A soccer ball with white and grey panels is positioned on a green grassy field. The background is a blurred landscape with trees and a bright sky, suggesting an outdoor setting. The text is overlaid on the right side of the image.

Prezentacja tematu pracy dyplomowej

Maciej Luberadzki
3 czerwca 2020 roku

Temat pracy

**Projekt i implementacja systemu
opartego na sztucznej sieci neuronowej,
umożliwiającego predykcję wyników
spotkań piłkarskich**

Promotor: dr hab. inż. Marcin Pluciński

Cel pracy

Celem pracy jest zaprojektowanie i implementacja systemu umożliwiającego predykcję wyników spotkań piłkarskich, bazującego na sztucznych sieciach neuronowych.

Oprogramowanie będzie umożliwiała gromadzenie i przygotowanie danych, uczenie sieci i testowanie jakości predykcji.

Zakres pracy

1. Przygotowanie danych historycznych o wynikach spotkań piłkarskich.
2. Synteza możliwych atrybutów wejściowych dla sieci neuronowej oraz analiza ich istotności.
3. Wybór typu sieci i jej implementacja.
4. Optymalizacja struktury sieci i jej uczenie.
5. Testowanie jakości systemu.
6. Wnioski

Wprowadzenie

Predykcja rezultatów meczów piłki nożnej to powszechnie znany problem, który dostarcza wielu emocji, zarówno osobom, które wykazują zainteresowanie meczami piłkarskimi tylko przy okazji dużych wydarzeń piłkarskich, jak i zagorzałym kibicom. Czy to w postaci zakładu „o przekonanie”, czy też zakładu u bukmachera o konkretne kwoty pieniężne, oceniamy jakość swoich przewidywań, najczęściej opartych na własnej intuicji. Często jednak nasze przewidywania okazują się wielu przyczyn błędne, stąd pomysł, aby usystematyzować sposób predykcji wyników meczów piłkarskich, przy pomocy opisów matematycznych i mocy obliczeniowej komputera.

Przygotowanie danych historycznych o wynikach spotkań piłkarskich

Krokiem podstawowym dla przebiegu pracy, będzie zebranie odpowiedniego zbioru danych, przy pomocy którego, będę mógł przeprowadzić efektywne uczenie sieci neuronowej oraz testowanie jej działania. Znalezienie, bądź skomponowanie własnego zbioru wciąż jeszcze jest przede mną. Najchętniej wykorzystałbym dane z meczów polskiej Ekstraklasy, jednak ze względu na ogólną opinię o niestabilności wyników, możliwe, że zdecyduję się na wykorzystanie statystyk z innych rozgrywek. Niemniej jednak, największy wpływ na postać mojego zbioru danych będzie miało to, do jakich zbiorów uda mi się dotrzeć, oraz jak „bogate” będą.

Statystyki z meczu Pogoń Szczecin – Zagłębie Lubin (29.05.2020 r.), zakończonego wynikiem 0:3 (źródło: www.sofascore.com)



	VS	
56%	Ball possession	44%
15	Total shots	10
4	Shots on target	6
11	Shots off target	4
10	Corner kicks	3
10	Fouls	18
0	Yellow cards	4
0	Red cards	0
3	Goalkeeper saves	4

Synteza możliwych atrybutów wejściowych dla sieci neuronowej oraz analiza ich istotności

Dobór odpowiednich atrybutów wejściowych będzie najprawdopodobniej kluczowym zadaniem, decydującym o jakości całego systemu. Realizację tego kroku planuję oprzeć na analizie istniejących rozwiązań, własnych eksperymentach, a także w miarę możliwości na własnych doświadczeniach. Jedno z rozwiązań, na które się natknąłem wstępnie analizując literaturę, zostało zawarte w artykule Huanga i in. z 2010 roku [1], w którym autorzy, w oparciu o własne doświadczenie oraz intuicję, jako atrybuty wejściowe przyjęli statystyki zawarte na kolejnym slajdzie. Jeśli chodzi o moje własne doświadczenia, obserwując pomeczowe statystyki, odniosłem wrażenie, że drużyna, której piłkarze przebiegli łącznie większy dystans, wygrywała, stąd na pewno nie omieszkałbym sprawdzić, czy taka zależność rzeczywiście zachodzi.

