멀티모달 기반 개인 맞춤형 피트니스 분석 및 LLM 코칭 시스템 설계에 관한 연구

박상현, 임재영, 한승민, 박성영, 윤수연*

국민대학교, 국민대학교, 국민대학교, 국민대학교, *국민대학교

tomshp@kookmin.ac.kr, dlawowo01@kookmin.ac.kr, handsomemin@kookmin.ac.kr, pou876@kookmin.ac.kr , *1104py@kookmin.ac.kr

A Study on the Design of a Multimodal Personalized Fitness Analysis and LLM-based Coaching System

Sanghyeon Park, Jae Yong Im, Seung Min Han, Park Seong Yeong , Sooyeon Yoon*

Kookmin Univ., Kookmin Univ., Kookmin Univ., Kookmin Univ., *Kookmin Univ.

요 약

본 연구는 기존 개인 트레이닝 서비스가 사용자의 실제 운동 자세를 정밀하게 분석하지 못하고, 개인별 운동 수행 수준을 반영한 맞춤형 코칭 제공에 한계가 있다는 문제를 해결하고자, Pose Estimation과 Pose Classification 기반의 자세 유사도 분석과 LLM을 결합한 멀티모달 개인 맞춤형 운동 코칭 시스템을 제안한다. 제안한 시스템은 사용자 운동 영상과 시범 영상을 입력으로 받아 MediaPipe Pose를 통해 프레임 단위로 33개의 인체 랜드마크를 추출하고, DTW 기반 시계열 정렬과 다양한 유사도 지표를 결합하여 관절 단위의 수행 정확도를 정량적으로 분석한다. 이후 분석 결과를 JSON 형태로 구조화하여 LLM에 입력함으로써, 관절 정확도에 기반한 자연어 운동 코칭 피드백을 자동 생성한다. 실험 결과, Bi-LSTM 기반 Pose Classification 단계에서 높은 분류 정확도를 달성하였으며, LLM을 활용한 코칭 생성 단계에서도 다양한 운동 동작과 촬영 각도 조건에서 구체적이고 일관된 피드백이 안정적으로 생성됨을 확인하였다. 본 연구는 정량적 자세 분석 결과를 자연어 코칭으로 연결하는 새로운 접근 방식을 제시하며, 향후 다양한 운동 동작 확장과 대규모 데이터셋을 활용한 고도화를 통해 실용적인 개인 맞춤형 피트니스 코칭 시스템으로 발전할 수 있을 것으로 기대된다.

Ⅰ. 서 론

1. 연구 배경 및 필요성

최근 보고에 따르면 글로벌 개인 트레이닝 소프트웨어 시장은 2025년 약 21억 달러 규모에서 2035년 약 30억 달러로 성장할 것으로 예측되며, 연평균 성장률은 약 4.1%에 달한다[1]. 하지만 기존 서비스는 사용자의 실제 운동 자세를 정밀하게 분석하거나 개인의 신체 특성과 운동 경험을 반영한 피드백을 제공하는 데 한계가 있다. 이를 Pose Estimation과 Pose Classification 등 유사도 분석 기술을 통해 영상으로부터 사용자의 운동 수행 상태를 정량적으로 분석하는 것이 가능하며, LLM은 이러한 분석 결과를 사용자 수준에 맞는 자연어 코칭으로 제공할 수 있다. 이에 본 연구에서는 Pose Estimation와 Pose Classification를 통해 Pose Similarity 분석한다. 그리고 분석 결과와 LLM을 활용하여 멀티모달 기반의 개인 맞춤형 피트니스 분석 및 코칭 시스템의 설계 방안을 제안하고자 한다.

Ⅱ. 관련 연구

1. Pose Estimation & Classification

Pose Estimation은 영상 입력으로부터 프레임 단위의 인체 관절 위치를 Keypoint 또는 Landmark 형태로 추출하는 기술로, 동작 인식 및 자세 분석을 위한 전처리 단계에서 널리 활용되고 있다. 기존 연구에서는 OpenPose, ViTPose, MediaPipe Pose[2] 등 다양한 방식이 제안되었으며, 이들은 RGB 영상으로부터 2D 또는 3D 관절 좌표를 추정한다. Pose Classification은 Pose Estimation을 통해 추출된 Landmark를 시퀀스로 입력받아 사용자의 동작 유형을 분류한다. Pose Classification을 위한 모델로는 LSTM, GRU, Bi-LSTM[3]과 같은 순환 신경망 기반 모델이 주로 사용된다.

