

이종의 라이프로그 정보분석을 위한 웰니스 인포메틱스엔진 설계

손창식*, 최락현*, 이상호*, 윤상훈*, 강원석*, 이동하*

*대구경북과학기술원 웰니스융합연구센터
e-mail:changsikson@dgist.ac.kr

Design of wellness informatics engine for analyzing heterogeneous lifelog information

Chang-Sik Son*, Rock-Hyun Choi*, Sang Ho Lee*, Sang Hun Yun*,
Won-Seok Kang*, Dong-Ha Lee*

*Wellness Convergence Research Center, DGIST

요 약

최근 다양한 웨어러블 기기의 등장으로 인해, 이종의 라이프로그 정보로부터 개인별 맞춤형 건강관리 서비스를 제공할 수 있는 분석 솔루션의 중요성이 대두되고 있다. 본 연구에서는 이기종의 웨어러블 기기를 통해 수집된 라이프로그 정보를 근간으로 개인의 건강상태를 모니터링하고 분석할 수 있는 웰니스 인포메틱스엔진에 대한 프레임워크를 소개한다. 또한 개발된 인포메틱스엔진의 효과성은 실증서비스 기간 동안에 수집된 다양한 라이프로그 정보를 활용하여, 2가지 벤치마크 방법들과의 예측능력을 비교함으로써 그 효과성을 제시한다.

1. 서론

라이프로그 관리 기술 (lifelog management technology)은 유비쿼터스 컴퓨팅 환경에서 실현 가능한 개인화된 서비스 기술 중 하나로, 개인이 일상생활에서 경험하는 모든 정보를 수집하고 이를 향후 활용하기 편리하도록 분류, 기록해두었다가 필요한 경우 관련 정보를 쉽게 찾아 볼 수 있는 기술이다[1]. 라이프로그는 단순히 개인의 일상을 스스로 기록하는 것에 그치지 않고 수집된 방대한 양의 정보를 분석하여 개인 맞춤형 서비스를 제공하는데 활용되고 있다. 대표적인 예로 운동거리, 운동시간, 칼로리 소모량을 측정할 수 있는 나이키의 퓨얼밴드 (fuel band)와 음식 먹는 속도를 조절해 주는 씨넛의 해피포크 (HAPifork)를 들 수 있다. 최근 다양한 웨어러블 기기 (wearable device)로부터 수집되는 데이터의 다양성과 복잡성 때문에 라이프로그 관리는 중요한 이슈가 되고 있으며, 특히 사용자에게 맞춤형 건강관리를 위한 양질의 서비스 제공을 위해서 이종의 정보 (heterogeneous information)를 포함하는 라이프로그를 분석할 수 있는 플랫폼의 중요성 또한 대두되고 있다.

본 연구에서는 다양한 웨어러블 기기를 통해 수집된 이종의 라이프로그 정보를 기반으로 개인의 건강상태를 분석할 수 있는 인포메틱스엔진의 프레임워크를 소개하고, 실제 서비스된 데이터를 중심으로 엔진의 성능을 평가하고자 한다.

2. 웰니스 라이프로그 정보

그림 1은 웨어러블 기기로부터 수집된 라이프로그에 대한 개인웰니스레코드(personal wellness record, PWR)를 나타낸다. 그림에서 ‘Vibra Image Based Stress’와 ‘PPG/RSP Based Stress’는 생리학 스트레스 정도 (physiological stress degree)를 측정하기 위한 라이프로그를 나타내고, ‘Activity’, ‘Dietary’, 및 ‘Electrocardiogram(ECG)’은 신체적 웰니스 정도를 측정하기 위한 라이프로그를, ‘Wellness Survey’는 사용자의 주관적인 웰니스 상태 (subjective wellness status)를 측정하기 위해서 설문조사 기반으로 수집된 라이프로그 정보를 나타낸다.

Vibra Image Based Physiological Stress											
종류 ID	이름	시작 시간	종료 시간	해당 시간	종료 시간	Stress	Stress	Stress	Stress	Stress	Stress
1	Stress	12:00	12:30	12:00	12:30	10	15	20	25	30	35
PPG/RSP Based Physiological Stress											
종류 ID	이름	시작 시간	종료 시간	해당 시간	종료 시간	HR	HR	HR	HR	HR	HR
2	Stress	12:00	12:30	12:00	12:30	70	75	80	85	90	95
Activity Using Fitbit Flex Tracker											
종류 ID	이름	시작 시간	종료 시간	시간	Age	Weight	Height	HeartRate	Step	Distance	Calory
3	Activity	12:00	12:30	30	25	60	160	1500	5000	1000	500
Dietary											
종류 ID	이름	시작 시간	종료 시간	시간	Age	Weight	Height	BMI	Food Category	Fat	Salt
4	Dietary	12:00	12:30	30	25	60	160	24.0	Carbohydrate	10g	5g
Electrocardiogram Using Liberal Wireless ECG, HeartCall U Heart											
종류 ID	이름	시작 시간	종료 시간	해당 시간	종료 시간	HeartRate	HeartRate	HeartRate	HeartRate	HeartRate	HeartRate
5	ECG	12:00	12:30	12:00	12:30	70	75	80	85	90	95
Wellness Survey											
종류 ID	이름	시작 시간	종료 시간	시간	Q1	Q2	Q3	Q4	Q5	Q6	Q7
6	Survey	12:00	12:30	30	1	2	3	4	5	6	7

