

Piotr Nowakowski  
Michał Kućko

## **WEDT - Dokumentacja końcowa**

### **Temat projektu:**

Głęboka sieć neuronowa do wyznaczania podobieństwa semantycznego zdań w języku angielskim.

### **Definicja problemu:**

Projekt polegał na przekształceniu istniejącej architektury sieci neuronowej tak, by umożliwiała ona realizację zadania nr 2 konkursu SemEval z 2015 roku. Polegało ono na połączeniu odpowiadających sobie semantycznie fraz dwóch zdań oraz wyznaczeniu dla każdej pary fraz oceny podobieństwa w skali 0-5 wraz z jedną z poniższych etykiet:

- **EQUI** - semantyczna ekwiwalentność (w tym przypadku ocena zawsze wynosi 5),
- **OPPO** - semantyczne przeciwieństwo,
- **SPE1, SPE2** - podobne znaczenie, ale jedna z fraz oferuje dokładniejsze informacje (np. kilka jabłek - pięć jabłek),
- **SIMI** - frazy mają podobne znaczenia, lecz nie zachodzi **EQUI, OPPO, SPE1**, ani **SPE2**,
- **REL** - frazy mają powiązane znaczenia, lecz nie zachodzi **EQUI, OPPO, SPE1, SPE2**, ani **SIMI**,
- **NOALI** - brak powiązania z żadną inną frazą z drugiego zdania (wówczas ocena nie jest liczbą, lecz przyjmuje wartość **NIL**).

### **Przygotowanie danych wejściowych:**

Nasz system nie dopasowuje samodzielnie fraz z dwóch zdań, lecz pobiera docelowe dopasowania z pliku udostępnionego przez organizatorów konkursu. Dlatego pierwszym niezbędnym krokiem było stworzenie parsera przekształcającego owy plik na pliki zrozumiałe dla rozwiązania z pracy magisterskiej: pliki z dopasowanymi frazami dla zbioru uczącego i testowego, pliki z ocenami oraz pliki z etykietami.

Interesujące nas dane w pliku konkursowym były zapisane w następującym formacie:  
token-id-seq1 <==> token-id-seq2 // type // score // comment,

gdzie *token-id-seq1* i *token-id-seq2* to numery fraz odpowiednio w pierwszym i drugim zdaniu, *type* to etykieta dopasowania, *score* to ocena podobieństwa, zaś *comment* to przedstawienie obu fraz w języku naturalnym o postaci *phrase1* <=> *phrase2*.

Dane wyciągnęliśmy za pomocą wyrażenia regularnego następującej postaci:

$\wedge .+ // (.+) // (.+) // (.+) <==> (.+) \$$ ,

gdzie nawiasami zaznaczone są grupy przechwytyjące interesujące nas informacje, czyli etykietę i ocenę oraz obie frazy w języku angielskim.

### Zmiana architektury sieci:

Zaimplementowana sieć neuronowa posiadała 6 neuronów wyjściowych, które określały typ podobieństwa (**EQUI**, **OPPO** itp.). Należy zauważyć jednak, że typów podobieństwa jest siedem. Dlatego do oryginalnej architektury dodano jeden neuron, aby sieć mogła realizować tę klasyfikację.

Pierwotną koncepcją na zrealizowanie oceny podobieństwa w skali 0-5 było dołączenie do neuronu realizującego kompresję zdań drugiego zestawu neuronów warstwy softmax. Spowodowało to znaczące pogorszenie jakości klasyfikacji. Jego przyczyną jest pojawienie się "konfliktu interesów" pomiędzy wyjściami sieci - okazuje się że określenie typu wymaga innych wag w węzłach sieci (innego rodzaju kompresji) niż dokonanie oceny w skali liczbowej. Z tego powodu gradientowa metoda uczenia sieci znalazła rozwiązanie uśrednione, dające słabe wyniki zarówno w określaniu typu jak i oceny w skali liczbowej.

Dlatego w ostatecznym rozwiązaniu trenujemy dwie oddzielne sieci, aby wagi połączeń neuronów zostały oddzielnie zoptymalizowane pod przydzielone im zadania. Spowodowało to znaczny wzrost jakości klasyfikacji.

Warto również zauważyć, że ocena podobieństwa w skali wbrew pozorom również wymaga 7 neuronów wyjściowych, gdyż oprócz ocen 0-5 istnieje również ocena **NIL** zarezerwowana dla braku powiązania fraz (typ **NOALI**).

