

논문 2018-13-25

연관 분류 마이닝 기법을 활용한 지식기반 신체활동 평가 모델

(A Knowledge Based Physical Activity Evaluation Model
Using Associative Classification Mining Approach)

손창식*, 최락현, 강원석

(Chang-Sik Son, Rock-Hyun Choi, Won-Seok Kang)

Abstract : Recently, as interest of wearable devices has increased, commercially available smart wristbands and applications have been used as a tool for personal healthy management. However most previous studies have focused on evaluating the accuracy and reliability of the technical problems of wearable devices, especially step counts, walking distance, and energy consumption measured from the smart wristbands. In this study, we propose a physical activity evaluation model using classification rules, induced from the associative classification mining approach. These rules associated with five physical activities were generated by considering activities and walking times in target heart rate zones such as 'Out-of Zone', 'Fat Burn Zone', 'Cardio Zone', and 'Peak Zone'. In the experiment, we evaluated the prediction power of classification rules and verified its effectiveness by comparing classification accuracies between the proposed model and support vector machine.

Keywords : Physical activity evaluation model, Physical activity criteria, Classification mining, Class association rule

1. 서론

최근 들어 스마트폰의 이용률이 급속적으로 증가함과 동시에 웨어러블 기기들이 출시되면서 건강과 관련된 다양한 제품과 서비스들이 쏟아져 나오고 있다. 예를 들어, 스마트폰에 내장된 가속도계를 이용한 신체활동 측정 애플리케이션 서비스와 스마트 밴드 (Fitbit, Jawbone, Nike Fuel Band, Xiaomi Mi Band 등)를 이용한 신체활동 모니터링

*Corresponding Author (changsikson@dgist.ac.kr)

Received: Mar. 23 2018, Revised: Apr. 30 2018, Accepted: Apr. 30 2018.

C. Son, R. Choi, W. Kang: DGIST

※ 본 논문은 산업통상자원부 산업핵심기술개발사업 (10063553)과 미래산업선도기술개발사업 (10044353) 그리고 DGIST R&D 프로그램 (18-IT-02)의 지원으로 수행되었습니다. 실험에서 사용된 데이터는 DGIST 생명윤리위원회의 승인 후 사용하였습니다 (IRB No. DGIST-160303-HR-012-01).

(활동량, 심박, 수면 등) 서비스가 대표적이다. 웨어러블 기기와 건강관리 시장이 꾸준히 증가하고 있음에도 불구하고 사용자들이 이러한 제품을 편리하게 사용하고 있는지, 또는 지속적으로 활용하는지에 대한 체계적인 연구는 그다지 많지 않다. 대부분의 선행연구는 웨어러블 기기의 기술적 문제점 특히 스마트 밴드로부터 측정된 걸음 수, 보행거리 및 에너지 소모량에 대한 정확성과 신뢰성을 평가하는데 중점을 두고 있다 [1-5]. 하지만 사용자들이 제품이나 서비스를 통해 지속적인 만족감을 얻기 위해서는 서비스 사용성과 지속성을 유지시켜줄 수 있는 전략이 필요하다 [6]. 이를 위해서 다양한 피드백 서비스 기능, 재미 혹은 흥미를 유발할 수 있는 인터랙션 방법, 직관적이고 공감할 수 있는 정보제공 방법, 적절한 운동중재 (intervention) 방법, 그리고 사용자의 성취감 향상을 위한 방법이 고려되어야 한다.

Cadmus-Bertram [7]는 25명의 과체중 (혹은 비만) 여성과 폐경기 여성을 대상으로 Fitbit 스마트 밴드를 착용하게 하고 주당 150분, 하루 10,000 걸

음의 증강도에서 고강도의 운동중재를 시행한 후, 16주간 여성 참여자의 신체활동에 대한 순응도를 관찰하였다. 그 결과 운동중재 기간 동안에 여성 참여자는 상당히 활동적인 신체활동 (주당 평균 신체활동 시간은 82분; 하루 평균 신체 활동량은 7,540 걸음)을 보여주었다. 이는 단기간 동안에 적절한 운동중재가 개인의 신체활동을 유지하거나 개선시키는 데 도움이 될 수 있다는 간접적인 결과를 보여준다. Pina [8]은 사용자가 오랜 시간 동안에 움직임이 적은 좌식 활동 (sedentary activity)을 줄이기 위한 방안으로, 15분 간격으로 수집된 데이터를 근거로 활동량이 적을 때 스마트폰에서 사용자에게 알림을 주는 시스템을 제안하였다. 이 시스템은 서비스의 사용성과 지속성을 유지시켜줄 수 있는 피드백 서비스로서, 사용자가 자신의 부족한 신체활동을 깨닫고 행동의 변화를 유도할 수 있다는 점에서 선행연구들과 차별성을 보여준다.

