https://nbn-resolving.org/urn:nbn:de:bsz:ch1-qucosa2-806704

Sports Scene Searching, Rating & Solving using AI

Robert Marzilger, Fabian Hirn, Raul Aznar Alvarez & Nicolas Witt

Fraunhofer Institute for Integrated Circuits IIS, Germany

Abstract

This work shows the application of artificial intelligence (AI) on invasion game tracking data to realize a fast (sub-second) and adaptable search engine for sports scenes, scene ratings based on machine learning (ML) and computer-generated solutions using reinforcement learning (RL). We provide research results for all three areas. Benefits are expected for accelerated video analysis at professional sports clubs.

Keywords: AI, ML, RL, tracking data, invasion games

Introduction

Video analysis of invasion games like soccer, ice hockey, and basketball, despite available software tools, is still a quite manual and challenging task, but very important for game preparation & post-game analysis. Identification of meaningful scenes for trainers, preparation for the playing style of opponents, as well as discussing different solutions for certain scenes are repeating tasks that could be supported by omnipresent tracking data and the use of machine learning techniques (ML). The following sections explain the use of ML for *Scene Search*, a scheme for *Scene Rating,* and the use of reinforcement learning for *Scene Solving*, which generates new solutions for scenes played in real games never seen before.

Methods and Results

Scene Searching is realized with a similarity-based search, that retrieves the most similar scenes to a query scene within fractions of a second using deep representation learning (Löffler et al., 2022a) and that adapts to expert annotations.

Fundamental problems with similarity search stem from a lack of available labels and are threefold (Löffler et al., 2022a). First, the unordered structure of samples from team sports like football is caused by the lack of a generalized role assignment between the two teams. Every player's position may change dynamically throughout the game. This affects the similarity calculation because the assignment between players of two scenes is undefined. Second, the high dimensionality of positional tracking sampled at 25Hz for 23 targets in football, and the combinatorial complexity of pairwise computations, limit the scalability of searching with raw samples.

In our search component, we address the assignment problem by calculating optimal assignments between pairs of scenes $(Scene_1, Scene_2)$ with the Hungarian algorithm (Kuhn, 1955). This produces pairwise optimal assignments of players from one scene to their counterparts in the other scene. However, while this necessary step solves the assignment, its computational complexity of $\mathcal{O}(n^3)$ is still infeasible, and the scalability problem even intensifies.

Hence, we jointly learn to estimate the assignment problem and a lower dimensional representation of the complex raw tracking data to solve the three fundamental problems. We employ Deep Siamese Metric Learning to learn both i) a distance preserving lower-dimensional embedding $f()$ and ii) an estimation of the assignment problem. We use the Euclidean distance $d()$ as a proxy for semantic similarity, since Sha et al. (2016) showed its efficacy despite its apparent simplicity, with the Siamese loss that minimizes the difference between the embedding and raw distances:

$$Loss_{Siamese}(Scene_1, Scene_2) = (\|f(Scene_1) - f(Scene_2)\|_2 - d(Scene_1, Scene_2))^2 \quad (1)$$

However, we individualize the similarity metric in a second step. Even though the Siamese embedding preserves the Euclidean well and accelerates search by orders of magnitude (Reeb et al., 2020), we seek to improve upon it. Specifically, the Euclidean distance suffers from the Curse of Dimensionality (Bellman, 1961), and in practice does not differentiate semantically, e.g., no weighting of ball or player trajectories. We propose to learn a metric from human annotations using triplets (Hoffer et al., 2015) of query $Scene_a$, similar $Scene_p$, and dissimilar $Scene_n$. This extends our Siamese Network by learning experts' notions of similarity, which may resemble semantics more closely. We implement this cost-effectively by leveraging transfer-learning and furthermore by only querying the most informative samples through Active Learning (Löffler et al., 2022b). We fine-tune via triplet loss:

$$Loss_{Triplet}(Scene_a, Scene_p, Scene_n) = max\left(\|f(Scene_a) - f(Scene_p)\|_2 - \|f(Scene_a) - f(Scene_n)\|_2 + 1, \quad 0\right) \quad (2)$$

Despite inconsistent annotations from noisy oracles, we see an increase in triplet accuracy of 5-10% triplet accuracy after only 20 queries.