### Wyznaczanie miar dopasowania:

W istniejącym rozwiązaniu zaimplementowano obliczanie współczynnika korelacji Pearsona między wynikami sieci a oficjalnymi ocenami lub etykietami. Naszym ostatnim celem było zastąpienie owej miary metodą stosowaną w konkursie, czyli współczynnikiem F1:

$$F_1 = 2 \cdot \frac{1}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

*Precision* określa stosunek dobrze przyporządkowanych przez system fraz do wszystkich zwróconych przezeń przyporządkowań. Z kolei *recall* oznacza stosunek dobrze przyporządkowanych przez system fraz do wszystkich przyporządkowań w zbiorze testowym.

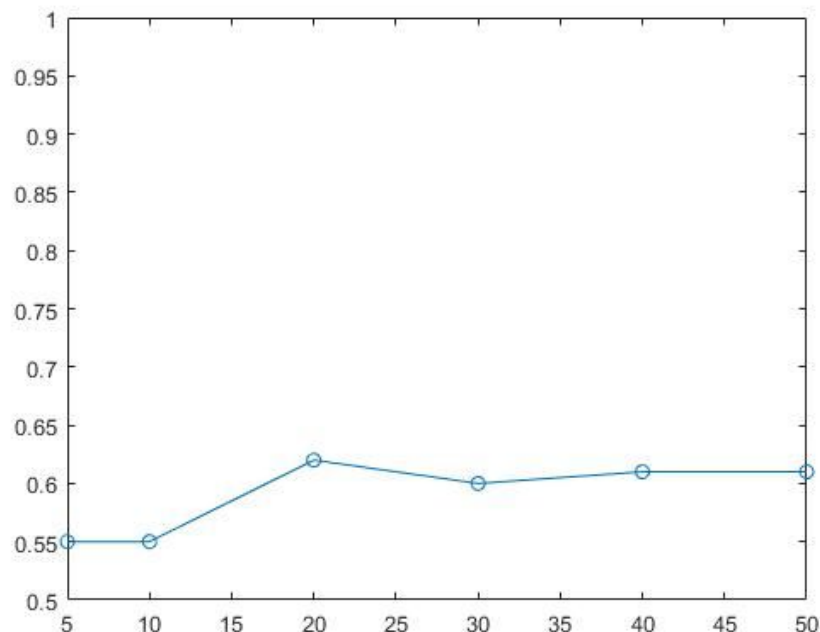
W konkursie obliczane są cztery wersje miary F1:

- ocena przypisanych etykiet oraz ocena w skali liczbowej 0-5 są ignorowane (F1A);
- dopasowana etykieta powinna być zgodna, ale oceny w skali liczbowej są ignorowane (F1AT);
- dopasowana etykieta jest ignorowana, ale każde dopasowanie jest karane, gdy ocena w skali liczbowej jest niezgodna (F1AS);
- dopasowana etykieta powinna być zgodna oraz każde dopasowanie jest karane, gdy ocena w skali liczbowej jest niezgodna (F1AST).

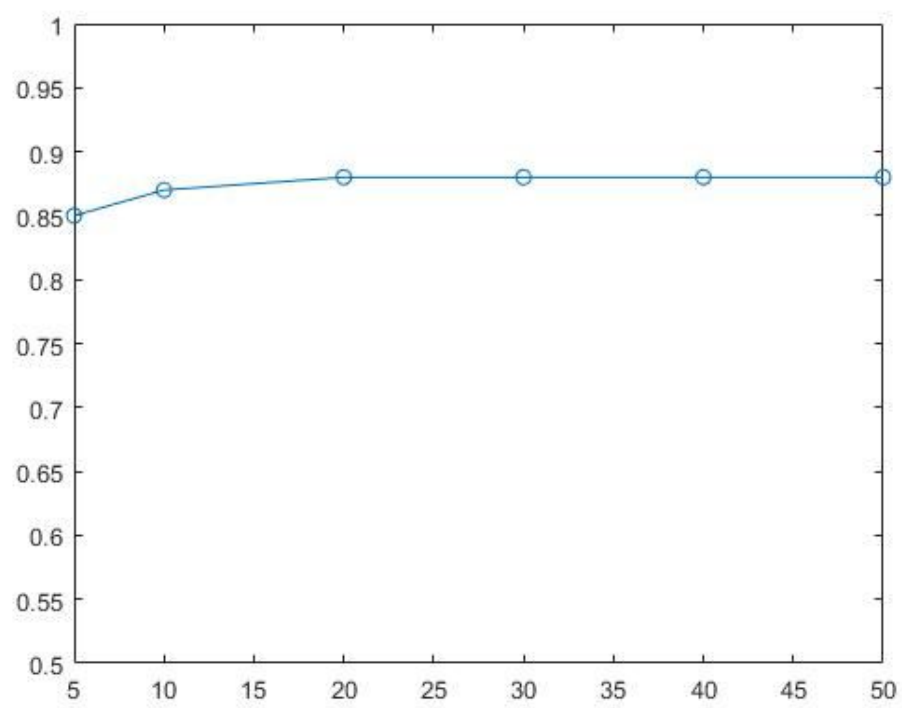
W naszym przypadku, gdy przyporządkowanie fraz nie jest wyznaczane przez system, lecz pobierane z pliku konkursowego, wyznaczone współczynniki nie będą dokładnie określać jakości całego systemu. Jednak pozwalają one na miarodajniejsze porównanie naszej implementacji z innymi rozwiązaniami konkursowymi.

