족도를 나타내었으며,

지속적으로 만족도와 정확도는 향상되었다. 값이 급격히 상승하는 구간은 신규 연구논문의 출판된 비율이 많은 경우이며, 완만한 변화를 보이는 구간은 신규 출판된 저널논문의 비율이 상대적으로 적은 경우이다. 시맨틱 분석과 사용자 관심분야를 이용하여 시스템 성능을 평가하고 검증하기 위하여 실험을 수행하였으며 만족도와 정확도는 각각 0.83와 0.96으로 매우 높은 수준을 나타냈으며, 키워드 확장을 하여 약 12%의 성능향상을 나타내었다.

V. 결론

본 논문에서는 시맨틱 분석과 사용자 관심 분야를 이용하여 APRPRS를 제안하였으며, 의미적 분석을 통하여 향상된 알고리즘을 제시하였고 구현하였다. 제안한 시스템이 높은 수준의 정확도로 연구자의 연구분야를 인지하고 높은 수준의 만족도로 신규 연구논문들을 추천할 수 있음을 확인하였다. 기존 저널분석에서 해당 저널에 나타나지 않은 키워드 또는 유사한 의미의 키워드인 경우 정확도를 증가시키는 연구가 필요하다. 향후 연구로는 단점을 개선하고 현재의 관심분야를 추천하는 지능기반 추천방법을 개발할 것이며, 의미유사도를 사용하여 더 높은 정확도와 만족도를 나타내고자 한다.

참고 문헌

- [1] Kwanghee Hong, Hocheol Jeon, Changho Jeon, "UserProfile-based personalized research paper recommendation system", 2012 8th International Conference On Computing and Networking Technology (ICCNT 2012), vol.8, No.1, pp.134-138, 2012.
- [2] Kwanghee Hong and Changho Jeon, "Automatic Research Area Recognition System for Researcher through Analyzing Previous Research Papers", Journal of Convergence Information Technology (JCIT), AICIT, vol.8, No.16, pp.117-127, 2013.
- [3] Asabere .N.Y., "Towards a Perspective of Hybrid Approaches and Methodologies in Recommender Systems", International Journal of Computer Science and Telecommunications, Vol.3, No.11, pp.23-32, 2012.
- [4] Cristiano Nascimento, Alberto H. F. Laender, Marcos André Gonçalves, Altigran S. da Silva, "A Source Independent Framework for Research Paper Recommendation", JCDL'11, 2011.
- [5] Asabere N.Y., "Towards a Perspective of HybridApproaches and Methodologies in Recommender Systems," International Journal of Computer Science and Telecommunications, Vol.3, No.11, pp.23-32, 2012.