Scene Rating of soccer scenes is done using three method types: i) distance functions such as the distance to the goal and the distance won (between the last and the first frame). ii) formation spread, average distance of players to the formation's center of mass. This is based on the tactical idea, that an increased distance between defenders makes it harder to defend and an attacking team must position itself to spread the opposing formation (Rafelt, 2021). iii) pressure metric, as proposed by Link et al. (2016). In addition, we implement an Expected Goals (xG) metric based on the distance to goal (Altman, 2015), and the pitch control metric (Fernandez et al., 2018).

We provide statistical breakdowns of scenes for professional soccer analysts. We can choose scenes in two different ways. The first lists 5s scenes that are pre-filtered scenes to satisfy the criteria that one team possesses the ball for at least 90% of the time. Alternatively, scenes can be selected before or after an annotated event (1s to 10s). We provide visualizations of the scene and rating, and the video clip. In Fig. 1, we show trajectories of players and ball (a), and radar plot with rating metrics (b). In addition, we show (animated dynamic) plots of the scene and an animation of the pitch control metric (c).

These animations can be played forwards and backward to allow analysts to gain insight into the play of either the attacking or defending team.

(a) (b) (c)

Fig. 1 (a) Scene trajectory plot; (b) Radar chart with rating; (c) Pitch control.

It is planned to extend the rating to data-driven models for the prediction (in the next five seconds) of box entry, shot on goal, and scoring for attacking scenes.

Scene Solving is realized using the Google Research Football Environment (Kurach et al., 2020) and Reinforcement Learning (RL) techniques to generate new solutions from real scenes. The generated artificial solutions can then be sorted by different rating criteria described above, to identify valuable new proposals. The RL-agent consists of a Multilayer-Perceptron (MLP) of two layers [512,512], trained for 15 million steps using Proximal Policy Optimization (PPO) (Schulman et al., 2017). The state is represented using a 169-dimensional vector (see Fig 2) and extends previous work by adding relative positions from the active player to other entities.

Agent Team	Opponent Team	Team ball owner	Ball Information	Controlled player	Sticky actions	Controlled player	Controlled player
Position & Direction	Position & Direction	One-hot encoder	Position & Direction	One-hot encoder	Binary	Relative pos&dir to all players	Relative pos&dir to ball & ref.points
44	44	3	6	11	9	42	10

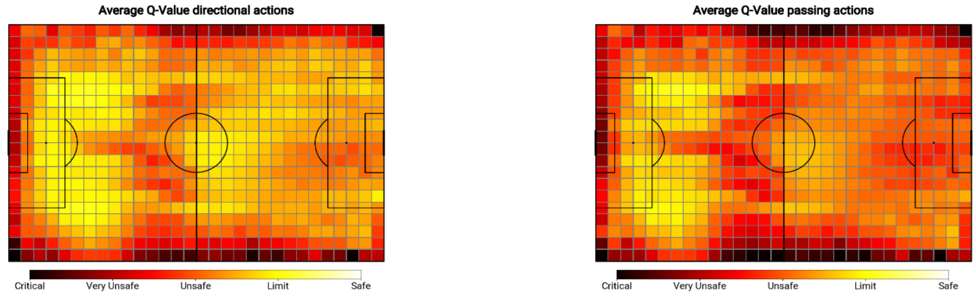
Fig. 2 State space representation

In soccer not losing the ball is of utmost importance. When risky actions are taken, the chance of losing the ball increases. Hence, we made the agent risk aware by defining constraints whenever the agent is in danger of losing the ball. Our risk estimator uses past transitions generated from the environment. The risk estimator allows the user of our application to tune the willingness of the playing agent to take a risk in its solutions. *Risk-tolerant* agents solutions advance faster, but the agents also lose the ball more easily, leading to counterattacks. *Risk-averse* agents focus on keeping the ball in possession which can be helpful if the team is already winning the match.