### Testy:

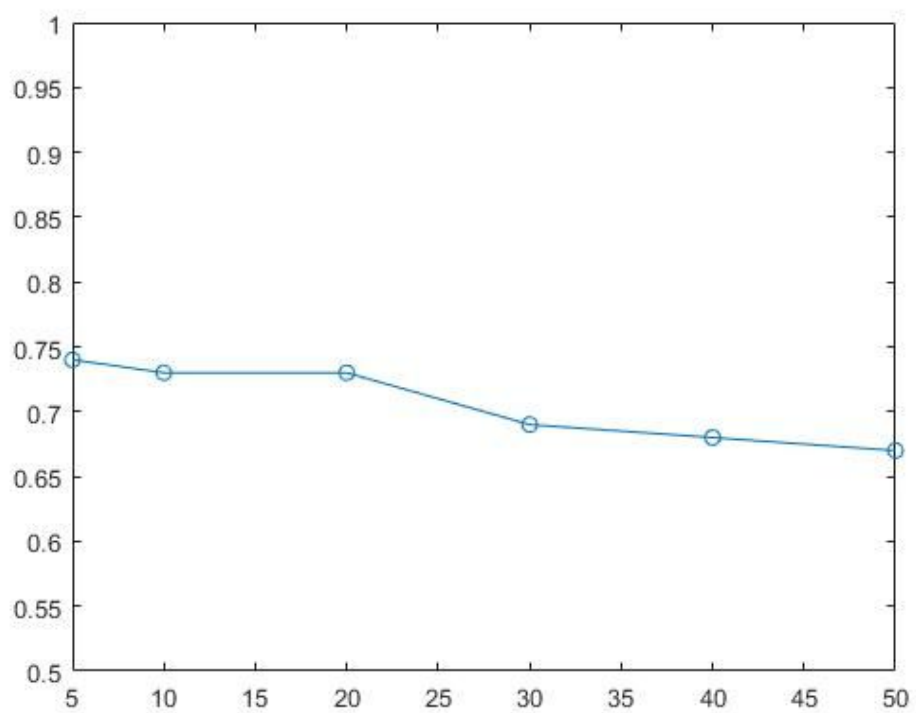
Poniższe wykresy przedstawiają zależność poszczególnych miar F1 od liczby iteracji algorytmu minimalizacji L-BFGS. Pominęliśmy pierwszą wersję tej miary (w której zarówno etykieta jak i ocena dopasowania są ignorowane), gdyż nasz system nie wyznacza przyporządkowania fraz z pierwszego i drugiego zdania i wartość F1 w tym przypadku zawsze wynosi 1.



Rys. 1. F1AT



Rys. 2. F1AS



Rys. 2. F1AST

Liczba iteracji powyżej 20 nie wpływa znacząco na wyniki, a wręcz w przypadku miary F1AST dalsze jej zwiększanie spowodowało spadek jakości. Zdecydowanie największe wartości są osiągane dla miary F1AS, zaś najmniejsze dla F1AT. Oznacza to, że system radzi sobie dobrze z przypisaniem liczbowej oceny podobieństwa, lecz ma trudności z dopasowaniem etykiety. Pewną poprawę można uzyskać, dodając proste reguły na współwystępowanie etykiety EQUI z oceną 5 oraz etykiety NOALI z oceną NIL (co jest zgodne z instrukcją konkursową).

#### **Bibliografia:**

- [1] Barbara Rychalska, *Sieci neuronowe w rozpoznawaniu podobieństwa semantycznego*, 2016
- [2] <http://alt.qcri.org/semeval2016/index.php?id=tasks>
- [3] <http://alt.qcri.org/semeval2015/task2/index.php?id=data-and-tools>