

웨어러블 라이프로그 기반 상황 인식과 응용 시나리오*

이충연¹○, 이상우¹, 곽동현², 장병탁¹¹서울대학교 컴퓨터공학부, ²서울대학교 뇌과학협동과정

{cylee, slee, dhkwak, btzhang}@bi.snu.ac.kr

Wearable Lifelog-based Context Awareness and its Application Scenario

Chung-Yeon Lee¹○, Sang-Woo Lee¹, Dong-Hyun Kwak², Byoung-Tak Zhang¹¹Department of Computer Science and Engineering, Seoul National University,²Interdisciplinary Programs in Neuroscience, Seoul National University

요 약

라이프로그는 유비쿼터스 환경에서 능동적인 에이전트 서비스 제공을 위한 개인의 특성 명세 데이터로 활용하기에 적합하다. 최근 스마트폰의 보급으로 라이프로그 관련 연구가 활발히 이루어지기 시작했으며, 사용자가 실제로 접하는 다양한 정보를 보다 자연스럽게 기록하기 위해 웨어러블 디바이스를 사용하려는 시도들도 있다. 본 논문에서는 이처럼 스마트폰과 웨어러블 디바이스를 통해 획득 가능한 개인의 라이프로그로부터 유의미한 핵심 정보를 추출하고, 이를 이용하여 장소, 행동, 사람, 사용자 의도 및 감정의 관계와 시간상에서의 패턴을 학습하여 상황을 인식하고 나아가 미래의 정보를 예측하는 연구와 응용 시나리오들을 소개한다. 또한 영상과 음향 데이터로부터 추출된 핵심 정보들이 가지는 상황 설명력을 검증하기 위해 간단한 상황 분류 실험 결과를 제시한다.

1. 서 론

라이프로그(Lifelog)는 개인의 일상생활에 대한 디지털 기록을 의미한다. 이러한 기록은 유비쿼터스(Ubiquitous) 컴퓨팅 환경에서 각종 지능형 계측 장비를 통해 수집되며, 이 과정에서 저장되는 데이터는 일인칭 기준 시점(Egocentric)에서 기록되는 영상, 음성, 신체 반응과 움직임, GPS를 통한 위치정보, 디바이스 사용 로그, 스케줄 내역, 이메일, 메신저 사용, 방문한 사이트 기록 등으로 그 범위가 매우 넓고 다양한 특성을 가진다 [1].

라이프로그가 기록된 정보를 검색하고 인출해내는 디지털 다이어리 형태의 단순한 서비스만을 제공할 수 있는 것은 아니다. 다가올 사물인터넷 환경에서 보다 능동적인 맞춤형 서비스 제공을 위한 개인의 특성 명세 데이터로 쓰일 수 있으며, 특정한 상황에서 개인의 의도 및 취향을 파악하는데 매우 중요한 정보로 활용될 수 있다. 단, 수시로 수집되는 순수 로그 데이터들은 신호 자체만으로는 그 의미를 명확하게 파악하기가 어렵고, 서로 다른 특성을 가지는 이질적(Heterogeneous)인 양상을 보이며, 시간에 따라 변화하는 동적인 특성을 가진다. 따라서 각각의 로그 데이터 특성에 맞는 핵심 정보를 추출해내는 전처리 과정이 필요하고, 또한 상황에 따라 실시간으로 변화하는 환경과 사용자 행동 간의 유의미한 연관성을 파악하는 방법이 개발되어야 한다 [2].

최근 다양한 센서들이 내장된 고성능 스마트폰이 여가에 보급되고 데이터 저장 공간의 가격도 점차 낮아지면서 라이프로그 수집이 용이해지고, 스마트폰을 이용한 라이프로그 관련 연구가 국내외에서 다양하게 수행되고 있다 [3, 4]. 하지만 스마트폰을 통해 기록 가능한 정보는 실제 사용자의 일상 정보를 충분히 기록하기에는 아직 부족하며, 특히 영상과 음성 정보는 보통 주머니에 들어가 있는 스마트폰으로 기록하는 것이 자연스럽지 않다.

