

Klasyfikacja tekstów z użyciem Transfer Learning i Active Learning

Michał Kuźba

B. Burny, M. Godek, M. Kuźba, P. Zięcik

Uniwersytet Warszawski & Sigmoidal

- Klasyfikacja tekstów
- Jak wytrenować model na małym zbiorze danych?
- *One-shot learning*

Stworzenie narzędzia do klasyfikacji na małych zbiorach danych przy użyciu dwóch technik:

- **Transfer Learning**

Trenujemy model na ogólnym, dużym zbiorze, luźno powiązanym z przyszłymi zadaniami. Dotrenowujemy model do następnych zadań używając małego, dostępnego zbioru.

- **Active Learning**

Powiększamy zbiór treningowy, w możliwie oszczędny sposób, wybierając przykłady, na których model jest najmniej pewny.

- Prosta sieć konwolucyjna z *Dropoutem*
- Reprezentacja tekstu z użyciem pretrenowanych embeddingów *Glove*

- **Zadanie:** binarna klasyfikacja na silnie niezrównoważonym zbiorze - wyszukiwanie rzadkiej kategorii
- **Metryka:** average precision
- **Korpus:** *Reuters RCV1* - 800,000 artykułów (podział na test/train według kategorii - symulacja nowych zadań)

- Warstwy reprezentacji i warstwy klasyfikacji
Przenosimy warstwy reprezentacji
- *Learning Discriminative Representations (LeaDR)*
Strumień małych zadań klasyfikacyjnych. Naprzemiennie zamrażamy warstwy klasyfikacji i reprezentacji

Podziel korpus na podzbiory T_1, \dots, T_n i powtarzaj dla $i \in \{1, \dots, n\}$

- Podziel podkorpus T_i na dwie części: C_i oraz R_i
- Zresetuj warstwy klasyfikacji
- Zamroź warstwy reprezentacji i wytrenuj model na C_i
- Odmroź warstwy reprezentacji, zamroź część klasyfikacyjną i wytrenuj model na R_i
- Ewaluacja

- Wybierz kategorię do rozpoznania
- Powtarzaj:
 - Wybierz "najciekawsze" artykuły i dodaj je do zbioru treningowego
 - Trenuj model na powiększonym zbiorze
 - Sprawdź skuteczność sieci

- Niepewność w modelu *Epistemic*
Może być wyjaśniona przy odpowiedniej ilości danych
- Niepewność w danych (Aleatoric)

Funkcje akwizycji:

- proste, np. entropia, max. prawdopodobieństwo

$$H_q = \sum_{i=1}^N -p_i \cdot \log_q p_i$$

- **Variational Dropout**

- entropia średniej
- **BALD** - *Bayesian Active Learning by Disagreement*

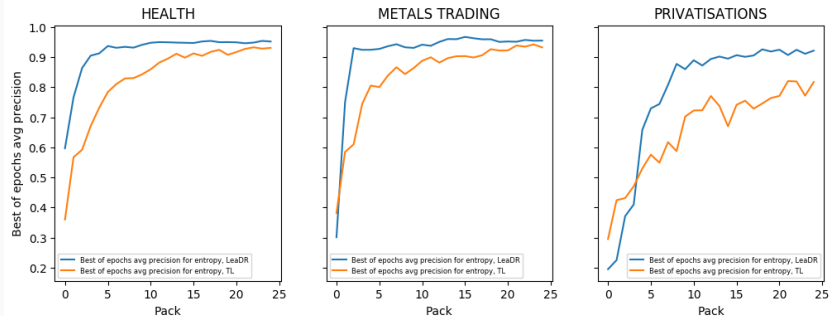
Liczymy:

entropia uśrednionej predykcji – średnia entropia predykcji

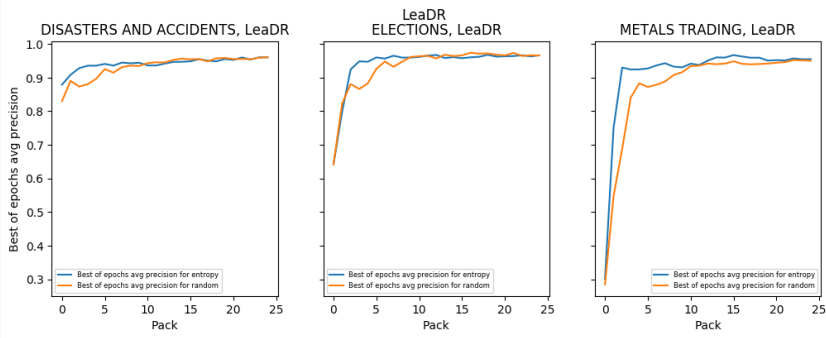
Intuicyjnie:

patrzemy na niezgodność różnych predykcji, raczej niż na
"niepewność" w obrębie pojedynczych predykcji

Eksperyment TL



Eksperyment AL



Po co jeszcze nam szacowanie niepewności?

- obserwacje *out of class*
- CTR - ukryty Active Learning
- *ensembling*
- podejmowanie decyzji na podstawie modelu
- poprawa modelu

- Jak szacować niepewność poza klasyfikacją?
- Jak robić to bez warstwy dropout?
- Model-agnostic?
- Na istniejącym modelu?



Lewis Smith and Yarin Gal.

“Understanding Measures of Uncertainty for Adversarial Example Detection”. In: CoRR abs/1803.08533 (2018).



Yarin Gal, Riashat Islam, and Zoubin Ghahramani.

“Deep Bayesian Active Learning with Image Data”. In: CoRR abs/1703.02910 (2017).



Ouais Alsharif, Philip Bachman, and Joelle Pineau.

“Lifelong Learning of Discriminative Representations”. In: CoRR abs/1404.4108 (2014).



Inbar Naor

Uncertainty in Deep Learning, PyData Tel Aviv, (2017)