



HAL
open science

QUELLES DONNÉES COLLECTER POUR UN BON MODÈLE D'EXPECTED GOAL AU HANDBALL ?

Alexis Mortelier, François Rioult, John Komar

► **To cite this version:**

Alexis Mortelier, François Rioult, John Komar. QUELLES DONNÉES COLLECTER POUR UN BON MODÈLE D'EXPECTED GOAL AU HANDBALL ?. Symposium MaDICS 2023, May 2023, Lyon, France. hal-04538449

HAL Id: hal-04538449

<https://hal.science/hal-04538449>

Submitted on 9 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

QUELLES DONNÉES COLLECTER POUR UN BON MODÈLE D'EXPECTED GOAL AU HANDBALL ?

Alexis Mortelier¹, François Rioult¹, John Komar²

¹Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, FRANCE

²National Institute of Education, Nanyang Technological University, Singapore, Singapore

alexis.mortelier@unicaen.fr, francois.rioult@unicaen.fr, john.komar@nie.edu.sg



1. La métrique Expected Goal (xG)

3. Méthode

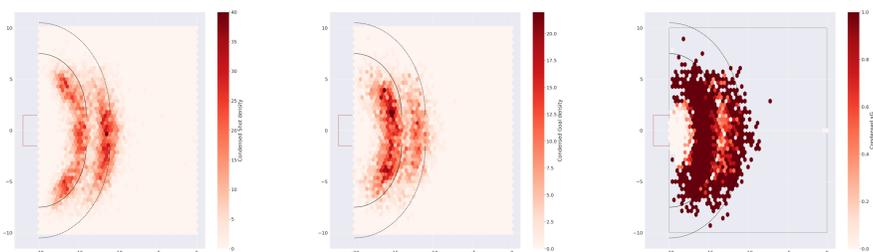


Figure 1: Densité des tirs (gauche), densité des buts (milieu) et xG (droite) au handball

Motivations

- modèles introduits en 2012 par Sam Green pour le site Opta
- xG : pour une position, indique la probabilité de transformer un tir en but
- Figure 1 : carte de chaleur d'xG pour les matchs du championnat du monde 2015 de handball masculin

Objectifs : déterminer les indices pertinents pour le calcul de la métrique xG.

2. Données



Tournoi : Championnat du monde de handball 2015

Lieu : Qatar

Acquisition : Vidéo

Type : Professionnel

Dataset : 88 matchs

Apprentissage automatique d'un modèle

- chaque expert a sa propre fonction xG dont l'expression exacte est inconnue → apprentissage automatique sur les données de la figure 1

Méthode wrapper de sélection d'attributs

- sélection des indices selon la performance de l'algorithme d'apprentissage
- wrapper effectue une recherche dans l'espace des sous-ensembles d'indices

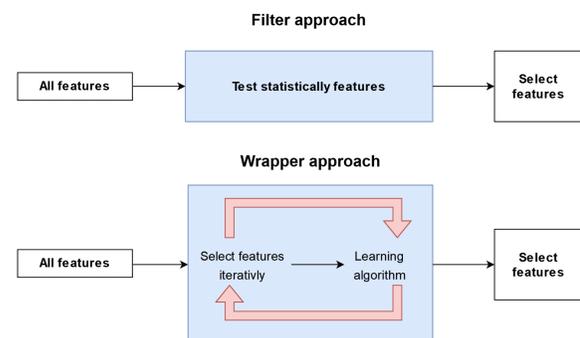


Figure 3: Approche wrapper

Dans la pratique, nous ne cherchons pas la configuration d'optimale d'indices. Nous nous contentons de comparer plusieurs combinaisons emblématiques :

pos : ne considère que la position du tireur

$(pos + \theta + d)$: ajoute l'angle d'ouverture θ du but et la distance au but d

données événementielles : données de position plus les informations sur les événements du match (temps écoulé depuis le tir précédent, différentiel de but, scores, etc.)

données spatiales : caractéristiques géométriques de la situation de jeu au moment du tir (position du gardien, vitesse du tireur, nombre de défenseurs, etc.)

données combinées : considèrent l'intégralité des indices

Résultats

input	accuracy	precision	recall	F1 score	AUC	learning time (s)
<i>pos</i>	0.72	0.77	0.74	0.75	0.71	258s
$pos + \theta + d$	0.73	0.78	0.74	0.76	0.72	+62%
données événementielles	0.74	0.79	0.76	0.77	0.73	+75%
données spatiales	0.75	0.77	0.78	0.78	0.75	+54%
données combinées	0.75	0.77	0.78	0.78	0.75	+67%

Table 1: Tableau de performance des modèles xG avec RANDOMFOREST

- il n'est pas nécessaire d'avoir un jeu de données complexe
- la position du tireur suffit pour obtenir un modèle d'xG obtenant des bons scores d'entraînement
- l'enrichissement des données d'entrée demande un coût trop élevé pour un faible gain de performance du modèle
- au handball, les événements de buts surviennent plus régulièrement dans une zone très petite par rapport au football, ce qui justifie que l'ajout d'indice influence peu la performance du modèle.