하지만 스마트 밴드로부터 수집될 수 있는 다양한 신체활동 정보 중에서 단순히 걸음 수만을 분석하여 신체활동을 구분하기란 쉽지 않다. 따라서 본 연구에서는 이러한 문제점을 고려하여 스마트 밴드로부터 수집된 신체활동 정보 (분 당 걸음 수, 심박수, 수면)를 활용하여 신체활동을 구분할 수 있는 평가기준을 제시하고, 이를 활용하여 사용자의 신체활동을 정성적인 관점에서 구분할 수 있는 방법을 제안한다.

II. 본 론

2.1 데이터 수집 과정

하루 일상생활 중에 개인의 라이프로그 (lifelog) 정보를 측정하기 위해서, 3축 가속도계 (3-axis accelerometer), 고도계 (altimeter)와 광학 심박수 센서 (optical heart rate sensor)로 구성된 Fitbit Charge HR을 활용하였다 [9].

그림 1에서 나타난 것처럼, Fitbit 클라우드 서버에서 제공하는 라이프로그는 신체활동과 관련된 걸음 수, 이동거리, 소모 칼로리, 심박수 그리고 수면 정보 (깨어난 횟수 및 시간, 뒤척인 횟수 및 시간, 실제 수면시간, 수면 효율) 등이 있다. Fitbit 서비스 정책에 따라 사용자의 라이프로그 정보를 수집하기 위해 PlanIT 앱 [10]을 서드 파티 앱 (third-party application)으로 등록한 후, 신체활동과 관련된 모든 시계열 데이터 (intraday time-series data)를 수집하였다. 데이터 특성에 대

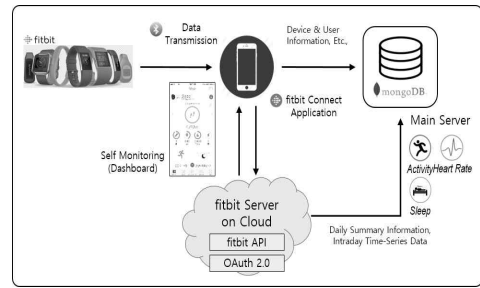


그림 1. 라이프로그 수집 과정
Fig. 1 Life-log collection process

한 보다 자세한 내용은 Fitbit Web API Reference를 참고하기 바란다 [11].

2.2 신체활동 평가기준

본 연구에서는 하루 동안에 개인의 신체활동을 평가하기 위해서, 분 당 걸음 수 (steps), 심박수 (heart rate), 그리고 수면 (sleep) 시계열 데이터를 활용하였다. 좀 더 정확한 신체 활동량을 계산하기 위해서 ‘하루 중 낮잠을 포함한 수면 동안에 움직임으로 인한 걸음 수’, ‘운동과 관련된 신체활동이라고 볼 수 없는 걸음 수’에 해당하는 분 당 40걸음 미만의 활동과 관련된 모든 샘플들을 제거하였다.

일반적으로 신체 활동량은 성별 혹은 나이와 같은 인구통계학적 특성에 따라 다르므로, 이를 고려하여 기준을 정의하는 것이 중요하다. 신체활동의 평가 기준은 William Haskell과 Smuel Fox에 의해서 고안된 식 (1)의 최대 심박수 (maximum heart rate) [12]를 기준으로 4가지 목표 심박영역 (target heart rate zones)을 정의하였다.

- 가) ‘좌식영역’ (Out-of Zone): 하루 중 분당 심박수가 최대 심박수의 50% 미만인 심박수
- 나) ‘지방연소영역’ (Fat-Burn Zone): 최대 심박수의 50%이상 70% 미만인 심박수
- 다) ‘심장강화영역’ (Cardio Zone): 최대 심박수의 70%이상 85% 미만인 심박수
- 라) ‘최대 심박영역’ (Peak Zone): 최대 심박수의 85%이상인 심박수

$$HR_{\max} = 220 - age \quad (1)$$

식 (1)에서 ‘age’는 만 나이 (western age)를 나타낸다.

개개인의 신체활동의 구분은 하루 중 분 당 심

박수가 ‘좌식영역’에 속한 걸음수를 제외한 나머지 3가지 목표 심박영역에서 전체 활동량 (total activity, TA , $TA \geq 0$)을 고려하여 5단계로 구분하였다.

