

Wykorzystanie sztucznych sieci neuronowych do prognozowania wyników meczów piłkarskich

1. Prognozowanie

Prognozowanie, czyli przewidywanie przyszłych zdarzeń stało się jednym z elementów życia człowieka. Prowadzimy prognozy meteorologiczne, gospodarcze i wiele innych, próbujemy przewidywać przyszłość praktycznie w każdej dziedzinie życia. Jedne prognozy są trafne, inne zupełnie mijają się z rzeczywistością. Wspólną cechą prognozowania jest przetwarzanie informacji i próba znalezienia elementów, które determinują przyszłe zdarzenia. W tym miejscu warto zadać pytanie, czy wszystkie zjawiska, wszystkie procesy są w pełni zeterminowane czy też nie? Czy stopień determinacji zależy od perspektywy czasowej? Najczęściej im bliżej chwili, dla której chcemy coś przewidzieć, tym więcej mamy informacji, a więc łatwiej jest znaleźć elementy determinujące. Ale tu kolejne pytanie: jaką wagę mają poszczególne determinanty? Można zadać jeszcze wiele innych pytań, dlatego od dawna próbowano wykorzystać w procesach prognozowania narzędzia matematyczne (głównie statystyczne), w większości z nich stosowano metody do opisu niepewności: regresywne i autoregresywne analizy, prawdopodobieństwo subiektywne w połączeniu z łańcuchami Markowa i metodami Monte-Carlo [Nazarko 2004]. W ostatnich czasach coraz częściej do prognozowania wykorzystuje się metody z grupy sztucznej inteligencji: sztuczne sieci neuronowe, zbiory rozmyte czy algorytmy genetyczne. Niezależnie od zastosowanej metody prognozowanie sprawdza się do znalezienia:

- przekształcenia pozwalającego wyznaczyć interesujące nas właściwości z przyszłości na podstawie dostępnych danych,
- zbioru elementów determinujących – mających wpływ na przyszłość. Zbiór ten nie musi być określony jawnie, wystarczy, że znajdziemy inne elementy, których kombinacja dostarcza potrzebnych nam informacji.

Można stwierdzić, że prognozowanie w ogólnym przypadku to nie tylko poszukiwanie związków zależności, które można traktować jako przekształcenie, ale również poszukiwanie argumentów dla tego przekształcenia, co szczególnie przy dużej liczbie danych wejściowych powoduje, iż wygodnym narzędziem prognostycznym mogą być sztuczne sieci neuronowe.

2. Prognozowanie wyników meczów piłkarskich

Przewidywanie wyników meczów piłki nożnej może być interesujące pod dwoma względami:

- wykazania przydatności różnych metod prognozowania, w tym najchętniej stosowanych metod matematycznych,
- możliwości uzyskania korzyści finansowych.

2.1. Specyfika problemu

Dla wielu fanów oraz specjalistów, zajmujących się prognozą wyników, matematyka jest podstawowym narzędziem wykorzystywanym przy podejmowaniu decyzji związanych z predykcją. To właśnie liczby przedstawiają wyniki poprzednich spotkań, osiągnięć drużyny czy też pojedynczego zawodnika. Stanowi to pewny materiał statystyczny, na podstawie którego dokonywane są analizy. Warto również podkreślić, iż przy wykorzystaniu metod matematycznych opartych tylko na danych statystycznych pozbywamy się elementu stroniczości w przewidywaniu. Jednak niezależnie od metody uzyskiwane predykcje nigdy nie będą w stu procentach poprawne.

Eksperti piłkarscy oraz kibice często dokonują prognozy rezultatu na podstawie prostych założeń oraz zdrowego rozsądku. Jeżeli pierwszy zespół wygrał wszystkie swoje poprzednie mecze i drugi zespół przegrał wszystkie wcześniejsze oraz pierwszy zespół wygrał poprzednie mecze z zespołem drugim, to należy się spodziewać, że bezpośredni pojedynek wygra zespół pierwszy. Jednak nawet przy takim założeniu nie można stwierdzić, że pierwszy zespół ma sto procent szans na wygraną. Zawsze jest możliwość, że wynik będzie remisowy lub zwycięstwo odniesie drużyna teoretycznie słabsza. Nawet niewielkie czynniki mogą wpłynąć na końcowy wynik. Szczególnie jest to zauważalne właśnie w sporcie, gdzie szczęście jest nieuniknione. Często zdarza się, że nierozsądnie „zarobiona” czerwona kartka, kontuzja na rozgrzewce przedmeczowej przechylają szalę zwycięstwa na drugą stronę. Pomimo występowania zdarzeń losowych obstawianie może być bardzo efektywne. Prognozowanie rezultatów meczów ma bardzo szeroki zakres. Możemy typować zwycięzcę pojedynku, dokładny wynik, margines wygranej oraz nawet bardzo szczegółowe zdarzenia, który zawodnik strzeli bramkę lub w jakim przedziale czasowym zostanie ona zdobyta.

