

0번째 순의 근삿값 (zeroth-order) [편집]

이는 최초의 근사값을 의미한다. 해당 결과값이 필요할 때 매우 단순한 식을 구성하여 빨리 뽑아내는 것을 의미한다. 가령, 통계학자로서의 당신은 실제로는 723,432명이 살고 있는 도시에 대해서 "이 도시에는 대략 수십만 명의 사람이 살고 있다."고 말할 수 있다.

1번째 순의 근삿값 (first-order) [편집]

단 하나의 주요 수치를 발견하여 좀 더 발전된 식을 통해 값을 정의하는 것을 의미한다. 예를 들면, 위에서 설명한 도시에 대해서 십만 단위를 정확하게 알아낸 것이 첫 번째 주요 수치로서, 약 70만 명이 살고 있다고 말할 수 있다.

2번째 순의 근삿값 (second-order) [편집]

두 개의 이상의 주요 수치를 적용한다. 예를 들면, 위에서 설명한 도시에 대해서 십만단위와 일만단위를(2가지) 정확히 알아 약 72만명이 살고있다고 말할 수 있다.

<요약>

- 3D 체형과 체지방 비율(BFP) 사이의 상관관계를 탐색하고 이러한 상관관계를 기반으로 3D 체형을 분석하여 BFP(Body Fat Percentage)를 추정하는 예측모델을 개발한다.
- 목표
 - 1) zeroth-order shape descriptors의 변수선택을 위한 회귀 모델(baseline regression)을 설정
 - 2) 예측력 향상을 위해 visual cue(=시각적 자료)를 이용하여 체형을 분류
- ➔ 본 연구에서는 모델 훈련 및 평가를 위한 근거 자료로 DXA BFA 사용
- ➔ 결과적으로, "BFP(=본 연구의 추정 도구)"와 "BOD, POD(=의료 수준에서의 체성분 평가 도구)"를 비교한다.
- ➔ 본 연구의 예측모델은 BOD, POD의 예측력보다 20.28% 증가함

<연구 순서>

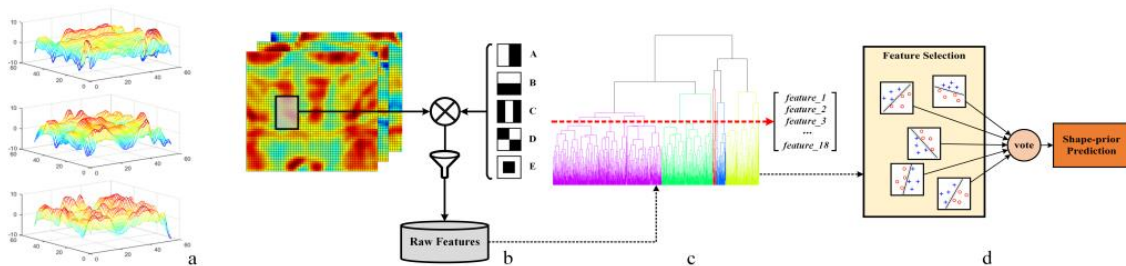
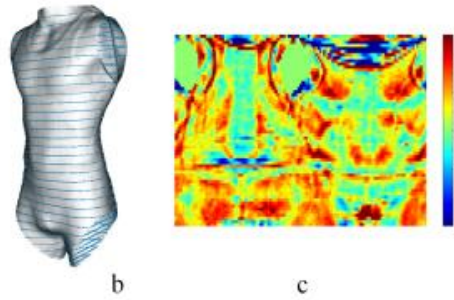


Fig. 2. Shape-prior classification. (a) A comparison of curvature distribution between three types of body shapes: obese (top), average (middle), and lean (bottom). (b) Raw feature extraction with Haar-like basis functions. (c) Using agglomerative hierarchical clustering to reduce the feature redundancy. (d) Feature selection and model training.

- (a) 체형을 세 가지로 분류(위에서부터, 비만-정상-저체중)
- (b) 변수 추출(Haar기법 기반)
- (c) 변수 중복성을 줄이기 위한 클러스터링
- (d) 변수 선택과 모델 훈련

<방법>



- 1) Shape descriptor를 세 가지로 정의
 - (1) Zeroth-order(그림b): 3D에서 직접 파생된 몸통부위 둘레
 - (2) First-order: 표면 (->second-order와 겹치므로 삭제)
 - (3) **Second-order(그림c): 곡률** → 곡률 함수를 이용하여 다양한 체형을 모델링한다.