- 가) ‘비활동’ (Inactive): 3가지 목표 심박영역의 전체 걸음 수가 목표 걸음 수의 20% 미만인 활동 상태
- 나) ‘좌식 활동’ (Sedentary Active): 전체 걸음 수가 목표 걸음 수의 20% 이상 40% 미만인 활동 상태
- 다) ‘저강도 활동’ (Moderately Active): 전체 걸음 수가 목표 걸음 수의 40% 이상 60% 미만인 활동 상태
- 라) ‘중강도 활동’ (Vigorously Active): 전체 걸음 수가 목표 걸음 수의 60% 이상 80% 미만인 활동 상태
- 마) ‘고강도 활동’ (Extremely Active): 전체 걸음 수가 목표 걸음 수의 80% 이상인 활동 상태

$$TA = \frac{Steps_{FatBurn} + Steps_{Cardio} + Steps_{Peak}}{GoalSteps} \quad (2)$$

식 (2)에서 $GoalSteps$ 은 하루 동안 목표 걸음수를 나타낸다. 일반적으로 하루 목표 걸음 수가 10,000 걸음으로 설정되어 있지만, 사용자가 Fitbit 앱이나 대시보드를 통해 수정할 수 있다. 또한 $Steps_{FatBurn}$, $Steps_{CardioZone}$, $Steps_{PeakZone}$ 은 각각 하루 동안에 ‘지방연소영역’, ‘심장강화영역’, 그리고 ‘최대 심박영역’의 걸음 수를 의미한다.

2.3 연관 분류 마이닝

연관 분류 (associative classification)는 데이터 베이스로부터 유용한 패턴을 추출하는 지식발견 (knowledge discovery)의 주요 단계들 중 하나로, 예측모델 혹은 분류기를 설계하기 위해서 2가지 잘 알려진 데이터 마이닝 기법인 연관규칙 마이닝 (associative rule mining)과 분류 과정을 하나로 통합한 것이다 [13-15].

연관 분류 마이닝에서 훈련 데이터 혹은 데이터-셋 T 는 m 개의 속성 A_1, A_2, \dots, A_m 과 클래스들의 리스트 C , $C = \{c_1, c_2, \dots, c_s\}$ 로 구성되고, 모든 속성 값은 카테고리 값으로 변환된 정성적인 값으로 표현된다.

[정의 1] T 의 샘플은 속성 A_j 와 값 a_{ij} , 그리고 클

래스 c_k 의 조합 (combination)으로 표현한다.

[정의 2] 항목 (item)은 속성-값 쌍 $\langle (A_j, a_{ij}) \rangle$ 으로 표현한다.

[정의 3] 항목집합 (item-set)은 데이터-셋 T 에 포함된 항목들의 집합 $\langle (A_{i1}, a_{i1}), \dots, (A_{im}, a_{im}) \rangle$ 으로 표현한다.

[정의 4] 규칙항목 (rule-item) r 은 $\langle itemset, c_k \rangle$, $c_k \in C$ 로 표현한다.

[정의 5] 규칙항목 r 의 지지도 빈도 (support count) 즉 $supp_count(r)$ 는 T 에서 클래스 레이블 c_k 을 포함하는 샘플들 중에서 항목집합들과 일치하는 샘플 수를 의미한다.

[정의 6] 규칙항목 r 의 지지도 (support)는 $supp_count(r)/|T|$ 로 정의하고, $supp_count(r)/|T| \geq minsupp$, $minsupp \in [0, 1]$ 을 만족하는 경우를 빈발 규칙항목 (frequent rule-item)이라고 한다.

[정의 7] 규칙항목 r 의 신뢰도 (confidence)는 $supp_count(r)/ant(r) \geq minconf$, $minconf \in [0, 1]$ 로 정의한다. 여기서 $ant(r)$ 은 T 에서 규칙항목 r 의 항목집합과 일치하는 샘플 수를 의미한다.

[정의 8] 클래스 연관 규칙 (Class Association Rule, CAR)은 $(A_{i1}, a_{i1}) \wedge \dots \wedge (A_{im}, a_{im}) \rightarrow c_k$ 로 표현되고, 규칙의 입력부 (antecedent part) $(A_{i1}, a_{i1}) \wedge \dots \wedge (A_{im}, a_{im})$ 는 m 개의 항목으로 구성된 항목집합들을 의미하고, 규칙의 출력부 (consequent part)는 클래스 c_k 를 나타낸다.