Synteza możliwych atrybutów wejściowych dla sieci neuronowej oraz analiza ich istotności

Statystyki meczów Mistrzostw Świata z 2006 roku, które Huang i in. [1] przyjęli jako atrybuty wejściowe sieci neuronowej:

1. Zdobyte bramki.
2. Oddane strzały.
3. Oddane strzały w światło bramki.
4. Liczba rzutów różnych dla drużyny.
5. Rzuty wolne z bezpośrednim strzałem na bramkę.
6. Rzuty wolne wykonane niebezpośrednio.
7. Posiadanie piłki.
8. Faule drużyny przeciwnej.

Wybór typu sieci i jej implementacja oraz optymalizacja struktury sieci i jej uczenie

Na tym etapie mam nadzieję zaimplementować co najmniej dwa rodzaje sieci neuronowych, aby móc porównać ich działanie i dokonać wyboru, która z nich przynosi lepsze efekty. Prawdopodobnie jedną z nich będzie sieć uczona z wykorzystaniem algorytmu wstecznej propagacji błędów, który polega na korygowaniu wag sieci, poprzez przesyłanie błędów otrzymanego na wyjściu sieci do warstw poprzedzających warstwę wyjściową, aż do warstwy wejściowej i modyfikowanie wag na jego podstawie. Jednak ostateczny wybór, które rodzaje sieci neuronowych zdecyduję się zaimplementować i przebadać wciąż pozostaje przede mną.

Wybór typu sieci i jej implementacja oraz optymalizacja struktury sieci i jej uczenie

W przytoczonym wcześniej artykule Huanga i in. [1], autorzy zdecydowali się na zastosowanie właśnie sieci trójwarstwowej uczonej algorytmem wstecznej propagacji błędów, o następującej strukturze:

- 8 neuronów w warstwie wejściowej,
- 11 neuronów w warstwie ukrytej,
- 1 neuron w warstwie wyjściowej.

Autorzy podają, że taka konfiguracja, wraz ze wspomnianym wcześniej dobozem atrybutów, osiągnęła 76,9% dokładność, po wyeliminowaniu meczów, które zakończyły się remisem. Wizualizacja znajduje się na kolejnym slajdzie.