2.2. Wybrane metody

Prostą metodę prognozowania wyników meczów piłkarskich opisuje Yue Weng Mak w swojej pracy *Prediction on Soccer Matches using Multi-Layer Perceptron* [Yue Weng Mak 2010]. Do otrzymania prognozy wykorzystuje pięć ostatnich meczów drużyny gości oraz gospodarzy, a także trzy bezpośrednie pojedynki pomiędzy drużynami. Na podstawie zgromadzonych danych przy-

znawane są punkty za mecz wygrany (1 punkt), za porażkę (-1) oraz za remis (0). Następnie punkty przemnażane są przez wagi – ostatnie spotkanie otrzymuje wagę 2,5, poprzednie 2 i tak kolejno aż do najstarszego spotkania, które otrzymuje wagę 0,5. Tak ważone wartości są sumowane dla obu drużyn. Oddzielnie sumowane są wartości za bezpośrednie mecze pomiędzy drużynami. W ten sposób uzyskujemy trzy wartości, które po przemnożeniu przez kolejne wagi (1 dla gospodarzy, -1 dla gości, 1 dla bezpośrednich spotkań) sumujemy. Jeżeli uzyskana suma jest znacznie większa od 0, typujemy zwycięstwo gospodarzy, jeżeli znacznie mniejsza od 0, zwycięstwo gości, w pozostałych przypadkach typujemy remis. Metoda osiąga skuteczność ok. 44%. Największą skuteczność, jaką udało się uzyskać, to 80% w danej kolejce, zaś najniższa to 20%.

Portal sportowy 1x2monster.com proponuje metodę, która umożliwi prognozowanie dokładnego bramkowego wyniku meczu. Metoda wykorzystuje wyniki obu przeciwników (drużyny gospodarzy oraz drużyny gości) z pięciu ostatnich meczów (ostatnie pięć pojedynków domowych gospodarzy oraz ostatnie pięć meczów wyjazdowych gości). Na podstawie tych wyników algorytm wyznacza: sumę bramek zdobytych przez gospodarzy z bramkami straconymi przez gości (HGP), sumę bramek zdobytych przez gości z bramkami straconymi przez gospodarzy (AGP). Różnica pomiędzy wyznaczonymi sumami (HGP – AGP) pozwala wystawić prognozę wyniku:

- $(HGP - AGP) \geq 2$ zwycięstwo gospodarzy,
- $(HGP - AGP) \leq -2$ zwycięstwo gości,
- $(HGP - AGP) \in (-2, 2)$ remis.

W celu uzyskania dokładnego wyniku bramkowego należy skorzystać z dodatkowych tabel (rys.1).

Gospodarze		Goście			
		≥ 0	≥ 11	≥ 17	≥ 26
≥ 49	6	0	1	2	3
≥ 43	5	0	1	2	3
≥ 37	4	0	1	2	3
≥ 30	3	0	1	2	
≥ 23	2	0	1		
≥ 0	1	0			

Gospodarze				Goście	
≥ 23	≥ 18	≥ 13	≥ 0		
3	2	1	0	5	≥ 34
3	2	1	0	4	≥ 29
	2	1	0	3	≥ 24
		1	0	2	≥ 18
			0	1	≥ 0

Gospodarze + Goście		
≥ 44	3	3
≥ 34	2	2
≥ 24	1	1
≥ 0	0	0

Rys. 1. Tabele przedstawiające prognozę wyniku meczu: a) wygrana gospodarzy, b) wygrana gości, c) remis

Źródło: 1x2monster.com

Metoda była testowana na angielskiej lidze Premiership. Na przestrzeni 21 kolejek uzyskano skuteczność 50,5%. Największa skuteczność, jaką udało się uzyskać w pojedynczej kolejce, to 80%, najniższa zaś to 30%.