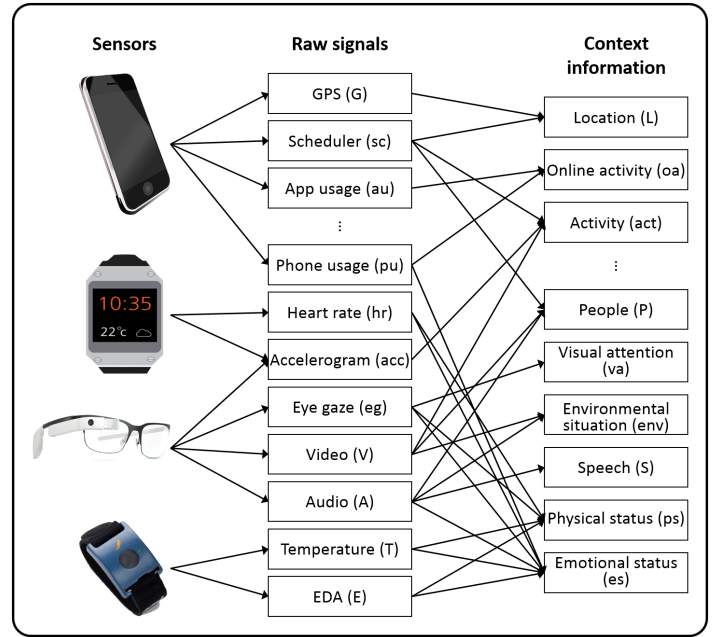
실제로 사용자가 접하는 다양한 정보를 보다 자연스럽게 기록하기 위해 웨어러블 디바이스(Wearable device)를 사용하는 방법이 있다 [5]. 최근 차세대 스마트 기기로 주목 받고 있는 웨어러블 디바이스는 우리가 일상적으로 몸에 걸치고 다닐 수 있는 형태로 설계된 정보 기기들을 통칭한다. 대표적으로 Google이 개발한 스마트 안경 형태의 Google Glass, 스마트 시계인 삼성의 Galaxy Gear, 클립이나 밴드 형태로 몸에 착용 가능한 운동량 측정 기기인 Fitbit과 Misfit Shine 등이 있다.

본 논문에서는 이처럼 다양한 웨어러블 디바이스를 통해 획득 가능한 개인의 라이프로그로부터 유의미한 핵심 정보를 추출하고, 이를 이용하여 장소, 행동, 사람, 사용자 의도 및 감정의 관계와 그 시간상에서의 패턴을 학습하여 상황을 인식하고 나아가 미래의 정보를 예측하는 연구와 응용 시나리오를 소개하고자 한다. 2장에서는 웨어러블 라이프로그의 획득 및 핵심 정보 추출에 대하여 자세히 설명한다. 3장에서는 이 핵심 정보들이 가지는 상황 설명력을 검증하기 위해 간단한 상황 분류 실험 결과를 제시한다. 4장에서는 웨어러블 라이프로그에서 중점적으로 연구되어야 하는 주제와 시나리오를 제시한다.

* 본 논문은 2014년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구이며(NRF-2010-0017734-Videome), 삼성전자와 정부(미래창조과학부 및 정보통신기술진흥센터)의 정보통신·방송 연구개발사업(10035348-mLife, 14-824-09-014-Machine Learning Center)의 지원을 일부 받았다.

