Akhil Santosh B

Project Report: Predicting Soccer Player Positions Using Facial Images and Player Attributes

Abstract

This project explores predicting soccer player positions using a multimodal approach that combines facial image data and numerical player attributes from the FIFA 2017 dataset. Utilizing a CNN for image processing and an MLP for numerical data, these modalities were fused to predict player positions across 24 classes. Data preprocessing, transfer learning with FaceNet, and feature engineering contributed to robust architecture. The multimodal model is demonstrated as an attempt to show its potential to perform well compared to single-modality models, highlighting the potential of integrating diverse data sources.

Problem Definition and Project Goals

The purpose of this project is to classify soccer players into positional roles (e.g., LW, RCB) using both facial images and player attributes from the FIFA 2017 dataset. The dataset was sourced from Kaggle and includes:

- **Features**: Physical statistics, technical skills, and ratings.
- **Images**: 432 labeled facial images.
- **Target Variable**: Derived from "Club_Position" columns, representing 24 unique classes.

The project aims to leverage both structured numerical data and unstructured image data to improve classification accuracy using a multimodal deep learning approach.

Related Work

Several studies have addressed player classification and performance analysis using structured data from FIFA datasets. For instance:

1. Papers employing numerical and image attributes for regression-based house price prediction: Link
2. Kaggle notebooks analyzing FIFA data for skill comparisons without incorporating image data. Link

This project distinguishes itself by incorporating facial images into the predictive framework, leveraging transfer learning via FaceNet, and integrating modalities for improved performance.

Data Exploration and Preprocessing

Dataset Challenges and Preparation:

1. **Data Alignment**: Manually Filtered CSV and Image datasets and took a subset of 317 players ensuring matching records between numerical and image data.
2. **Feature Engineering**: Encoded categorical variables (e.g., "Preferred Foot") using one-hot and binary methods. Excluded irrelevant fields such as "Name" and "Nationality".
3. **Image Labeling**: Initially, it is just a folder with images and player names as labels. Organized images into 24 subfolders based on positions derived from the CSV file.

Numerical Data Preprocessing:

- Identified and removed columns with a high percentage of missing values (e.g., "National_Position" with 48.58% missing data).
- Encoded categorical features such as "Preferred Foot" and "Work Rate" using mappings and one-hot encoding, respectively.
- Removed suffixes from height and weight columns, converting them to numeric types.
- Standardized numerical predictors using MinMaxScaler.

Image Data Preparation:

- Organized images by player positions.

- Split data into training (70%) and validation (30%) directories for each position class.
- Applied Data Augmentation with random transformations like flipping, rotation, and zoom for increased variability and robustness during training.

Dataset Splits:

- **Numerical Data**: Stratified sampling ensured consistent class distribution, splitting data into training (75%) and validation (25%) sets.
- **Image Data**: Augmented and split into training and validation datasets using Image dataset from directory.

Data Analysis

Model Architectures:

1. **Multi-Layer Perceptron (MLP):**
   - Input Layer: Processes standardized numerical data.
   - Hidden Layers: Five dense layers with ReLU activation, batch normalization, and dropout for regularization.
   - Output Layer: Softmax layer predicting probabilities for 24 classes.
2. **Convolutional Neural Network (CNN):**
   - **Pretrained Feature Extraction:**
     - Model: Pre-trained FaceNet extracts 128-dimensional embeddings from input images.
     - Input Layer: Processes images of size (160, 160, 3).
     - Dense Layers: Incorporate batch normalization and dropout for fine-tuned classification.
   - **Data Augmentation:** Applied transformations like flipping, rotation, and zoom during model training to enhance generalization.
   - **Output Layer:** Softmax activation predicts class probabilities for 24 positions.
3. **Multimodal Model:**

- **Dual Input Branches:** The model consists of two separate branches:
    - **Image Branch:** Processes facial images through the CNN, using FaceNet embeddings as input features.
    - **Numerical Branch:** Processes numerical data using the MLP.
- **Feature Fusion:** The outputs from the CNN and MLP branches are concatenated to create a unified representation.
    - This concatenation step allows the model to leverage both visual and structured data simultaneously.
- **Post-Fusion Layers:**
    - Dense layers with ReLU activation refine the concatenated features.
    - Dropout layers are applied to prevent overfitting.
    - A final softmax layer predicts probabilities for the 24 position classes.
- **Techniques Used:**
    - Batch normalization after concatenation ensures stable learning.
    - Regularization via dropout to improve generalization.
    - Adam optimizer and sparse categorical cross-entropy loss for efficient training.

Training Workflow:

- Loaded, cleaned, and preprocessed data for numerical and image inputs.
- Encoded categorical features and transformed numerical data into consistent formats.
- Trained MLP and CNN models independently before integrating into a multimodal architecture.
- Evaluated using Adam optimizer and sparse categorical cross-entropy loss over 150 epochs.

Evaluating, Tuning, and Improving Model

Evaluation Metrics:

- Accuracy and F1-score were used to assess performance on validation data.
- Learning curves indicated initial overfitting mitigated through dropout, batch normalization, and data augmentation.

Improvements:

1. Adjusted CNN learning rate and added batch normalization.
2. Increased data augmentation to address limited image diversity.
3. Fine-tuned the multimodal architecture to balance contributions from both modalities.
4. Stratified sampling maintained class distributions during data splits.

Conclusion

The multimodal approach effectively combined numerical attributes and facial image features for player position classification. The model hasn’t demonstrated improved performance over single-modality baselines yet, probably due to challenges such as class imbalance and limited image data-constrained accuracy.

Key Findings:

- Multimodal architectures can leverage diverse data for enhanced feature representation.
- Data preprocessing and augmentation are critical in addressing small and imbalanced datasets.

Future Directions:

1. Expand the dataset to include more player images and balanced class distributions.
2. Experiment with additional pre-trained image models.
3. Explore alternative fusion techniques for combining modalities.
4. Address class imbalance with synthetic data generation or oversampling techniques.

References

1. Kaggle: FIFA 2017 Dataset - https://www.kaggle.com/datasets/artimous/complete-fifa-2017-player-dataset-global/data
2. Relevant technical papers and Kaggle notebooks analyzing FIFA datasets for player classification and performance prediction.