

K리그-서울시립대 공개 AI 경진대회

Track 2: 아이디어 개발 부문

팀명	team deTACter
프로젝트명	deTACter

최종 산출물 - Summary

K리그-서울시립대 공개 AI 경진대회

팀명	team deTACTer
팀원	김태현
서비스명	deTACTer
한줄소개	이벤트 데이터로부터 주요 공격 시퀀스를 추출하여, 고급 행동평가 지표인 VAEP와 함께 애니메이션으로 제공함으로써 코치, 감독, 팬 모두가 특정 팀의 위협적인 시퀀스를 쉽게 알 수 있게 함

서비스 제안 배경

배경

- 현대 축구에서는 경기당 수천 개 이상의 이벤트 데이터가 발생하며, 이를 사람이 일일이 분석하는 것은 시대 착오적입니다. 훨씬 간단하게 코치, 팬, 선수가 특정 팀의 주요 공격 시퀀스를 선택해서 볼 수 있게 합니다.

문제정의

- 자동화 : 수천 개의 시퀀스를 자동 클러스터링하여 팀의 핵심 전술을 요약/제공합니다.
- 객관화 : VACP 점수를 통해 어떤 움직임이 득점 확률을 얼마나 높였는지 수치로 증명합니다.
- 시각화 : 추상적인 데이터 좌표를 애니메이션으로 변환하여 직관적인 전술 이해를 돕습니다.

데이터 분석 및 활용

데이터 분석 전략

- SPADL 변환: 각 액션을 '누가, 언제, 어디서(시작/종료 좌표), 어떤 행동(패스, 드리블 등)을 했는지'에 대한 고정된 속성값으로 재정의하여, 데이터를 표준화된 액션 언어인 SPADL (Soccer Player Action Description Language) 포맷으로 변환하고 분석 호환성을 확보합니다.
- 공격 방향 통일: 모든 시퀀스에서 공격시 방향을 왼쪽에서 오른쪽(L->R) 공격 방향으로 통일합니다. 이를 위해 매 경기/전반/후반별로 골키퍼(GK)의 평균 위치를 계산하여, 골키퍼가 0.5(하프라인)보다 오른쪽에 위치할 경우 해당 진영의 모든 좌표를 반전시킵니다.
- 좌표 정규화 및 클리핑: 피치 규격(105m x 68m)에 의존하지 않도록 모든 좌표를 0~1 범위로 정규화(Normalization)합니다. 또한, 측정 오류로 인해 피치 경계를 벗어난 좌표는 0.0과 1.0으로 클리핑(Clipping)하여 모델 입력의 안정성을 확보합니다.

데이터 분석 및 활용

데이터 분석 전략

- Disconnected Event 제거: 데이터의 특성으로 인해, 특정 이벤트(tackle, take-on 등)가 이전 이벤트의 종료 지점과 현재 이벤트의 시작 지점이 물리적으로 연결되지 않는 경우가 있고 이는 시퀀스 역추적 시 큰 영향을 주지 않으므로, 허용 오차를 초과하면 is_dis=True로 마킹하여 시퀀스 역추적 시 해당 지점에서 끊어지도록 처리, 학습 데이터의 연속성을 보장합니다.
- 정지 이벤트 제거 및 리시브 보정 : 전술적 의미가 약한 이벤트를 정제합니다.
 - 제자리 리시브 삭제: 이동 거리 없이 제자리에서 공을 받고 끝나는 'Pass Received' 이벤트는 궤적 분석에 불필요하므로 제거합니다.
 - 이동 리시브 변환: 공을 받은 후 이동하는 리시브 이벤트는 실질적인 전진 드리블로 간주하여 'Carry' (Dribble)로 변환합니다.
 - 궤적 연결: 정지된 이벤트와 다음 액션 사이에 위치 차이가 발생할 경우, 인위적인 'Carry' 이벤트를 생성하여 궤적의 연속성을 보완합니다.

데이터 분석 및 활용

전술 시퀀스 추출 전략

- 성과 기반 역추적: 단순히 시간 순으로 데이터를 나열하는 것이 아니라, '유효 슈팅' 또는 '상대 진영 박스 진입'이라는 명확한 성과가 있어야, 유의미한 공격 시퀀스로 인정하여 해당 이벤트를 기준으로 과거 액션을 역추적합니다.
- 중복 방지: 데이터의 통계적 독립성을 위해, 한 번 추출된 액션이 다른 시퀀스에 중복 포함되지 않도록 배타적으로 처리합니다.

핵심 활용 방안

- 데이터에 시작 좌표와 끝 좌표가 존재하고, 이벤트들의 좌표를 선으로 이었을 때 직선의 조합이 나오는 것을 이용하여 공의 궤적을 추론합니다. 이 공의 궤적을 바탕으로 공격 전술을 추론하고 클러스터링 합니다.
- 시퀀스와 VAEP값(추후 설명)을 결합하여 공격 시퀀스의 객관적인 가치를 평가합니다.