표 1은 클래스 연관 규칙 생성 알고리즘의 의사 코드 (pseudo code)를 나타낸다. 라인 1의 R_1 은 하나의 항목 (즉, 속성-값 쌍)으로 구성된 집합들 중에서 최소 지지도를 만족하는 후보 규칙들을 나타내고, 라인 2의 CAR_1 은 이들 중에서 최소 신뢰도를 만족하는 규칙들을 의미한다. 라인 3에서 7은 이전 단계에서 선별된 $k-1$ 개의 항목으로 구성된 후보 규칙들을 조합함으로써, 새로운 k 개의 항목으로 구성된 후보 규칙들을 생성하고 (라인 4), 최소 지지도를 만족하는 규칙들을 추출하는 과정을 나타낸다 (라인 5). 이 과정은 최소 지지도와 신뢰도를 만족하는 규칙이 더 이상 존재하지 않을 때까지 반복한다.

표 1의 과정으로부터 생성된 클래스 연관 규칙은 훈련 데이터 샘플 수와 특성에 따라, 적게는 수십 개에서 많게는 수천 혹은 수만 개의 후보 규칙들이 생성될 수 있다. 본 연구에서는 보다 최소화된

표 1. 클래스 연관 규칙 알고리즘

Table 1. Class association rule algorithm

Input:	train data T ;	Output:	CAR
1	$R_1 = \{1\text{-rule items}\}$		
2	$CAR_1 = \text{generate_rules}(R_1)$		
3	FOR ($k = 2; R_{k-1} \neq \emptyset; k++$) do		
4	$R_k = \{k\text{-rule items from } R_{k-1}\}$		
5	$CAR_k = \text{generate_rules}(R_k)$		
6	end FOR		
7	$CAR = \bigcup_k CAR_k$		
8	Sort the rules, i.e., CAR , according to the criteria shown in Table 2 (see Table 2)		
9	Remove the redundant rules or rules that don't cover any training case (see Table 3)		

표 2. 규칙 순위화 기준

Table 2. Rule ranking criteria

Given two associative rules, r_i and r_j , r_i precedes r_j ($r_i > r_j$) if:
1) The confidence of r_i is greater than that of r_j .
2) The confidence values between r_i and r_j are the same, but the support of r_i is greater than that of r_j .
Confidence and support values between r_i and r_j are
3) the same, but r_i has fewer conditions, i.e., number of items, in its left hand side than that of r_j .
Confidence, support and cardinality of r_i and r_j are the
4) same, but r_i is associated with a more representative class than that of r_j .
All above criteria are identical for r_i and r_j , but r_i was
5) generated from items that have higher order in the training data than that of r_j .

연관 규칙을 얻기 위해서 표 2에서 제시된 규칙의 우선순위 기준 (rule ranking) [13]을 적용한 후, 표 3의 규칙선택 기준 (rule selection criteria)을 사용하여 최종 규칙을 획득하였다 (표 1의 라인 8과 9).

표 3은 생성된 모든 후보 규칙들 중에서 중복된 규칙을 제거하기 위한 선택기준과 하나의 디폴트 출력을 생성하는 과정을 보여준다. 라인 1에서 8은 중복된 규칙을 제거하는 단계로서, 보다 높은 우선순위를 가진 이전의 후보 규칙에 의해서 분류되지 않고 최소한 한 개 이상의 훈련 샘플을 분류할 수 있는 규칙을 찾을 때까지 반복한다. 디폴트 출력 (라인 9에서 10)은 이전에 선택된 규칙에 의해서 분류되지 않고 남은 훈련 샘플들의 클래스 빈도를 고려하여 가장 높은 절대빈도를 가진 클래스로 할당한다. 만약 이 과정에서 동일한 클래스 빈도를 가진 레이블이 둘 이상인 경우, 훈련 데이터 T 에서 상대적으로 높은 빈도를 나타낸 클래스로 할당한다.

표 3. 규칙 선택기준과 디폴트 출력

Table 3. Rule selection criteria and a default output

Input:	The set of rules, R , sorted to descending order and the training data T ;	Output:	R'
1	FOR each rule $r \in R'$ with the highest priority		
2	find all samples in T that match r antecedent part		
3	IF no sample matches r antecedent part		
4	remove r		
5	ELSE		
6	insert r into R'		
7	delete all samples in T that are covered by r		
8	end FOR		
9	IF $len(T) > 0$ then		
10	select the majority class as a default class from the remaining samples in T		

표 4. 예측 절차

Table 4. Prediction procedure

Input:	test data T_S and ranked rules R ;
Output:	predicted labels P
1	FOR each sample t_s in test data T_S
2	FOR each rule r in the ranked rules R
3	IF rule r fully matches t_s condition
4	insert class label of r into P
5	break
6	end FOR
7	IF there is no rule r that matches t_s condition
8	insert default output into P
9	end FOR

표 4는 클래스 레이블을 포함하지 않는 새로운 샘플 데이터의 출력 클래스를 예측하는 과정을 나타낸다. 표 4의 라인 1에서 9는 표 3으로부터 추출된 연관 규칙으로부터 샘플 데이터 t_s 의 출력 클래스를 결정하는 과정으로, 만약 규칙집합 R 내에서 임의의 규칙 r 의 조건과 일치하는 경우 규칙 r 의 출력을 샘플 데이터 t_s 의 출력 클래스로 할당하고, 그렇지 않은 경우 순차적으로 다음번의 규칙과 비교한다. 만약 일치하는 규칙이 존재하지 않는다면 디폴트 출력을 샘플 데이터의 클래스로 할당한다 (라인 7과 8).