Ⅲ. 실험

1. 데이터셋

본 연구를 위해서 [표 1]와 같이 운동 동작 분석을 위한 데이터셋을 구축하였다. 실제 사용자 환경을 반영하기 위해 YouTube에서 Squat, Situp 등을 포함한 총 8종의 운동 동작 영상을 수집하였다. 수집된 데이터는 총 145개 영상으로 구성되며, 전체 영상 길이는 약 2시간으로 총 1,203회의 반복 동작을 포함되어 있다. 각 영상에 대해 MediaPipe Pose를 적용하여 프레임 단위로 33개의 Keypoint를 추출하였다.

표 1. 데이터셋 개요

수집 매체	Youtube
수집 영상 개수	145개
총 영상 시간	약 2시간
총 동작 수	1203회
동작 구분	Squat,Lunge,Situp,High Knees,Bridge,Cobra,Side Lunge,Jumping Jack
Landmark 추출 방식	MediaPipe

2. 운동 코칭 파이프라인

본 연구는 [그림 1]와 같이 사용자 운동 영상과 시범(Original) 영상을 입력으로 받아, 자세 추출, 동작 인식, 자세 유사도 분석을 거쳐 LLM 기반 자연어 코칭을 생성하는 코칭 파이프라인을 제안한다. 먼저 사용자 영상과 시범 영상에 대해 MediaPipe Pose를 적용하여 프레임 단위 33개의 각 관절 및 신체 부위에서 33개의 Landmark를 추출하고, 체형·해상도·촬영 거리 차이의 영향을 줄이기 위해 좌표 정규화를 수행한다.

이후 두 영상 간 수행 속도와 길이 차이를 보정하기 위해 DTW 기반 시계열 정렬을 적용하며, 추가적으로 시범 영상 프레임을 기준으로 사용자 프레임 범위 내에서 코사인 유사도가 최대가 되는 매칭을 선택하여 각 영상의 속도 차이를 보완한다. 정렬된 Landmark 시퀀스에 대해서는 코사인 유사도를 이용해 자세의 형태적 유사도를 계산하고, L2 Norm과 OKS(Object Keypoint Similarity)를 결합한 복합 지표로 공간적 위치 오차를 함께 반영한다. 유사도 결과는 프레임 단위 및 관절의 쌍에 대한 평균 점수를 포함하는 JSON 형태로 출력된다. 마지막으로 LLM 코칭 모듈은 유사도 측정 결과(JSON)를 입력으로 받아, 코칭 문장을 생성한다.

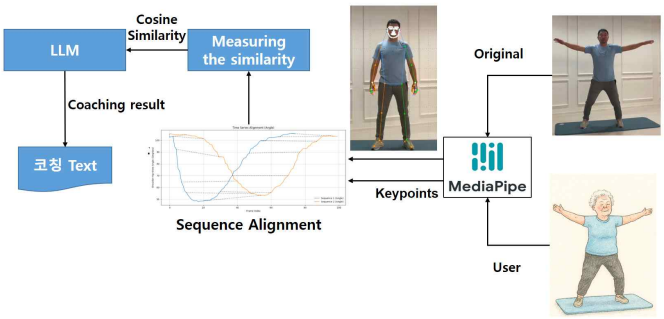

그림 1. LLM 기반 개인 맞춤형 코칭 시스템 파이프라인

3. 실험 및 결과 분석

본 연구에서는 Bi-LSTM 모델을 통해 Pose Classification을 수행하였다. 실험에 사용된 데이터셋은 사전에 구축한 데이터를 기준으로 학습 데이터와 테스트 데이터를 8:2 비율로 분할하였다. 모델 학습을 위해 batch size는 16, dropout은 0.3, epoch 수는 80, learning rate는 1e-4로 설정하였다.

실험 결과, 전체 분류 정확도(Accuracy)는 0.9520으로 높은 성능을 달성하였다. [그림 2]에서 확인할 수 있듯이 대부분의 동작 클래스에서 우수한 분류 성능을 보였으나, 비교적 단순한 동작에 비해 동작의 변화가 크거나 유사한 자세 전환을 포함하는 동작에서는 오분류가 상대적으로 많이 발생하였다. 특히, side lunge, squat, lunge와 같이 앉았다가 일어나는 동작 패턴이 유사한 운동들 사이에서 혼동이 발생하는 경향이 관찰되었다.

