Mattia Stival and Lorenzo Schiavon's contribution to the Discussion of 'Flexible marked spatio-temporal point processes with applications to event sequences from association football' by Narayanan, Kosmidis, and Dellaportas

Mattia Stival and Lorenzo Schiavon

Department of Economics, Ca' Foscari University of Venice, Venice, Italy

Address for correspondence: Mattia Stival, Department of Economics, Ca' Foscari University of Venice, San Giobbe - Cannaregio 873, Fondamenta San Giobbe, 30121 Venice, Italy. Email: mattia.stival@unive.it

In association football, there exist two types of in-game data: event-sequence data provide qualitative information on the succession of ball-related events in time and space outlining a *local* view of the match; tracking data report with fine temporal granularity the positions of the ball and every player, allowing for a *global* view of the match, in which the ball is just one of many interacting objects. Using event-sequence data, the authors place themselves within a local perspective, with the possible undesirable consequence of missing part of relevant information. Figure 1 illustrates this idea, by showing two alternative situations in which red Player 1 has different probabilities of scoring a goal due to the different positions of the defenders.

By construction, event-sequence data do not report the positions of defenders, producing estimates potentially affected by the presence of unobserved information correlated with the observed outcome. To mitigate the impact of this shortcoming, multiple solutions can be accounted for. One option consists in enriching the event-sequence data with additional qualitative knowledge regarding the game situation. If also player tracking data are available, an alternative solution would be merging any observed event in the event-sequence data with the corresponding snapshot of player and ball positions of the tracking data set.

When this extra information is missing, a careful model specification is required. The process defined by the authors represents a brilliant answer to this challenge. Since any sequence of ball-related events is partially determined by the player's locations on the pitch, the observation of a certain sequence carries with it additional implicit information about team positioning. Unlike most of the recent literature based on Markovian assumptions (Schulte et al., 2017; Rudd, 2011; Singh, 2019), the Hawkes-fashioned specification in equations (10)–(12) recognises all past events as relevant factors in determining match evolution. Recognising event sequences as partially linked to the positioning of players may justify why, according to the authors, 'event sequences in football have a significant dependence on their history'.

With this model, predictive probability density functions of the occurrence of any marked event can be derived. Hence, one could reconstruct via simulation the distribution of the number of any event combination observable in a limited amount of time. This is allowed by joint modelling the event sequence and the time between subsequent events, with temporal modelling standing as a crucial feature to formulate in-game forecasts, and representing a key difference with respect to other frameworks based on a discrete-time game-states representation (Decroos et al., 2019; Fernández et al., 2021). To exploit such potentialities, fast updates of the parameters are needed, requiring new computational approaches (Panos et al., 2021).

A final remark concerns how model complexity is addressed. It would be interesting to compare the association rule learning method with alternative strategies in which the modelling assumptions (e.g. Su et al., 2016) or prior distributions (Ishwaran & Rao, 2005) directly account for sparsity.

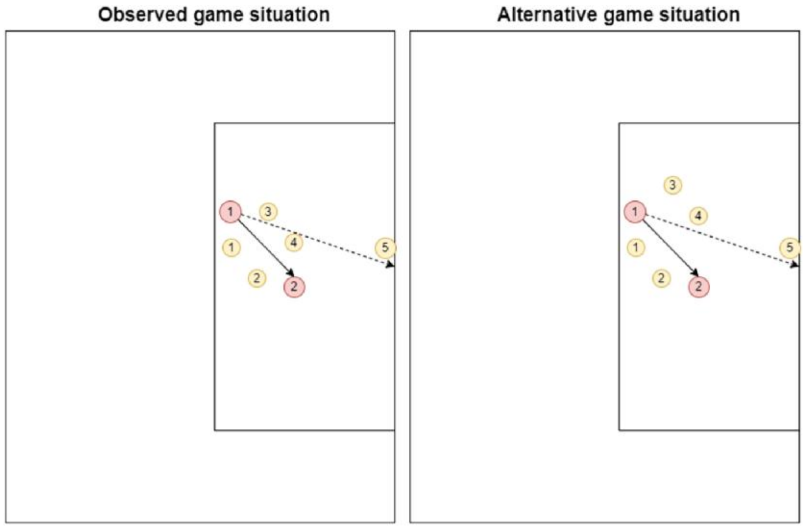

Figure 1. On the left, a sketch of the pass (continuous arrow) that led to the goal scored by Jack Wilshere (red circle, player 2) for Arsenal against Norwich, recalled in Figure 1 of the paper. Norwich player positions are reported with smaller yellow circles. On the right, a sketch of an alternative game situation, in which defenders 3 and 4 of Norwich City are placed, arbitrarily, in different positions with respect to the observed ones. Dashed arrows denote unobserved shot events.

To conclude, we thank the authors for their contribution, hoping that our thoughts may enrich the discussion.

Conflict of interest: None declared.

References

Decroos T., Bransen L., Van Haaren J., & Davis J. (2019). Actions speak louder than goals: Valuing player actions in soccer. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 1851–1861). Association for Computing Machinery. https://doi.org/10.1145/3292500.3330758

Fernández J., Bornn L., & Cervone D. (2021). A framework for the fine-grained evaluation of the instantaneous expected value of soccer possessions. *Machine Learning*, *110*(6), 1389–1427. https://doi.org/10.1007/s10994-021-05989-6

Ishwaran H., & Rao J. S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *The Annals of Statistics*, *33*(2), 730–773. https://doi.org/10.1214/009053604000001147

Panos A., Kosmidis I., & Dellaportas P. (2021). 'Scalable and interpretable marked point processes', arXiv, arXiv:2105.14574, preprint: not peer reviewed.

Rudd P. (2011, September 24). A framework for tactical analysis and individual offensive production assessment in soccer using Markov Chains. In *New England Symposium on Statistics in Sports*. Cambridge, MA: The Harvard University Science Center.

Schulte O., Khademi M., Gholami S., *et al.* (2017). A Markov Game model for valuing actions, locations, and team performance in ice hockey. *Data Mining and Knowledge Discovery*, *31*, 1735–1757. https://doi.org/10.1007/s10618-017-0496-zt

Singh K. (2019). *Introducing expected threat (xt)*. https://karun.in/blog/expected-threat.html.

Su Z., Zhu G., Chen X., & Yang Y. (2016). Sparse envelope model: Efficient estimation and response variable selection in multivariate linear regression. *Biometrika*, *103*(3), 579–593. https://doi.org/10.1093/biomet/asw036

https://doi.org/10.1093/jrsssc/qlad068
Advance access publication 26 July 2023