III. 실험결과

3.1 연구 대상 및 데이터 특성

연구 대상자는 2016년 4월 1일부터 4월 30일까지 웰니스휴먼케어서비스플랫폼 실증사업에 참여한 경북 구미지역의 사업장 2곳에 종사하는 사무직 근

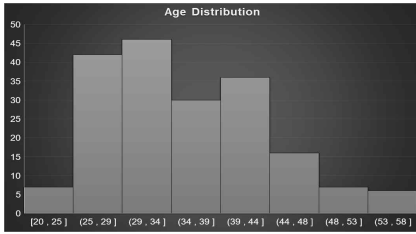


그림 2. 연령 분포
Fig. 2 Age distribution

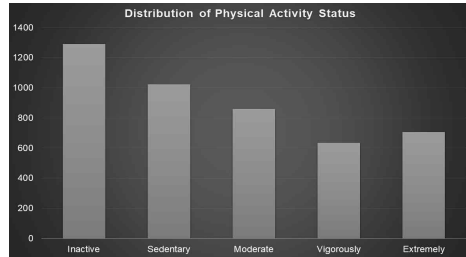


그림 4. 신체활동 상태의 분포
Fig. 4 Distribution of physical activity status

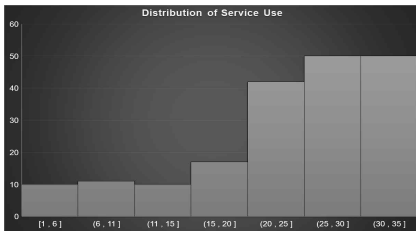


그림 3. 서비스 사용일 분포
Fig. 3 Distribution of service use

로자와 전남 여수지역의 사업장 2곳에 종사하는 제조 및 생산직 근로자 190명을 대상으로 하였고, 하루 단위를 기준으로 약 한 달 동안 4,514건의 시계열 데이터와 이와 관련된 추가적인 라이프로그 정보를 수집하였다. 수집된 데이터로부터 개인의 신체 활동량을 평가하기 위해서 근로자의 ‘만 나이’, ‘성별’, ‘체질량 지수’, ‘흡연’ 및 ‘음주 유무’ 뿐만 아니라, 4가지 목표 심박영역 즉 ‘좌식영역’, ‘지방연소영역’, ‘심장강화영역’, ‘최대 심박영역’에서의 걸음 수와 보행시간(분)에 대한 특징들을 추출하였다.

그림 2에서 4는 약 한 달 동안 웰니스휴먼케어 서비스플랫폼을 이용한 대상자의 연령, 서비스 사용일, 신체활동 상태에 대한 경향을 보여준다. 연령 분포에서는 30대와 40대의 참여자가 가장 높은 분포를 보였으며, 그 다음으로 20대와 50대 순으로 조사되었다. 또한 서비스 사용일 분포에서는 한 달 중 20일 이상의 이용분포가 약 70% 차지하는 것으로 조사되었고, 신체활동 상태에서는 ‘비활동’이 가장 높았고, ‘좌식활동’, ‘저강도 활동’, ‘고강도 활동’, ‘중강도 활동’ 순의 분포를 보였다.

3.2 신체활동 평가를 위한 분류 지식

클래스 연관규칙 마이닝에서 최소 지지도와 신뢰도에 대한 적정 임계값의 결정은 연관규칙의 수 뿐만 아니라 분석모델 혹은 시스템의 성능에 직접

적인 영향을 준다. 일반적으로 지지도의 수준을 높게 결정할 경우, 생성 가능한 연관규칙의 수는 최소화할 수 있지만 대부분의 경우 확률적으로 낮은 신뢰도를 가진 규칙을 획득하기 때문에 과소적합(under-fitting) 문제를 야기하고, 그 반대의 경우 방대한 수의 연관규칙으로 인하여 과적합(over-fitting) 문제가 발생된다. 이러한 문제점을 최소화하기 위해서 여러 선행 연구에서는 지지도의 수준을 비교적으로 낮게, 반면에 신뢰도의 수준은 가능한 한 높게 정의를 한다 [13-15].