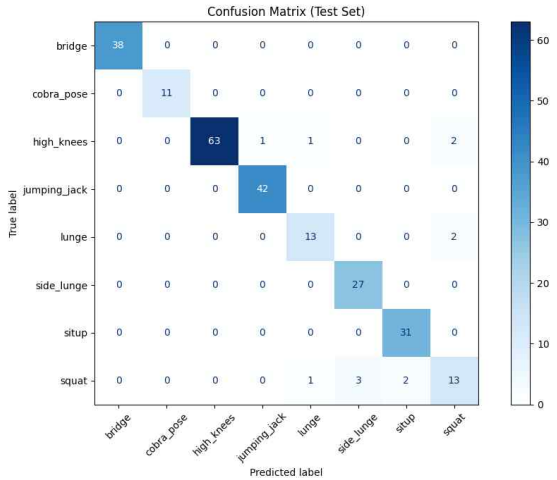

그림 2. Bi-LSTM 모델 Pose Classification Confusion Matrix

4. LLM 입력 결과

본 연구에서는 이전 단계인 Pose Classification 및 Pose Similarity 과정에서 생성된 JSON 기반 결과 데이터를 LLM에 입력하여, 사용자에게 제공할 자연어 기반 맞춤형 운동 코칭 피드백을 자동으로 생성하였다. 이를 위해 운동 코치 페르소나를 정의하고, 운동 동작별 특성, MediaPipe 기반 인체 랜드마크 체계, 관절별 점수 해석 규칙을 프롬프트에 반영하였다. 관절별 평균 점수는 해석 가능한 구조로 변환되어 운동 맥락 정보와 함께 입력되었으며, 그 결과 관절 정확도에 기반한 구체적이고 실행 가능한 코칭 피드백이 생성됨을 확인하였다. 또한, 다양한 운동 동작과 촬영 각도 조건에서도 일관된 피드백 결과가 유지됨을 통해 제안한 LLM 기반 피드백 생성 방식의 유효성을 검증하였다. [그림 3]은 이러한 LLM 기반 운동 코칭 피드백의 출력 예시를 보여준다.

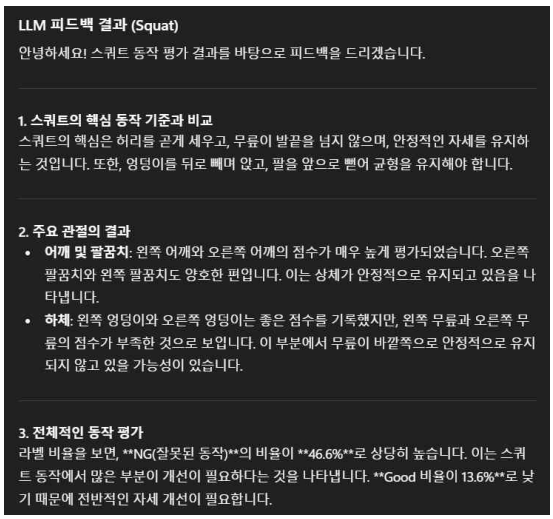

그림 3. Squat 동작에 대한 LLM 결과

Ⅴ. 결론

본 연구에서는 Pose Estimation과 Pose Classification을 기반으로 사용자의 운동 동작을 정량적으로 분석하고, DTW와 다양한 유사도 지표를 결합한 Pose Similarity 분석을 통해 관절 단위의 수행 정확도를 산출하였다. 또한, 이러한 수치 기반 분석 결과를 LLM에 입력하여 자연어 기반의 맞춤형 운동 코칭으로 변환함으로써, 영상 분석 결과를 사용자에게 직관적이고 이해하기 쉬운 형태로 제공하는 멀티모달 코칭 파이프라인을 제안하였다. 실험 결과, Pose Classification 단계에서 높은 분류 정확도를 달성하였으며, LLM을 활용한 코칭 생성 단계에서도 관절 정확도에 기반한 구체적이고 실행 가능한 피드백이 안정적으로 생성됨을 확인하였다. 향후 연구에서는 분석 대상 운동 동작의 종류를 확장하고, 더 대규모의 데이터셋을 활용한 학습 및 검증을 통해 다양한 운동 환경에서도 신뢰성 있는 코칭 결과를 제공할 수 있도록 연구를 확장할 예정이다.

ACKNOWLEDGMENT

본 연구는 2022년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음.(2022-0-00964)

This research was supported by the MIST(Ministry of Science, ICT), Korea, under the National Program for Excellence in SW), supervised by the IITP(Institute of Information & communications Technology Planning & Evaluation) in 2022 (2022-0-00964)

참고문헌

[1] Research Nester, “Personal Training Software Market Size, Growth Analysis and Forecast (2026 - 2035),” 2025.

[2] C. Lugaresi et al., “MediaPipe: A Framework for Building Perception Pipelines.”

[3] A. Graves, S. Fernández, and J. Schmidhuber, “Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition.”