3. Prognozowanie wyników meczów z wykorzystaniem Sztucznych Sieci Neuronowych

Przedstawione w części 2 metody prognozowania wyników meczów arbitralnie określały wartości progów decyzyjnych. Wykorzystanie do prognozowania Sztucznych Sieci Neuronowych dzięki ich zdolności do uczenia się uwalnia nas od tego problemu. Pozostaje kwestia doboru danych uczących. Zdecydowano, iż jako dane wejściowe do systemu prognozującego wyniki meczów piłkarski zostaną wykorzystane:

- rezultaty pięciu ostatnich meczów drużyny gospodarzy,
- wyniki pięciu ostatnich meczów drużyny gości,
- wyniki trzech ostatnich bezpośrednich pojedynków obydwu drużyn.

Decyzja o wyborze takiego zakresu danych wejściowych podyktowana jest następującymi przesłankami: wyniki ostatnich meczów drużyn dają obraz o jej aktualnej formie – co bezsprzecznie ma wpływ na uzyskiwane przez drużynę rezultaty. Należy podkreślić, iż brane pod uwagę są tylko wyniki meczów ligowych (co eliminuje możliwość, iż któraś z drużyn ostatnie mecze rozgrywała z przeciwnikami innej klasy lub w mocno zmienionym składzie). Wyniki bezpośrednich spotkań pomiędzy drużynami obrazują zjawisko często określane mianem, iż „drużyna leży innej lub też jej nie leży”. Ponadto przy doborze danych uczących wykorzystano doświadczenia opisane w innych pracach [Bartman 2013: 358–365].

Zanim sieć neuronową wykorzystamy do prognozowania, musi zostać ona odpowiednio zaprojektowana oraz nauczona. Projekt sieci jest częściowo zeterminowany parametrami danych uczących, jednak pewne jej atrybuty są dobierane metodą prób i błędów. Analiza literatury [Tadeusiewicz 1993; Masters 1996] pozwala stwierdzić, iż:

- sieć powinna składać się z trzech warstw (wejściowa, ukryta, wyjściowa),
- warstwa wejściowa winna być zbudowana z 13 neuronów (dokładnie tyle danych jest w wektorze wejściowym),
- w warstwie wyjściowej konieczny jest tylko 1 neuron (oczekujemy odpowiedzi typu zwycięstwo, remis, porażka).

Precyzyjne określenie liczby neuronów warstwy ukrytej metodami analitycznymi jest niemożliwe, można jedynie dokonać pewnych jej oszacowań [Masters 1996]. Dlatego też liczbę neuronów w warstwie ukrytej ustalono metodą prób i błędów.

Na potrzeby SSN dane wejściowe muszą być odpowiednio zakodowane, i tak zwycięstwo zakodowano wartością 1, remis wartością 0, a porażkę (–1). W zależności od odległości czasowej meczu odbytego do meczu, którego wynik prognozujemy, dane wejściowe powinny być odpowiednio ważone, im bliżej do prognozowanego meczu, tym dane mają większą wagę.

Dane do budowy systemu prognozowania wyników meczów piłkarskich pozyskano z serwisu internetowego www.livesports.pl.

3.1. Wyniki prognoz

Prognozy przeprowadzono dla angielskiej Premiership, przewidując wyniki 10 kolejek spotkań. Badania prognostyczne były prowadzone równoległe z badaniem wpływu wybranych parametrów architektury sieci na efektywność jej pracy. Badano wpływ tych parametrów sieci, których nie można określić analitycznie (dobiera się je metodą prób i błędów, bazując na wiedzy eksperckiej lub pewnych wstępnych oszacowaniach). W ten sposób ustalano:

- liczbę neuronów w warstwie ukrytej,
- funkcje przejścia dla neuronów z warstwy ukrytej,
- funkcje przejścia dla neuronów warstwy wyjściowej,
- końcową wartość funkcji celu.

Wyniki badań zestawiono w tabeli 1.