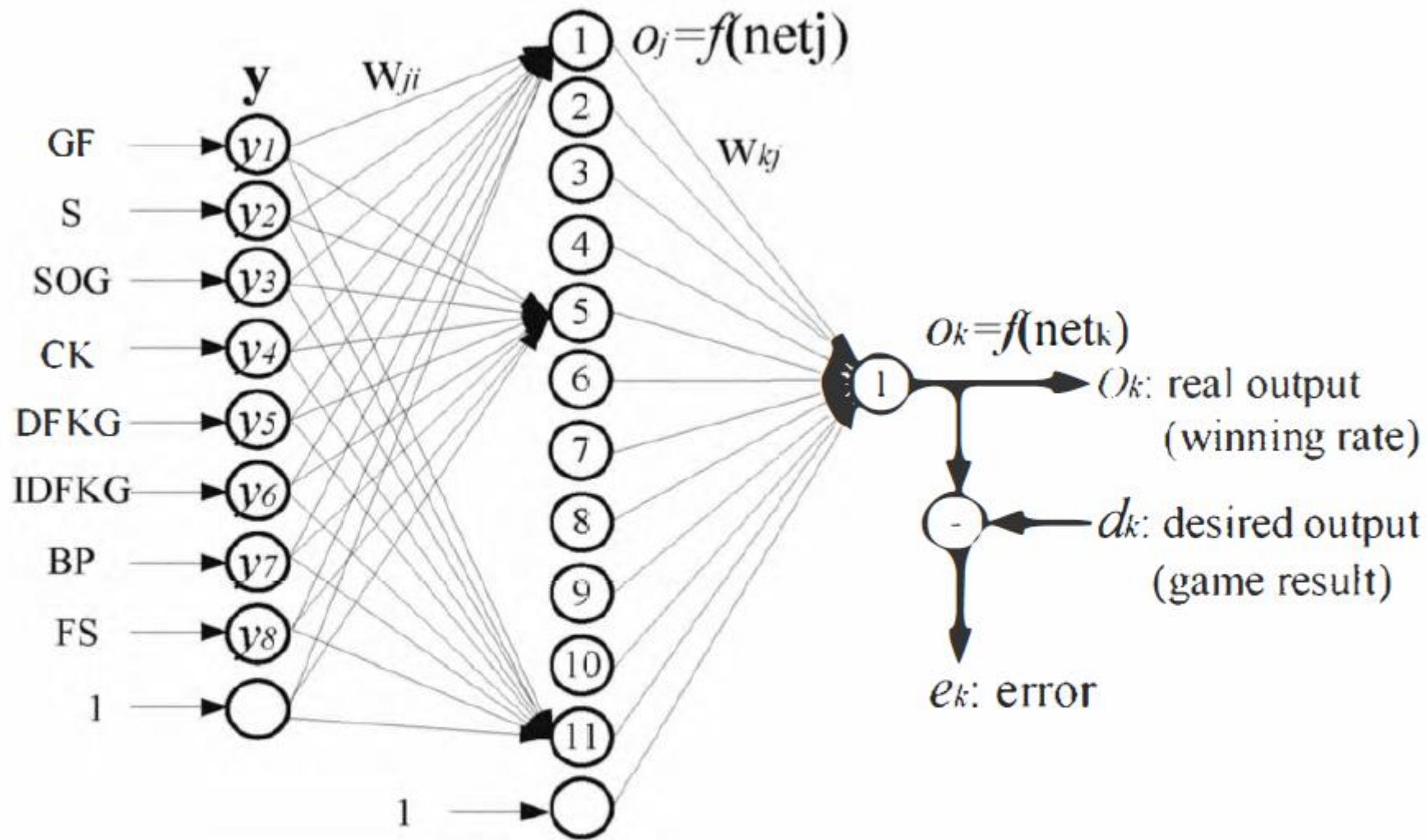


Fig. 3. MLP network architecture used in predicting the winning rate of 2006 WCFG.

Wybór typu sieci i jej implementacja oraz optymalizacja struktury sieci i jej uczenie

W kwestii optymalizacji struktury sieci neuronowej, ciekawym podejściem jest zastosowanie algorytmu przycinającego i optymalizującego strukturę sieci, przykładowo algorytmu Brain Damage Optimization. Takie rozwiązanie w dziedzinie sieci neuronowych zastosował Yusiong [2] w artykule z 2013 roku, w którym przedstawił nowy algorytm uczenia sieci neuronowej z wykorzystaniem Cat Swarm Optimization oraz wspomnianego algorytmu do optymalizacji struktury sieci. Sądzę, że zastosowanie takiego, bądź podobnego algorytmu przycinającego sieć, może być ciekawym rozszerzeniem, a także sporym wyzwaniem.

Testowanie jakości systemu

Podobnie, jak poprzednie punkty, testowanie jakości systemu jest tematem, który będę musiał zgłębić, natomiast mam nadzieję także wykorzystać w odpowiednim momencie system, do predykcji dopiero nadchodzących meczów piłkarskich. Jednym z czynników, który można już teraz wymienić jako kryterium oceny systemu, jest dokładność predykcji.

A soccer ball with white and dark grey panels is positioned on a green grassy field. The background is a blurred landscape with trees and a bright sky, suggesting an outdoor setting.

Dziękuję za uwagę

Ewentualne pytania proszę kierować na adres e-mailowy: lm41429@zut.edu.pl

Bibliografia

1. Huang, K.Y. and Chang, W.L., 2010, July. A neural network method for prediction of 2006 world cup football game. In *The 2010 international joint conference on neural networks (IJCNN)* (pp. 1-8). IEEE.
2. Yusiong, J. P. Y., „Optimizing Artificial Neural Networks Using Cat Swarm Optimization Algorithm,” *International Journal of Intelligent Systems And Applications*, vol. 01, 2013, pp. 69-80.