SPRAWOZDANIE

Zajęcia: Matematyka Konkretna

Prowadzący: prof. dr hab. Vasyl Martsenyuk

|  |  |
| --- | --- |
| Laboratorium Nr 8  Data 28.06.2025  Temat: „Algorytm LSTM dla tekstu”  Wariant 10 | Anna Więzik  Informatyka  II stopień, niestacjonarne,  2 semestr, gr.1a TTO |

1. Polecenie:

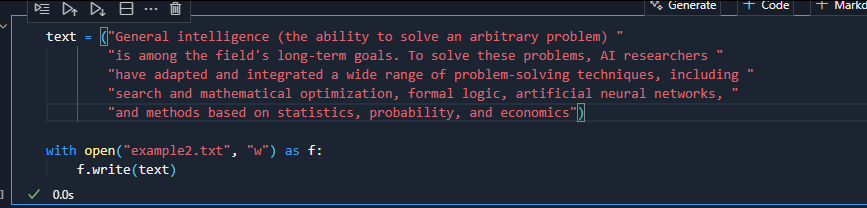
Link do repozytorium: <https://github.com/AnaShiro/MK_2025>

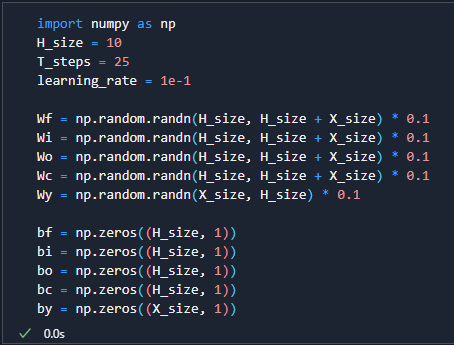
Celem zadania było stworzenie i zbadanie działania sieci LSTM (Long Short-Term Memory), uczonej z wykorzystaniem metody propagacji wstecznej w czasie (BPTT), w kontekście analizy sekwencji tekstowych. LSTM stanowi rozwinięcie klasycznych rekurencyjnych sieci neuronowych, zaprojektowane w celu skuteczniejszego radzenia sobie z problemem zanikającego gradientu, który utrudnia uczenie się długoterminowych zależności w tradycyjnych RNN. Dzięki mechanizmom bramek – wejściowej, wyjściowej i zapominania – LSTM potrafi selektywnie gromadzić oraz odrzucać informacje w pamięci wewnętrznej, co umożliwia jej zapamiętywanie kluczowych fragmentów sekwencji.

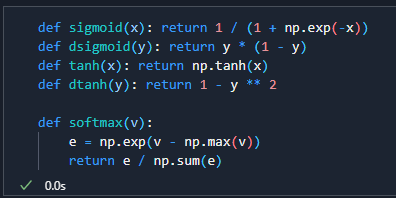
Proces uczenia tej sieci wymaga szczególnego podejścia do wyznaczania gradientów, ponieważ uczenie rozciąga się w czasie. Metoda BPTT pozwala śledzić błąd wzdłuż kolejnych kroków czasowych, by dokładnie ocenić wpływ każdego z nich na modyfikację wag. W trakcie pracy wykorzystano funkcje aktywacyjne sigmoid oraz tangens hiperboliczny i ich pochodne, co umożliwiło prawidłowe obliczenia sygnałów i gradientów. Wagi początkowe zainicjowano losowo z rozkładu normalnego, a dla bramek sigmoid zastosowano przesunięcie średniej, co pozwoliło na lepszą kontrolę początkowych aktywacji i ułatwiło start procesu nauki.

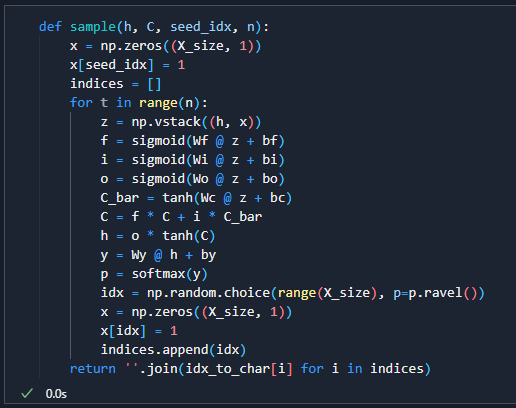
Dodatkowo przygotowano odpowiedni proces wstępnej obróbki danych tekstowych: przekształcono dane na zestaw unikalnych znaków (alfabet) i zakodowano je w sposób umożliwiający podanie do sieci i dekodowanie wyników. Dzięki temu możliwe było trenowanie modelu tak, by przewidywał kolejne znaki na podstawie poprzedniego kontekstu, systematycznie aktualizując parametry w celu minimalizacji funkcji kosztu.