Tabela 1

Zestawienie jakości prognoz dla różnych parametrów SSN

Liczba neuronów	Funkcja przejścia neuronów wyj.	Efektywność uczenia [%]	Efektywność		
			minimalna [%]	średnia [%]	maksymalna [%]
4	Liniowa	65,0	45,6	55,2	64,6
	Tangensoidalna	76,0	54,5	60,3	66,7
6	Liniowa	98,0	46,5	54,0	66,7
	Tangensoidalna	85,5	52,5	60,5	68,7
8	Liniowa	100	44,4	54,1	62,6
	tangensoidalna	90,0	51,5	59,9	65,6
10	Liniowa	100	48,5	54,7	63,6
	tangensoidalna	92,6	50,6	60,1	66,7
12	Liniowa	100	47,5	54,8	62,8
	Tangensoidalna	91,7	53,5	59,5	66,7

Analiza danych zebranych w tabeli 1 pozwala ustalić brakujące parametry sieci neuronowej: liczbę neuronów w warstwie ukrytej na 4–10, funkcje przejścia neuronów w warstwie wyjściowej na tangensoidalną. Dodatkowe badania wykazały, iż warstwa ukryta powinna zawierać neurony z tangensoidalną lub sigmoidalną funkcją przejścia. Ustalono również, iż zmniejszanie oczekiwanej wartości funkcji celu poniżej 0,01 nie poprawia efektywności pracy sieci, a jedynie wydłuża czas uczenia. Należy jednak zaznaczyć, iż dla krótszych warstw ukrytych uczenie sieci było mniej efektywne (patrz tabela 1).

W celu poprawy jakości prognoz rozszerzono dane wejściowe o pozycje drużyn w tabeli. Okazało się, iż nie wpłynęło to na jakość prognoz.

Dla poszczególnych kolejek skuteczność prognoz mieści się w przedziale od 30% do 80%.

4. Podsumowanie

Analiza prognoz metodami przedstawionymi w pracy pozwala zauważyć, iż:

- średnia jakość prognoz wynosi od 44% do 61%,
- najlepszą skuteczność prognoz uzyskała SSN,
- istnieją duże rozbieżności w skuteczności prognoz dla poszczególnych kolejek spotkań (od 20% do 80%).

Przeprowadzone badania wykazują, iż Sztuczna Sieć Neuronowa może być wykorzystana do prognozowania wyników meczów. W celu poprawy efektywności konieczne są jednak dalsze prace nad tą metodą. Wydaje się, iż możliwości poprawy skuteczności prognoz należy szukać w rozszerzeniu zakresu danych wejściowych o dodatkowe elementy: np. absencje zawodników, częstotliwość rozgrywania spotkań, można również rozważyć wyeliminowanie ze zbioru uczącego wyników sensacyjnych.

Literatura

- Bajda K. (2014), *Wykorzystanie sztucznych sieci neuronowych do prognozowania wyników sportowych – projekt systemu*, praca inżynierska, Rzeszów.
- Bartman J. (2009), *Regula PID uczenia sztucznych neuronów*, „Metody Informatyki Stosowanej”, 3/2009.
- Bartman J. (2013), *Wpływ opisu danych na efektywność uczenia oraz pracy sztucznej sieci neuronowej na przykładzie identyfikacji białek*, „Edukacja – Technika – Informatyka”, 4/2013-2.
- Masters T. (1996), *Sieci neuronowe w praktyce: programowanie w języku C++*, Warszawa.
- Nazarko J. (2004), *Wprowadzenie do metodyki prognozowania*, Białystok.
- Tadeusiewicz R. (1993), *Sieci neuronowe*, Warszawa.
- Yue Weng Mak (2010), *Prediction on Soccer Matches using Multi-Layer Perceptron*.
www.1x2monster.com

Streszczenie

W pracy zaprezentowano koncepcję wykorzystania Sztucznych Sieci Neuronowych do prognozowania wyników meczów. Przedstawiono architekturę sieci oraz skuteczność realizowanych przez nią prognoz. Uzyskane wyniki zestawiono z wynikami otrzymanymi przy wykorzystaniu innych metod.

Słowa kluczowe: Sztuczne Sieci Neuronowe, prognozowania.

Using artificial neural networks to predict the results of football matches

Abstract

The paper presents the concept of using Artificial Neural Networks to predict the results of football matches. Authors presents the architecture of the network and the effectiveness of the implementation by the forecasts. The results were compared with results obtained using other methods.

Key words: Artificial Neural Networks, prediction.