AI 모델 및 기능 설계

AI 모델 구조

- 전술적 움직임을 컴퓨터가 이해할 수 있는 Latent Space으로 변환하기 위해 SequenceAutoencoder를 설계하였습니다.
 - Encoder (Transformer/GRU Based): 가변 길이의 시퀀스 좌표(x, y)를 입력받습니다. 실험 결과, 긴 의존성(Long-term Dependency) 학습에 유리한 Transformer Encoder 구조를 최종 채택하였습니다.
 - Decoder: 인코더가 압축한 임베딩 벡터를 다시 원본 좌표 시퀀스로 복원합니다. 모델은 MSE Loss를 최소화하는 방향으로 학습되며, 이 과정에서 전술의 핵심 특징(공간 활용, 이동 궤적, 템포)을 압축된 벡터에 담게 됩니다.
- TPE 최적화 : 하이퍼파라미터는 TPE(Tree-structured Parzen Estimator) 알고리즘을 통해 탐색하며, 각 모델의 의도에 맞게 MSE, Silhouette Score 등 을 최적화 방향으로 모델을 진화시킵니다.
- 가치 평가 모델 : 액션의 가치를 평가하기 위해 VAEP프레임워크를 적용하였습니다.
 - 구조: CatBoostClassifier(XGboost, LGBM을 포함한 실험에서 가장 안정적)를 사용하여 현재 게임 상태(State)에서 득점 확률(P_{scores})과 실점 확률($P_{concedes}$)을 예측합니다.
 - 계산: 어떤 액션 a_i 의 가치는 해당 액션 수행 전후의 득점/실점 확률 변화량으로 정의됩니다.
 - 검증: 5-Fold Cross Validation을 수행하고 Brier Score를 통해 확률 예측의 정확도를 검증합니다.

AI 모델 및 기능 설계

- OPTICS 클러스터링 : Encoder의 임베딩 벡터를 아주 빈번한 패턴부터 드물게 발생하는 패턴까지, 데이터의 밀도차를 스스로 파악하여 유연하게 클러스터링합니다. min_samples 파라미터를 통해 전술적 의미가 없는 일회성 우연 사례들을 자동으로 걸러내어 분석의 순도를 높입니다.

적용 방안 (Tactical Clustering)

- 학습된 임베딩 벡터에 Grid Grouping (1차)과 OPTICS (2차) 클러스터링을 적용하여 유사한 시퀀스들을 클러스터링합니다. 클러스터링 된 시퀀스들을 VAEP 값과 함께 애니메이션으로 만들어 정량적 인사이트를 코칭 스태프와 선수, 팬들에게 직관적으로 제공합니다. 또한, 웹페이지를 통해 서비스하여 접근성을 높이고, VAEP에 대한 간략한 설명을 통해 이해도를 높일 수 있습니다.

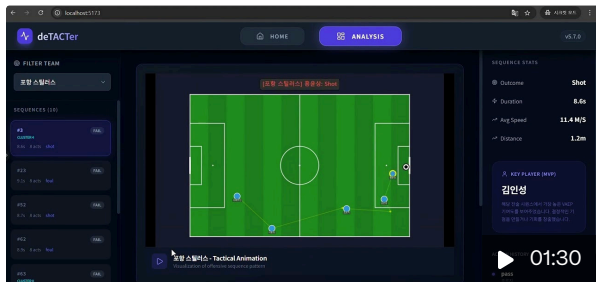
MVP 개발 구현 과정

개발 스택

- Data Processing: Python, Pandas , Socceraction (SPADL 변환/VAEP 피처 엔지니어링)
- Deep Learning: PyTorch (Transformer/GRU Autoencoder)
- Machine Learning: CatBoost (VAEP 확률 예측), Scikit-learn (OPTICS, K-Means 등 클러스터링)
- Optimization: Optuna (TPE 기반 하이퍼파라미터 최적화)
- Backend/Pipeline: Modular Python Scripts, YAML Config, Batch Processing
- Frontend (Web): React 18, Vite, TailwindCSS (Styling), Framer Motion (Animation), Recharts (Data Visualization)

MVP 작동 과정

데모 링크



deTACTer 데모영상



실용성 및 확장 가능성

가치와 효율성

- 코칭 스태프: 분석 시간 단축 및 데이터 기반의 객관적 전술 지시 가능합니다.
- 선수: 자신의 움직임이 팀 승리에 얼마나 기여했는지(VAEP) 시각적으로 확인 가능합니다.
- 팬: 단순 관람을 넘어, 골이 만들어지는 전술적 빌드업 과정의 심층적인 이해로 이어져, K리그에 대한 흥미와 팬심 상승을 유도할 수 있습니다.

현재 한계와 확장 가능성

- 이벤트 데이터가 많아질수록 분석 가능한 시퀀스 수가 증가하며, 전술 패턴 추출의 신뢰도와 서비스 완성도가 함께 향상 가능합니다.
- 트래킹 데이터가 결합될 경우, 온더볼 행동뿐 아니라 오프더볼 움직임까지 평가 범위에 포함 가능
 - 실제 경기 기준으로 선수의 온더볼 관여 시간은 90분 중 평균 약 3분에 불과하여, 보다 체계적 분석에는 오프더볼 분석이 필수적입니다.
- 경기 시작 전 또는 중계 전 프리뷰 형태로 제공 시, 시청자의 전술 이해도와 경기 몰입도 상승을 기대할 수 있습니다.