본 연구에서는 이들 선행연구에서 사용된 기준을 근거로 최소 지지도와 신뢰도를 각각 0.01 (1%)과 0.50 (50%) 수준에서 결정하였다. 개인의 신체활동 상태를 평가하고 해석할 수 있는 규칙을 유도하기 위해서는 수치적인 값들로 구성된 정량적인 특징들을 정성적인 값으로 변환하는 데이터 범주화(discretization) 과정이 선행되어야 한다. 그러므로 본 연구에서는 CAIM 알고리즘 [16]을 활용하여 수집된 데이터의 특징들 가운데에 정성적인 값으로 구성된 성별, 음주 및 흡연 유무를 제외한 나머지 정량적인 특징들에 우선 적용하였다. 그 결과 ‘만 나이’, ‘체질량지수’, ‘좌식영역’, ‘지방연소영역’, ‘심장강화영역’과 ‘최대 심박영역’에서의 걸음 수와 보행시간(분)은 표 5와 같이 범주화되었다. 표 5에서 두 번째 열은 실험에서 사용된 특징들을 나타내고, 세 번째 열은 CAIM 알고리즘에 의해서 결정된 각 특징에 대한 범주화 기준을 보여준다. 예를 들어, 나이의 경우 4개의 기준값을 근거로 다음과 같이 5가지 구간으로 범주화할 수 있다: 1) 나이 \leq 23.5; 2) 23.5<나이 \leq 25.5; 3) 25.5<나이 \leq 27.5; 4) 27.5<나이 \leq 43.5; 5) 나이>43.5.

이러한 방법으로 정량적인 값으로 표현된 모든 특징들을 범주화된 구간으로 변환한 후 클래스 연관규칙을 추출하였다. 실험결과 2,879개의 후보 규

표 5. 범주화 기준
Table. 5 Discretization criteria

No	Features (units)	Cut-Points	No	Features (units)	Cut-Points
1	Age (years)	23.5, 25.5, 27.5, 43.5	8	Total Steps in Fat-Burn Zone	1999.5, 3855.5, 5976.5, 7822.5
2	Gender	-	9	Total Walking Time in Fat-Burn Zone (minutes)	23.5, 48.5, 82.5, 102.5
3	Body Mass Index (kg/m ²)	24.25, 24.45, 24.55, 47.2	10	Total Steps in Cardio Zone	43.5, 407.0, 648.5, 2206.0
4	Smoking	-	11	Total Walking Time in Cardio Zone (minutes)	0.5, 3.5, 7.5, 14.5
5	Drinking	-	12	Total Steps in Peak Zone	53.5, 608.5, 634.5, 3245.5
6	Total Steps in Out-of Zone	20.5, 186.5, 622.0, 11981.5	13	Total Walking Time in Peak Zone (minutes)	0.5, 8.5, 18.5, 20.5
7	Total Walking Time in Out-of Zone (minutes)	0.5, 3.5, 9.5, 60.5			

표 6. 분류 성능
Table 6. Classification performance

Confusion Matrix	Target Status					
	S1	S2	S3	S4	S5	Sum
S1	1291	27	1	1	0	1320
S2	0	957	56	10	11	1034
S3	0	34	802	96	2	934
S4	0	4	1	493	35	533
S5	0	1	0	33	659	693
Sum	1291	1023	860	633	707	4514

S1, Inactive; S2, Sedentary Active; S3, Moderately Active; S4, Vigorously Active; S5, Extremely Active.

칙을 찾았고 그 중에서 중복된 규칙을 제거한 후 271개의 규칙과 한 개의 디폴트 출력을 획득하였다. 신체활동 평가를 위한 연관 규칙의 분류성능을 계산하기 위해서 혼동행렬 (confusion matrix)을 구성하였고, 신체활동별 정확도를 분석하였다 (표 6). 신체활동 평가모델의 정확도는 약 93.1%, 각 신체 활동 상태의 정확도는 ‘비활동’ (100.0%), ‘좌식활동’ (93.6%), ‘저강도 활동’ (93.3%), ‘중강도 활동’ (77.9%), ‘고강도 활동’ (93.2%)을 보여주었다. 그리고 각 신체활동 상태에서 가장 높은 신뢰도를 보여준 연관 규칙은 다음과 같이 조사되었다.