- 곡률 데이터를 2D 이미지로 평탄화
 - Haar와 같은 기본함수를 사용하여 lean(날씬한 체형)과 non-lean으로의 매핑을 구별

⇒ 이 방법으로 체형을 분류한다.
- 2) 변수선택
 - (1) 계층적 클러스터링을 사용하여 변수 중복 줄이기
 - (2) Cut-off 임계값을 18개의 클러스터로 설정하고 클러스터 중심에 가장 가까운 18개의 변수 선택
 - (3) 18개의 변수를 4개로 조합하고, SVM을 사용해 각 조합을 평가(그림2-d)
 - (4) 일반화 오류를 평가하기 위해 LOOCV(Leave-one-out-CrossV) 수행(결과 섹션 참고)
- 3) Baseline Regression 회귀 모형 만들기
 - BFP 예측의 기준 회귀 모델을 훈련하기 위해 zeroth-order descriptor를 사용
 - ➔ zeroth-order descriptor: 3D에서 zeroth-order descriptor는 몸통 둘레, 표면적 등을 포함한 47개의 0차적 특징이다. 여기에 세 가지 특징을 더해(체중, 키, 체질량 지수) 총 50가지의 features
 - 가우스 프로세스 회귀(GPR) 선택
- 4) 예측 모델= Baseline Regression + shape-prior
 - Second-order(곡률)을 사용하여 체형을 lean(마른 체형)과 non-lean으로 분류
 - ➔ 이 분류를 통해 BFP를 정량화 한다.

<결과>

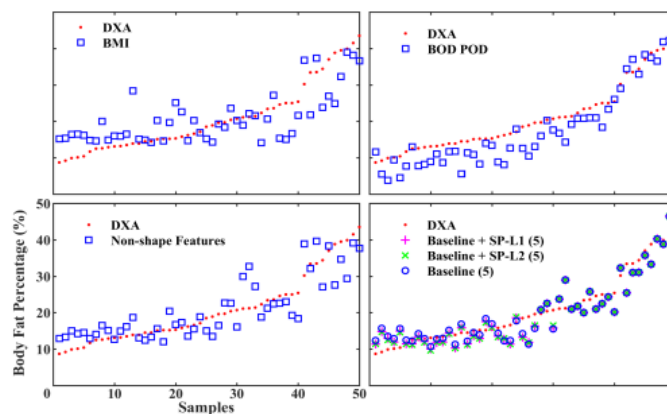
- 성인 남성 50명의 표본 크기로 데이터 수집

A. 체형 분류(shape-prior)

- 곡률 분포에 따라 lean(군살 없는)과 non-lean 체형을 라벨링하여 24개의 lean 체형과 26개의 non-lean 체형에 대한 샘플 추출하여 모델을 훈련시킨 결과, 정확도는 92%

B. 회귀(Baseline regression)

- 모델 훈련 및 테스트를 위해 LOOCV 수행
 - LOOCV(Leave-One-Out Cross-Validation 교차검증): 단 하나의 관측값만을 validation set으로 사용하고, 나머지 n-1개 관측값은 train set으로 사용하는 방법
 - 비교를 위해 두 가지 테스트 추가적으로 실시
 - 1) BFP 예측을 위한 회귀 모델의 유일한 예측변수로 BMI 사용
 - 2) BFP를 예측하기 위해 체중, 키, BMI, 가슴둘레, 엉덩이 둘레와 같은 non-shape features 사용
- ➔ 이 두 비교 모델의 일반화 오류를 평가하기 위해 LOOVC 수행



➔ DXA와 다른 측정지표들과의 비교

➔ 결과는 DXA가 평균적으로, BOD POD를 18.6%, non-shape features 모델을 28.76%, BMI 모델을 44.31% 능가하는 것으로 보아, DXA의 BFP 예측력이 더 좋음

C. 예측 모델: 회귀(B)+체형 분류(A)=회귀의 일반화 오류 평가

- 회귀 모형의 정확도를 향상시키기 위해 체형분류(shape-prior)를 사용할 것을 제안
- 회귀 모형이 non-lean(날씬하지 않은) 체형의 BFP를 과대평가하고 lean(날씬한) 체형의 BFP를 과소평가한다는 경향이 있다고 가정하여 일반화 오류를 평가한다.
- 예측 정확도 결과
 - 1) l1 norm을 가진 예측 모델: 평균적으로 BOD POD를 20.16%, non-shape 모델을 30.12%, BMI 모델을 45.36% 능가함
 - 2) l2 norm을 가진 예측 모델: 평균적으로 BOD POD를 20.39%, non-shape 모델을 30.32%, BMI 모델을 45.52% 능가함