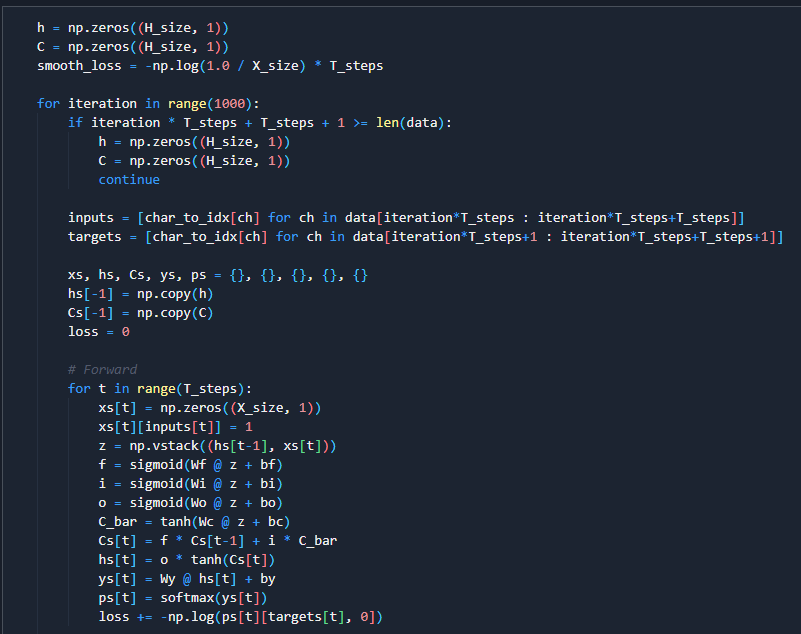
1. Opis programu opracowanego

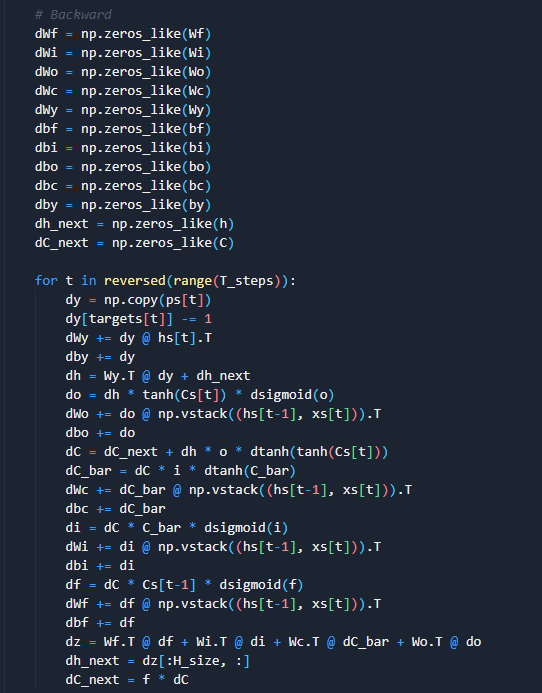


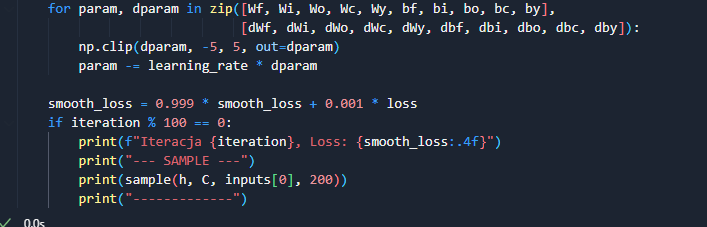


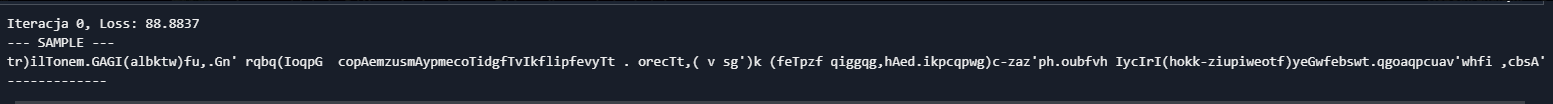












1. Wnioski

Zrealizowane badanie wykazało, że sieci LSTM znacząco usprawniają modelowanie zależności w danych sekwencyjnych, szczególnie w kontekście problemu zanikającego gradientu, który w tradycyjnych RNN skutecznie hamuje proces nauki. Dzięki zastosowanym bramkom model był w stanie świadomie przechowywać lub pomijać informacje, co pozytywnie wpłynęło na dokładność przewidywania kolejnych znaków w tekście, zwłaszcza w miarę postępu treningu. Kluczową rolę odegrała tu metoda BPTT, umożliwiająca przekazywanie informacji o błędach wstecz przez całą sekwencję, co zapewniło precyzyjną aktualizację wag, zarówno w krótkim, jak i długim horyzoncie czasowym.

Wyniki doświadczenia pokazały również, jak duże znaczenie ma odpowiednie zainicjalizowanie wag — dostosowane do charakteru używanych funkcji aktywacji — w kontekście szybkości zbieżności algorytmu. Połączenie architektury LSTM z metodą BPTT okazało się zatem skutecznym rozwiązaniem w konstruowaniu modeli zdolnych do analizy długich ciągów danych i wykrywania w nich subtelnych, długoterminowych relacji. Mimo złożoności implementacyjnej, rozwiązanie to sprawdza się znakomicie w zadaniach takich jak przetwarzanie języka naturalnego czy predykcja sekwencji.