S1) ‘비활동’

IF Total Steps in Fat-Burn Zone ≤ 1999.5 and Total Steps in Cardio Zone ≤ 43.5 THEN

표 7. 10-겹 교차검정 실험 동안에 분류 성능
Table 7. Classification performance during 10-fold cross validation experiment

Evaluation Criteria	Models	
	SVM(%)*	Proposed Model(%)
Avg. Accuracy	91.8	92.2
Avg. Accuracy of Each Physical Activities		
Inactive	95.6	99.7
Sedentary Active	94.0	93.2
Moderately Active	90.9	91.8
Vigorously Active	86.3	77.2
Extremely Active	88.0	90.8
Avg. Number of Rules	-	255.4

* Experimental Parameters: Kernel Function, Polynomial; c=1.0; epsilon=1.0E-12; exponent=10; tolerance=0.001.

Inactive with Confidence (1.000), Support (0.264).

S2) ‘좌식활동’

IF 1999.5 < Total Steps in Fat-Burn Zone ≤ 3855.5 and Total Steps in Cardio Zone ≤ 43.5 THEN Sedentary Active with Confidence (1.000), Support (0.154).

S3) ‘저강도 활동’

IF Smoking is Yes and 3855.5 < Total Steps in Fat-Burn Zone ≤ 5976.5 and 48.5 < Total Walking Time in Fat-Burn Zone ≤ 82.5 and Total Steps in Cardio Zone ≤ 43.5 THEN Moderately Active with Confidence (0.973), Support (0.032).

S4) ‘중강도 활동’

IF $622.0 < \text{Total Steps in Out-of Zone} \leq 11981.5$ and $5976.5 < \text{Total Steps in Fat-Burn Zone} \leq 7822.5$ and $\text{Total Steps in Cardio Zone} \leq 43.5$ THEN Moderately Active with Confidence (1.000), Support (0.037).

S5) ‘고강도 활동’

IF $\text{Age} > 43.5$ and $\text{Total Steps in Fat-Burn Zone} > 7822.5$ and $\text{Total Walking Time in Fat-Burn Zone} > 102.5$ THEN Extremely Active with Confidence (1.000), Support (0.033).

제안된 평가모델과 SVM (Support Vector Machine with polynomial kernel function) 모델 [17]과의 객관적인 성능지표를 제공하기 위해서, 통계적 10-겹 교차검정 (10-fold cross validation) 실험 [18]을 수행하였다. 이때에 훈련과 실험 데이터는 전체 샘플 데이터를 랜덤으로 섞은 다음, 10개의 하위-데이터 그룹 (sub-data group)을 구성한 후 각 그룹에서 훈련과 실험 샘플의 비율을 90%와 10%로 정의하였고, 10번의 반복 실험을 수행한 후 평균 분류 성능을 계산하였다 (표 7).

제안된 평가모델이 SVM 모델에 비해 ‘좌식활동’과 ‘중강도 활동’에서 약간 낮은 식별능력을 보였지만, 나머지 신체 활동 상태에 대해서 더욱 개선된 성능을 제공하였다. 또한 전체 신체활동 상태의 식별능력은 제안된 모델이 평균 0.4% 개선된 결과를 보여주었다.

IV. 결론

본 연구에서는 Fitbit 스마트 밴드로부터 수집된 라이프로그 정보를 이용하여 개인의 신체활동을 평가할 수 있는 모델을 제안하였다. 개인의 신체활동을 정성적인 관점에서 해석하기 위해서 4가지 목표 심박영역에서 걸음 수와 보행시간(분)의 특징정보를 추출하였고, 이를 활용하여 5가지 신체활동 상태를 구분하였다. 실험에서는 제안된 모델의 객관적인 성능지표를 제공하기 위해서, 통계적 교차검정 실험을 수행하여 SVM 모델과 성능을 비교하였다. 향후 연구에서는 본 평가 모델에서 사용된 신체활동에 대한 평가기준의 통계적 신뢰성 검정과 다양한 분류 혹은 연관 규칙 마이닝 방법들과의 성능 비교 연구를 수행할 계획이다.