In Fig 3, we show the average safety value for passing and directional actions. Passing actions carry a higher risk than directional actions, especially when the agent is closer to the opponent's goal. Consequently, risk-averse agents do not pass when they are close to the goal, whereas risk-tolerant agents may perform a pass in such situations.

Conclusion

We think that a fast and adaptable search engine based on scene ratings and tracking data in combination with advances in computer aided scene solving will help to make video analysis more effective and may lead to new playing strategies.



(a) (b)

Fig. 3 (a) Safety values of directional actions; (b) Safety values of passing actions.

Conflict of interest We declare no conflicts of interest.

Funding We would like to acknowledge support for this project from the Bavarian Ministry of Economic Affairs, Infrastructure, Energy and Technology as part of the Center for Analytics-Data-Applications (ADA) within the framework of “BAYERN DIGITAL II”.

References

D. Altman. (2015). Beyond Shots: A new approach to quantifying scoring opportunities. OptaProForum. https://northyardanalytics.com/Dan-Altman-NYA-OptaPro-Forum-2015.pdf

J. Fernandez, & L. Bornn. (2018). Wide Open Spaces: A statistical technique for measuring space creation in professional soccer. *MIT Sloan Sports Conference*.

D. Link, S. Lang, & P. Seidenschwarz. (2016). Real Time Quantification of Dangerousity in Football Using Spatiotemporal Tracking Data. *PLoS ONE 11*(12).

M. Rafelt. (2021, March 27). Wie Guardiolas 3-2-2-3 (letztlich) das Abwehrspiel löst. *Spielverlagerung.de.* https://spielverlagerung.de/2021/03/27/wie-guardiolas-3-2-2-3-letztlich-das-abwehrspiel-loest/

Löffler, C., Reeb, L., Dzibela, D., Marzilger, R., Witt, N., Eskofier, B. & Mutschler, C. (2022a). Deep Siamese Metric Learning: A Highly Scalable Approach to Searching Unordered Sets of Trajectories. *ACM Transactions of Intelligent Systems and Technology, 13*(1), Article 6.

Sha, L., Lucey, P., Yue, Y., Carr, P., Rohlf, C. & Matt hews, I. (2016). Chalkboarding: A New Spatiotemporal Query Paradigm for Sports Play Retrieval. *21st International Conference on Intelligent User Interfaces*, pp. 336–347.

Reeb, L., Dzibela, D., Marzilger, R., & Witt, N. (2020) Effiziente Suche und Bewertung von Szenen in Spielsportarten. *spinfortec digital*, pp. 16-17.

Bellman, R. (1961). Adaptive Control Processes. A Guided Tour. *Princeton University Press, 255*.

Hoffer, E. & Nir A. (2015). Deep Metric Learning Using Triplet Networks. Similarity-Based Pattern Recognition SIMBAD. Lecture Notes in Computer Science, 9370.

Löffler, C., Fallah, K., Fenu, S., Zanca, D., Eskofier, B., Rozell, C. J., Mutschler, C. (2022b) *Active Learning of Ordinal Embeddings: A User Study on Football Data. ArXiv.* 2207.12710

Kuhn, H. (1955). The Hungarian Method for the Assignment Problem. *Naval Research Logistics Quarterly, 2*, pp. 83-97.

Kurach, K., Raichuk, A., Stanczyk, P., Zajac, M., Bachem, O., Espeholt, L., Riquelme, C., Vincent, D., Michalski, M., & Bousquet, O. (2020). Google research football: A novel reinforcement learning environment. *AAAI Conference on Artificial Intelligence, 34*(4), pp. 4501–4510.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal Policy Optimization Algorithms. *ArXiv*. 1707.06347