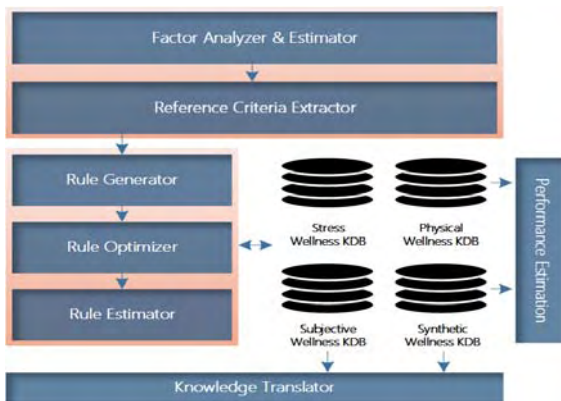
(그림 1) 라이프로그 정보 기반 PWR 명세

특히 신체적 웰니스 라이프로그 정보들 중에서 ‘Activity’와 ‘ECG’와 관련된 라이프로그 정보는 Fitbit사의 Fitbit Flex와 (주)유하트의 1채널 무선 홀터심전계인 HeatCall-B (Sampling frequency, 300Hz) 제품을 활용하였고, 통합 앱 (wellness4u)을 활용하여 수집하였다[2].

3. 웰니스 인포메틱스엔진의 구조

웰니스 인포메틱스엔진은 6종류의 이종의 웰니스 라이프로그 정보를 근간으로 개인의 웰니스 상태, 즉 이벤트(event) 웰니스와 데일리(daily) 웰니스 서비스를 모니터링하고 추천할 수 있도록 설계되어있다. 이벤트 웰니스 서비스는 ‘Vibra Image Based Stress’, ‘PPG/RSP Based Stress’, ‘ECG’와 같이 사용자가 측정 후 생리학적 스트레스 상태 혹은 모니터링 결과를 즉시 확인할 수 있는 서비스를 의미하고, 데일리 웰니스 서비스는 ‘Activity’, ‘Dietary’, ‘Wellness Survey’와 같이 하루 동안 측정된 누적 결과들의 경향을 근거로 추정된 웰니스 상태를 나타낸다.

그림 2는 개발된 웰니스 인포메틱스엔진의 개념적 프레임워크의 구조를 나타내며, 4가지 주요 핵심 컴포넌트(‘요인분석 및 평가부’, ‘규칙생성 및 최적화부’, ‘지식변환부’, ‘성능평가부’)로 구성된다. ‘요인분석 및 평가부’에서는 이종의 라이프로그 정보를 근간으로 이벤트 혹은 데일리 웰니스 상태를 추론하는데 있어서 가장 중요한 요인들(즉 라이프로그들) 간에 연관요인(associative factor)들을 추출할 수 있는 기능을 제공한다.



(그림 2) 인포메틱스엔진의 프레임워크 구조

‘규칙생성 및 최적화부’에서는 선별된 라이프로그 정보들을 활용하여 규칙베이스(rule-base)를 자동으로 생성할 수 있는 기능과 누적된 라이프로그 정보들로부터 각 규칙의 확률적 근거(probabilistic evidence)를 추정(‘Rule Estimator’)할 수 있도록 설계하였다. 또한 ‘지식변환부’는 ‘규칙생성 및 최적화부’를 통해 자동 생성된 규칙베이스로부터 언어적인 형태로 해석 가능한 지식으로 표현할 수 있는 기능을 제공하며, ‘성능평가부’에서는 지식베이스의 예측능력을 평가하기 위한 하나의 도구로서 정확도(accuracy), ROC(receiver operating characteristic) 등을 분석할 수 있는 기능을 제공한다. 개발된 인포메틱스엔진의 모든 컴포넌트는 JDK 1.8(java development kit) 64-bit 환경에서 개발되었다.

4. 실험결과

2015년 3월 2일부터 시행된 웰니스휴먼케어 1차 실증 시범서비스(wellness human-care pilot test), 즉 대구지역 15개 기관의 사회복지사 110명을 대상으로 약 60일 동안 수집된 라이프로그 데이터(수집기간, 2015년 3월 2일 - 4월 30일)를 활용하였다. 표 1은 데일리 웰니스 서비스를 위해서 하루 단위를 간격으로 수집된 데이터 현황을 보여준다.

<표 1> 수집된 데이터 현황 (단위, 건/하루)

Device	Vibra	PPG/RSP	Activity	Dietary	ECG	Survey
Subject	2,120	2,256	2,448	948	319	98

인포메틱스엔진의 객관적인 성능평가를 위해서, 통계적 10-fold 교차검증[3](훈련과 실험 샘플의 비율, 7대 3) 동안에 2가지 벤치마크 방법들(표준 k-NN[4], 퍼지 k-NN[5])과의 예측 성능을 비교하였다.

