Hoofdstuk 5

Van echte naar virtuele loopjes

> "Kan je de computer leren een gevaarlijkspelsituatie te herkennen?" SciSports is een bedrijf dat voetbaldata analyseert. Mede-oprichter Anatoliy Babic vroeg de Studiegroep Wiskunde met de Industrie om mee te denken over de analyse van positiedata van spelers.

Hij was nog onbekend, Wout Weghorst, toen zijn contract bij FC Emmen afliep in 2014. Voetbaldatabedrijf SciSports gaf Heracles destijds een aankoopadvies over de speler. Anatoliy Babic, medeoprichter van SciSports vertelt in zijn beginpresentatie tijdens de Studiegroep Wiskunde met de Industrie dat Weghorst in die tijd werd onderschat. "Zijn statistieken waren opvallend goed, zo zagen wij, vandaar dat wij een aankoopadvies gaven." De inschatting bleek juist, Weghorst werd topscoorder en later gekocht door AZ. Het is een van de succesverhalen die het jonge bedrijf SciSports graag vertelt. Dit laat namelijk zien dat het analytisch bestuderen van voetbaldata duidelijk toegevoegde waarde heeft.

Bij SciSports draait alles om data. Dat kunnen data zijn over de acties van de spelers in het veld: aangekomen passes, succesvolle dribbels, schoten op doel, en zo meer. De volgende stap zijn positiedata van alle spelers en de bal. Babic: "Van de topspelers zijn veel data beschikbaar, ook positiedata, maar van de semi-professionals veel minder. Dat is lastig, want wij willen bijvoorbeeld graag aankoopadvies geven over aankomend talent." Het bedrijf werkt daarom zowel aan methoden om data te verwerven als aan het analyseren van beschikbare data.

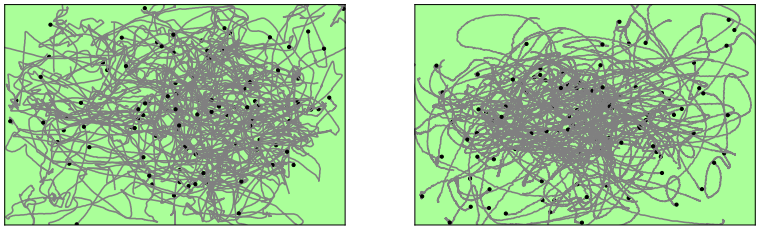
Op de lange termijn wil SciSports toe naar 3D-data van spelers: waar zijn ze op het veld, hoe staat hun lichaam, hoe bewegen ze. Op basis van al die gegevens kan dan een FIFA-achtige simulaties worden gemaakt, en kan je het spel vanuit alle posities bekijken. Om te beginnen wil het bedrijf aan de slag met 2D-positiedata van de spelers en de bal. Babic kwam naar de Studiegroep Wiskunde met de Industrie met een open vraag: willen jullie meedenken over wat er met die data mogelijk is, en wat niet? Hij wil graag weten of het op basis van 2D-tracking data mogelijk is om de kwaliteiten

van een speler in te schatten - zonder extra kennis uit voorgaande wedstrijden. Kun je iets zeggen over zijn versnellingsvermogen, zijn conditie, misschien zelfs iets over speelstijl? En nog een stapje verder, is het mogelijk om ook het spel zelf te analyseren? Zou je met software bijvoorbeeld gevaarlijke spelsituaties kunnen herkennen? Babic is zelf wiskundige, een jaar eerder deed hij nog als deelnemer mee aan de Studiegroep Wiskunde met de Industrie. De wiskundige inzichten uit de studiegroep gebruikt hij om de koers voor de onderzoeksafdeling van het bedrijf verder uit te zetten.