2. 웨어러블 라이프로그 획득 및 핵심 정보 추출

개인의 일상을 반영하는 라이프로그 획득을 위해 스마트폰과 웨어러블 디바이스들로부터 그림 1과 같이 센서 정보와 이를 가공하여 얻을 수 있는 상황 정보를 추출한다. 예를 들어 스마트폰의 GPS 센서로부터 위치 좌표를 구하고 해당 좌표의 위치 정보를 온라인 지도 데이터로부터 추출한다. 이 위치 정보는 해당 시간을 포함하는 일정에 대한 장소 정보가 스마트폰 사용자의 스케줄러에 기입되어 있는 경우 교정이 가능하다. 이밖에도 스마트폰 사용 정보 및 통신 기록 등을 획득함으로써 사용자의 온·오프라인 활동 및 만나는 사람에 대한 정보를 추출한다. 스마트워치로부터는 사용자의 심박수, 가속도, 체온과 피부전도도를 측정 가능하며, 이를 통해 신체 반응과 간접적인 감정 상태 정보를 추출한다. 또한 스마트글래스를 이용하여 일인칭 시점의 동영상과 나아가 시선추적 정보를 기록하며, 각각의 데이터로부터 주변 환경 정보와 사용자의 행동 정보 및 주의 집중 정보 등을 추출한다. 본고에서는 이렇게 획득한 데이터 중 영상 및 음성에 대한 전처리 과정과 상황 인식 결과를 제시한다.



2.1 영상 데이터 특징값 추출

스마트글래스를 이용하여 녹화한 일인칭시점 비디오로부터 1초에 한 장씩의 이미지를 추출한 후, 각 이미지를 대표하는 20개의 설명 태그(Descriptive tags)와 설명에 대한 확률값을 구한다. 표 1은 실험을 통해 획득한 설명 태그들 중 일부를 상황별로 열거해놓은 것이다.

이 과정은 기존에 이미 학습된 모델을 이용하여 계산되며, 현재까지 가장 높은 이미지 분류 성능을 보인 Convolutional Deep Learning [6]을 이용하여 구할 수 있다. 이렇게 추출된 설명 태그가 기존에 컴퓨터 비전 연구에서 사용하는 Bag-of-Words 모델과 비교하여 가지는 장점은 각 성분이 의미적 설명이 가능하고, 보다 적은 차원의 특징값으로 사용 가능하며, 개념망 형태의 모델을 구성하는데 유리하다는 것이다. 이어질 실험 결과에서 설명 태그의 상황 설명력을 검증한다.

2.2 음향 데이터 특징값 추출

위와 동일한 비디오 데이터에서 음향 데이터를 분리한 후, 사람의 소리 인지 주파수를 반영하는 것으로 알려져 있는 MFCC 계수를 추출한다. 이때 44100 Hz의 샘플링비트수로 녹음된 음향 데이터를 영상 데이터와 마찬가지로 1초 간격으로 분할하고, 2048개 샘플로 구성된 프레임을 1024개 간격으로 이동시키면서 이전 프레임과 50% 중첩되는 방식으로 분할한다. 이렇게 총 42개 프레임으로 나뉜 데이터를 FFT (Fast Fourier Transform)를 이용하여 주파수 영역으로 변환한 후, Mel-scale의 필터 뱅크(Filter Bank)에 통과시켜 파워스펙트럼을 구한다. Mel-scale 필터 뱅크를 통과한 데이터에 로그 함수와 DCT (Discrete Cosine Transform)를 수행하여 12개의 주파수 특성과 하나의 프레임 로그 에너지 특성으로 구성된 총 13차의 MFCC 계수를 추출한다. 최종적으로 음향 데이터로부터 1초마다 42개의 13차원 특징 벡터들이 구해지게 된다.

