

# Detectarea Limbajului Figurat

Iulia Pincu

`iulia-maria.pincu@s.unibuc.ro`

Andrada Bejenaru

`andrada-elena.bejenaru@s.unibuc.ro`

## Rezumat

Acest proiect își propune să traducă expresii figurative sau metaforice în forma lor literală, identificând automat semnificația intenționată. Setul de date conține enunțuri metaforice urmate de două interpretări posibile, iar sarcina modelului este să aleagă interpretarea corectă. Prin antrenarea unui model TF-IDF + Regresie Logistică pe aceste date, am obținut o acuratețe ridicată în maparea metaforei la înțelesul literal.

## 1 Introducere

Limbajul figurativ este prezent în comunicarea cotidiană și necesită un proces cognitiv de traducere a sensului metaforic în semnificație literală. Sarcina noastră a fost să construim un model capabil să interpreteze expresii metaforice și să aleagă dintre două opțiuni posibile pe cea care reflectă sensul corect. Această abordare nu implică identificarea metaforei în sine, ci recunoașterea efectului ei semantic.

## 2 Lucrări Asemănătoare

Majoritatea abordărilor privind metaforele implică clasificarea expresiilor ca fiind figurative sau literale. În schimb, noi urmăm o linie puțin explorată: traducerea automată a metaforei în forma ei intenționată. Similar cu abordările de tip commonsense inference, această sarcină presupune înțelegerea sensului implicit. Modele precum BERT sau GPT ar putea fi aplicate, dar aici testăm un model simplu bazat pe TF-IDF.

## 3 Metodologie

### 3.1 Setul de Date

Datele conțin perechi:

- o expresie figurativă (ex: *"The voyage will be a walk through fire."*),

- două interpretări (una corectă, una incorectă),
- și o etichetă binară ce indică dacă prima interpretare este cea corectă.

Datele sunt organizate ca rânduri cu patru coloane: enunț metaforic, interpretare A, interpretare B și etichetă.

## 3.2 Preprocesare

Fiecare rând a fost transformat în două exemple: **enunț + interpretare A** și **enunț + interpretare B**. Eticheta este 1 pentru varianta corectă și 0 pentru cea greșită. Această structurare permite un model de clasificare binară.

## 3.3 Modelul

Am folosit vectorizare TF-IDF (cuvinte + caractere) și un clasificator Logistic Regression. Modelul a fost antrenat pe setul de train, optimizat prin GridSearch și evaluat pe setul de validare. Am folosit de asemenea un pipeline cu **FeatureUnion** pentru a combina trăsături la nivel de cuvânt și caracter.

## 4 Rezultate

Modelul a obținut o acuratețe de aproximativ 50% pe setul de test, ceea ce indică faptul că a reușit să învețe unele patternuri între expresiile metaforice și interpretările lor corecte, dar că există și multe cazuri în care predicțiile nu au fost corecte.

Având în vedere că este un model simplu, bazat pe TF-IDF și Regresie Logistică, fără înțelegere contextuală profundă, aceste rezultate sunt totuși promițătoare. Ele oferă o bază pe care se pot construi soluții mai avansate în viitor.

Modelul s-a descurcat mai bine cu metafore explicite și interpretări clare, dar a întâmpinat dificultăți atunci când diferențele dintre opțiuni erau subtile sau când sensul figurativ necesita cunoștințe culturale sau contextuale.

## 5 Lucrări Viitoare

- Adăugarea de date din alte domenii (emoții, relații, afaceri);
- Evaluarea sensibilității modelului la metafore culturale sau ambigue;
- Compararea performanței cu un model transformer românesc antrenat pe același set.

## 6 Concluzie

Traducerea automată a limbajului figurativ este o sarcină fezabilă chiar și cu metode clasice. Prin utilizarea TF-IDF și logistic regression, am demonstrat că înțelegerea semantică implicită poate fi obținută și fără modele complexe, dacă datele sunt bine structurate.

## 7 Limitări

Modelul nu funcționează bine pentru metafore rare sau ambigue și nu oferă o explicație pentru alegeri. De asemenea, modelul nu este robust la schimbări semantice subtile sau exprimări ironice.

## 8 Declarație Etică

Acest proiect nu implică date personale și nu produce concluzii morale. Totuși, interpretarea automată a metaforelor poate fi sensibilă în contexte sociale sau culturale, necesitând atenție în aplicații reale.

## Mulțumiri

Mulțumim cadrului didactic coordonator pentru încurajarea originalității și a gândirii critice în procesarea limbajului natural.

## Bibliografie

- [1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*.