"Doordat de vraag zo breed was, duurde het even voor we hadden besloten wat we gingen doen", zo vertelt wiskundige Clara Stegehuis van de Technische Universiteit Eindhoven. "We wilden iets doen wat in een week te doen was, én wat meerwaarde had voor SciSports." De groep begon met de analyse van de bewegingen van één speler. Stegehuis: "Kunnen we vaststellen wat typische looppatronen zijn, kunnen we de bewegingen modelleren? Dat was de vraag waarmee we begonnen. Vanuit die basis wilden we dan uitbouwen naar alle spelers, en uiteindelijk ook naar de interactie tussen de spelers."

De groep besloot gebruik te maken van machine learning voor de analyse van de loopbewegingen. "Je voert de computer dan een hele hoop data, en die probeert zich patronen in die data eigen te maken", zo vat Stegehuis machine learning snel samen. "Door de computer loopbewegingen van een speler te geven, kan hij uiteindelijk zelf looppatronen genereren die typisch zijn voor die speler." Babic: "Het leuke is dat de wiskundigen kwamen met een methode van machine learning, waarvan ik al een aantal keer had gehoord dat ze veelbelovend is, maar waar ik me nog niet in had verdiept. Het kwam er niet van. Nu weet ik dat deze methode inderdaad de moeite waard is om ons in te verdiepen."

De groep was groot genoeg om in tweeën op te splitsen; zo konden twee methoden van machine learning worden uitgeprobeerd op de 2D-loopdata van Sci-Sports. Stegehuis: "Beide methoden zijn ontwikkeld voor het herkennen van foto's." Stegehuis: "Je kan een loopbeweging ook zien als een plaatje. In plaats van foto's voerden wij de coör-



Figuur 5.1: Links honderd gemeten loopbewegingen, rechts honderd door de computer (VAE) gegenereerde loopbewegingen.

dinaten in van een loopbewegingen van 10 seconde." De eerste methode heet GAN, wat staat voor Generative Adversarial Network. "Daarbij gebruik je twee netwerken: het ene netwerk probeert zo goed mogelijk de speler te simuleren, de ander probeert onderscheid te maken tussen de echte speler en de speler die het eerste netwerk heeft gesimuleerd. De twee netwerken versterken elkaar, als twee vrienden die elkaar steeds te slim af willen zijn", aldus Stegehuis. GAN werkt uitstekend voor het genereren van nieuwe, vergelijkbare plaatjes op basis van bestaande plaatjes: zou het ook werken om loopbewegingen te genereren die bij een speler passen?

De andere methode was Variational Autoencoder, VAE. Dat is een methode die de data zoveel mogelijk comprimeert, om die informatie weer om te zetten naar een beeld. De VAE leert eerst om de beweging van een voetballer in de data zo efficiënt mogelijk te beschrijven. Door daarna deze beschrijvingen iets te variëren, kan de VAE ook nieuwe bewegingen van diezelfde voetballer genereren. Deze methode is wat eenvoudiger dan de eerste.

Met beide methoden produceerde de groep looppatronen. Babic: "Eigenlijk maak je zo een digitale variant van de speler." De tijd om de netwerken te trainen was kort, dus de resultaten nog wat grof. Maar de resultaten leken daadwerkelijk op de bewegingen van een voetballer. "Vooral bij de tweede methode", aldus Stegehuis. Dat de tweede methode, VAE, beter werkte dan de eerste, GAN, wijt ze vooral aan de korte tijd die er was om de netwerken te trainen. "Dat trainen kost veel rekentijd, en we hadden tijdens de week maar twee nachten tijd." De tweede methode leert het snelst, maar Stegehuis verwacht van de eerste methode uiteindelijk betere resultaten.

Team SciSport

Anatoliy Babic (Sci-Sports), Harshit Bansal (Technische Universiteit Eindhoven)[1], Gianluca Finocchio (Universiteit Leiden)[2], Julian Golak (Universiteit Maastricht)[3], Mark Peletier[1], Jim Portegies[1], Clara Stegehuis[1], Anuj Tyagi[1], Roland Vincze[3], William Weimin Yoo [2]