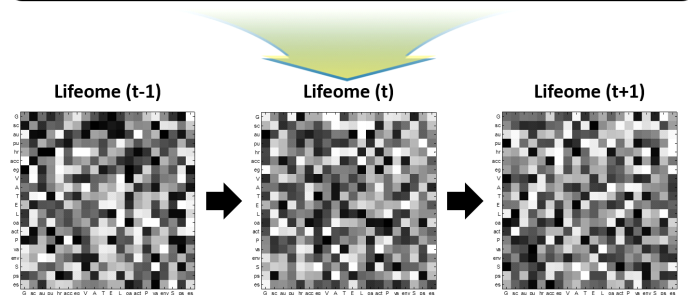


그림 1. 웨어러블 라이프로그와 상황 정보

표 1. 사용자 활동별 상위 20개 영상 데이터 설명 태그

take a class	have fun	rest at home	in the lab	in the library	moving
pachyderm	night	room	baguette	room	light
room	concert	indoors	room	shadow	transportation
indoors	light	window	electronics	indoors	road
business	festival	furniture	business	black&white	night
education	performance	business	technology	light	vehicle
school	bridge	technology	indoors	nobody	street
furniture	stage	electronics	pachyderm	floor	motion
airport	music	chair	screen	furniture	airport
desk	motion	baguette	furniture	storm	subway
classroom	energy	museum	computer	motion	police
baguette	reflection	floor	window	airport	shadow
class	asleep	interior design	telephone	food court	business
transportation	evening	shadow	portable	dark	competition
chair	travel	light	graveyard	stairs	automobile
museum	band	hotel	desk	texture	traffic
sitting	city	luxury	laptop	rain	indoors
technology	celebration	screen	display	water	dark
floor	party	contemporary	education	winter	speed
window	north america	pachyderm	empty	business	car
interior design	singer	lamp	interior design	weather	recreation

3. 영상 및 음향 데이터를 이용한 상황 인식

영상과 음향 데이터로부터 추출한 특징값들의 상황 설명력을 확인하기 위해 사용자가 직접 입력한 시간대별 활동 내역을 분류해보는 실험을 하였다. 실험을 위해 스마트폰글래스를 착용한 사용자가 3일 간 일인칭 비디오 데이터를 수집하였으며, 이 기간 동안 사용자의 활동 내역 및 데이터 수집 시간은 표 2에서 확인할 수 있다.

표 2. 사용자 활동 내역 및 데이터 수집 시간 (단위: 초)

수업	여가	휴식	연구	도서관	이동
6,232	1,064	1,576	8,430	983	13,629

비디오에서 추출된 영상들로부터 설명 태그와 확률값들을 추출하기 위해 2013년도 ImageNet Classification Competition의 분류 과제에서 가장 좋은 성능을 보인 Convolutional Neural Networks 기반의 Clarifai API [7]를 사용하였다. 이후 추출된 전체 설명 태그들 중 평균 확률값이 가장 높은 200개의 설명 태그들만 사용하여 각 영상 데이터의 특징값을 구하였다. 표 1은 상황별로 가장 높은 확률 평균값을 보이는 20개 태그를 열거한 것이다.

영상 데이터의 설명 태그와 음향 데이터의 MFCC 계수들을 특징값들로 사용하여 사용자의 활동 내역을 분류한 결과는 그림 2와 같다. Naive Bayes (N/B), Support Vector Machine (SVM), k-Nearest Neighbor (KNN; k=1), 그리고 Decision tree 계열의 J48과 Random Forest (R/F)를 각각 사용하여 분류한 결과 Naive Bayes를 제외한 나머지 분류기에서 모두 높은 정확도를 확인할 수 있었다. 특히 영상 데이터 설명 태그는 90%가 넘는 정확도를 보임으로써 높은 상황 설명력을 가지고 있음이 검증됐다.

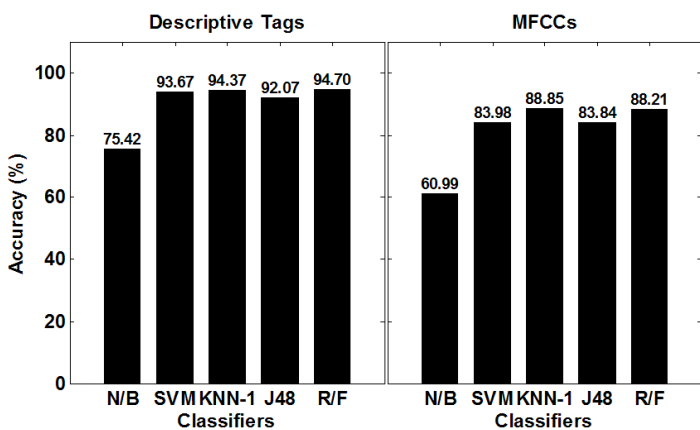


그림 2. 사용자 활동 내역 결과 (10-Cross Validated)