<표 2> 바이브라 이미지 기반 생리학적 스트레스 성능평가

Vibra	인포메틱스 엔진	표준 k-NN (k=1)	퍼지 k-NN (k=1)
평균 정확도	83.29±0.90	83.33±0.65	83.38±0.67
웰니스 상태의 조화평균	75.73	71.59	71.74
지식 수	455	-	-

<표 3> 맥파/호흡 기반 생리학적 스트레스 성능평가

PPG/RSP	인포메틱스 엔진	표준 k-NN (k=1)	퍼지 k-NN (k=1)
평균 정확도	95.83±0.65	96.07±0.41	96.07±0.41
웰니스 상태의 조화평균	84.04	79.07	79.07
지식 수	88	-	-

<표 4> 신체적 웰니스 성능평가 (Fitbit Flex)

Activity	인포메틱스 엔진	표준 k-NN (k=1)	퍼지 k-NN (k=1)
평균 정확도	94.20±0.77	92.39±1.20	
웰니스 상태의 조화평균	95.43	93.27	93.27
지식 수	45	-	

<표 5> 신체적 웰니스 성능평가 (식습관)

Dietary	인포메틱스 엔진	표준 k -NN ($k=1$)	퍼지 k -NN ($k=1$)
평균 정확도	80.70±1.45	79.25±3.09	
웰니스 상태의 조화평균	75.80	63.20	63.20
지식 수	171	-	

<표 6> 신체적 웰니스 성능평가 (1채널 무선 홀터 심전계)

ECG	인포메틱스 엔진	표준 k -NN ($k=1$)	퍼지 k -NN ($k=1$)
평균 정확도	99.06±1.16	98.85±1.35	
웰니스 상태의 조화평균	98.21	97.88	97.88
지식 수	5	-	

<표 7> 웰니스 설문조사를 통한 주관적 웰니스 성능평가

Survey	인포메틱스 엔진	표준 k -NN ($k=1$)	퍼지 k -NN ($k=1$)
평균 정확도	54.67±2.98	63.14±2.66	50.03±11.58
웰니스 상태의 조화평균	51.74	52.97	52.69
지식 수	287	-	-

<표 2-7>은 통계적 교차검증 동안, 훈련 데이터로부터 설계된 지식베이스로부터 실험 데이터를 예측하였을 때 평균 예측 정확도와 웰니스 상태들의 조화평균 (harmonic mean) 값을 보여준다.

실험결과에서 볼 수 있듯이, 개발된 인포메틱스 엔진의 분류능력은 2가지 벤치마크 방법들을 기반으로 추론된 성능과 비교하였을 때, 평균 예측 정확도에서는 유사하거나 보다 좋은 예측 성능을 보여주었고, 각 웰니스 상태별 예측 정확도의 조화평균에서는 웰니스 설문조사를 제외한 모든 서비스 부분에서 보다 좋은 예측 성능을 보여주었다. 위의 결과는 인포메틱스엔진이 불균형 분포 (imbalanced or unbalanced distribution)를 포함한 라이프로그 정보에서 추가적인 샘플링 기법 (즉 over-sampling 혹은 under-sampling)을 활용하지 않더라도 균형화된 예측 결과를 제공해 줄 수 있음을 보여준다.

5. 결론

본 연구에서는 다양한 웨어러블 기기를 통해 수집된 이종의 라이프로그 정보를 기반으로 개인의 건강상태를

분석할 수 있는 인포메틱스엔진의 프레임워크를 소개하였고, 웰니스휴먼케어 1차 실증시험서비스 기간 동안에 수집된 라이프로그 정보를 활용하여 성능을 평가하였다. 향후 연구에서는 이종의 라이프로그 정보들이 서로 조합되었을 때에 성능비교가 필요할 것으로 사료된다.

ACKNOWLEDGEMENT

본 연구는 산업통상자원부에서 지원하는 미래산업선도 기술개발사업(10044353)에 의해 수행되었으며, 수집된 데이터의 활용은 DGIST 생명윤리위원회 (DGIST-150113-HR-001-02)의 규정을 준수하여 수행되었습니다.

참고문헌

- [1] Yi M. K, Whangbo T. K, "A selective lifelog management scheme for personalized healthcare services," Journal of Korean Institute of Next Generation Computing, Vol. 10, No. 6, pp. 14-21, 2014.
- [2] Wellness4u, <https://play.google.com/store/apps/details?id=com.sdinfo.wellness4u&hl=ko>
- [3] Son C. S, Kim Y. N, Kim H. S, Park H. S, Kim M. S, "Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches," Journal of Biomedical Informatics, Vol. 45, pp. 999-1008, 2012.
- [4] Cover T. M, Hart P. E, "Nearest neighbor pattern classification," IEEE Transactions on Information Theory, Vol. IT-13, No. 1, pp. 21-27, 1967.
- [5] Keller J. M, Gray M. R, Giverns J. A, "A fuzzy k-nearest neighbor algorithm," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-15, No. 4, pp. 580-585, 1985.