References

- [1] J. Takacs, C.L. Pollock, J.R. Guenther, M. Bahar, C. Napier, M.A. Hunt, “Validation of the Fitbit One Activity Monitor Device During Treadmill Walking,” *Journal of Science and Medicine in Sport*, Vol. 17, No. 5, pp. 496-500, 2013.
- [2] M.A. Case, H.A. Burwick, K.G. Volpp, M.S. Patel, “Accuracy of Smartphone Applications and Wearable Devices for Tracking Physical Activity Data,” *Journal of the American Medical Association*, Vol. 313, No. 6, pp. 625-626, 2015.
- [3] S.A. Petro, K.L. Dannecker, E.L. Melanson, FACSM, R.C. Browning, “Accuracy of Research and Consumer Physical Activity Monitors in Estimating Energy Expenditure,” *Journal of Medicine & Science in Sports & Exercise*, Vol. 43, No. 1, pp. 61-62, 2011.
- [4] J.M. Lee, Y.W. Kim, G.J. Welk, “Validity of consumer-based physical activity monitors,” *Medicine & Science in Sports & Exercise*, Vol. 46, No. 9, pp.1840-1848, 2014.
- [5] K.M. Diaz, D.J. Krupka, M.J. Chang, J. Peacock, Y. Ma, J. Goldsmith, K.W. Davison, “Fitbit®: an Accurate and Reliable Device for Wireless Physical Activity Tracking,” *Journal of Cardiology*, Vol. 185, pp. 138-140, 2015.
- [6] Y. Doh, S. Keum, S. Lee, J. Lee, “The Exploratory Study of Factors Which Influence to the Maintenance of Using Wearable Device for Healthcare - Interdisciplinary Approach to User Experiences Combined With Technology, Psychology and Interaction Perspectives,” *Journal of the Korean Institute of Information Scientists and Engineers*, Vol. 32, No. 11, pp. 37-45, 2014 (in Korean).
- [7] L.A. Cadmus-Bertram, B.H. Marcus, R.E. Patterson, B.A. Parker, B.L. Morey, “Use of the Fitbit to Measure Adherence to a Physical Activity Intervention Among Overweight or Obese, Postmenopausal Women: Self-monitoring Trajectory During 16 Weeks,” *Journal of Medical Internet Research MHealth and UHealth*, Vol. 3, No. 4, pp. 1-7, 2015.

- [8] L.R. Pina, E. Ramirez, W.G. Griswold, "Fitbit+: a Behavior-based Intervention System to Eeduce Sedentary Behavior," Proceedings of Pervasive Computing Technologies for Healthcare, pp.175-178, 2012.
- [9] Available on Fitbit, <http://www.fitbit.com/>
- [10] Plan-IT, Available on <https://play.google.com/store/apps/details?id=com.dgist.planit>
- [11] Fitbit Web API Reference, Available on <https://dev.fitbit.com/build/reference/web-api/>
- [12] R.A. Robergs, R. Landwehr, "The Surprising History of the HRmax=200-age Equation," Journal of Exercise Physiology Online, Vol. 5, No. 2, pp. 1-10, 2002.
- [13] F. Thabtah, "A Review of Associative Classification Mining," Journal of the Knowledge Engineering Review, Vol. 22, No. 1, pp. 37-65, 2007.
- [14] K. Song, K. Lee, "Predictability-based Collective Class Associative Rule Mining," Journal of Expert Systems With Applications, Vol. 79, No. 15, pp. 1-7, 2017.
- [15] B. Liu, W. Hsu, Y. Ma, "Integrating Classification and Association Rule Mining," Proceedings of on Knowledge Discovery and Data Mining, pp. 80-86, 1998.
- [16] L.A. Kurgan, K.J. Cios, "CAIM Discretization Algorithm," Journal of IEEE Transactions on Knowledge and Data Engineering, Vol. 16, No. 2, pp. 145-153, 2004.
- [17] C. Cortes, V. Vapnik, "Support-vector Networks," Journal of Machine Learning, Vol. 20, No. 3, pp. 273-297, 1995.
- [18] C. Son, Y. Kim, H. Kim, H. Park, M. Kim, "Decision-making Model for Early Diagnosis of Congestive Heart Failure Using Rough set and Decision Tree Approaches," Journal of Biomedical Informatics, Vol. 45, No. 5, pp. 999-1008, 2012.

Chang-Sik Son (손창식)

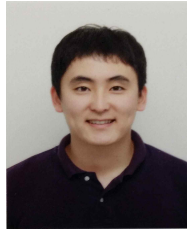


He is received his M.S. and Ph.D. degrees in intelligent systems from Catholic University of Daegu, Daegu, Korea in 2002 and 2006, respectively. He is currently

a senior researcher in the Convergence Research Center for Wellness at DGIST. His research interests include artificial intelligence areas in biomedical informatics and biomedical engineering.

Email: changsikson@dgist.ac.kr

Rock-Hyun Choi (최락현)



He is received his B.S and M.S degrees in information and communication system engineering from Daegu University, Korea in 2010 and 2012, respectively. He is currently a researcher

in the Convergence Research Center for Wellness at DGIST. His research interests include data mining, machine learning, and wireless network control system.

Email: choimosi@dgist.ac.kr

Won-Seok Kang (강원석)



He is received his B.S. and M.S. degrees in computer engineering from Yeungnam University, Korea in 1998 and 2000, respectively. He is currently a senior researcher in the Convergence Research

Center for Wellness at DGIST. His research interests include distributed simulation algorithms, mobile robot systems, and intelligent embedded systems.

Email: wskang@dgist.ac.kr