4. 웨어러블 라이프로그 연구 주제 및 응용 시나리오

웨어러블 라이프로그를 통해 표 3과 같이 크게 3가지 문제를 다뤄볼 수 있다. 각각의 문제는 동적인 환경에서 실시간 멀티모달 센서 데이터에 기반을 둔 빠르고 유연한 Life-long 학습 및 추론 모델을 연구하기에 적합하다.

또한 제시된 응용 시나리오들은 산업적 응용 관점에서

웨어러블 디바이스가 단순히 스마트폰의 일부 기능을 대체하는 수준에서 그치지 않고, 사용자에게 특화된 경험을 제공하는 서비스로 포지셔닝하기 위한 킬러앱이 될 수 있을 것으로 기대된다.

표 3. 웨어러블 라이프로그 연구 주제 및 응용 시나리오

문제	연구 주제 및 응용 시나리오
상황인식	<ul style="list-style-type: none"> • 상황의 변화 및 예외적(abnormal) 상황을 감지 • 학교에 가는 버스 탑승, 친구를 만나 대화 시작, 수업 시작 등의 상황을 실시간으로 인식하고, 사용자의 주의가 필요한 예외적 상황을 감지하여 알림
일상요약	<ul style="list-style-type: none"> • Temporal clustering/segmentation • Micro/Macro 스케일의 상황 및 일상 요약, 기억 증강
상황예측	<ul style="list-style-type: none"> • 재생성 학습 기반 상황 시뮬레이션 및 예측 추론 기법 • 이동 경로 및 활동, 만날 가능성이 있는 사람, 나누게 될 대화 주제 등을 예측

4. 논의 및 결론

웨어러블 라이프로그는 특정 시간대에서 개인이 얻게 되는 대다수의 정보를 포함하며, 이들은 인간의 기억과 학습에 사용되는 데이터를 대표하는 것으로 볼 수 있다. 따라서 이론적으로는 이러한 데이터 집합을 이용하여 사람과 유사하게 학습 가능한 지능형 에이전트를 개발 가능하며, 이를 위해 지속적으로 입력되는 서로 다른 센서 데이터간의 시공간적 연관 관계에 기반을 둔 개념망 형태의 모델(그림 1)을 학습하는 방법에 대한 연구가 이루어져야 할 것이다. 본 논문에서는 일인칭 시점의 영상과 음향 데이터를 이용한 상황 인식 실험을 통해 이러한 가능성을 확인하였고, 중점적으로 연구되어야 하는 주제와 응용 시나리오를 제시하였다.

참고 문헌

- [1] G. Bell and J. Gemmell, "A digital life," *Scientific American*, pp. 58-65, 2007.
- [2] 장병탁, "차세대 기계학습 기술," *정보과학회지*, 제 25권, 제 3호, pp. 96-107, 2007.
- [3] 서호, 조성배, "모바일 라이프로그 네트워크에서의 의미관계기반 검색방법," *정보과학회논문지: 소프트웨어 및 응용*, 제 41권, 제 1호, pp. 80-87, 2014.
- [4] R. Rawassizadeh, M. Tomitsch, K. Wac and A. M. Tjoa, "UbiqLog: a generic mobile phone-based life-log framework," *Personal and Ubiquitous Computing*, Vol. 17, No. 4, pp. 621-637, 2013.
- [5] M. Stikic, D. Larlus, S. Ebert and B. Schiele, "Weakly supervised recognition of daily life activities with wearable sensors," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 12, pp. 2521-2537, 2011.
- [6] M. D. Zeiler, *Hierarchical convolutional deep learning in computer vision*, Ph.D. Thesis, New York University: NY, USA, 2014.
- [7] Clarifai, <http://www.clarifai.com/>