

Analiza porównawcza wyników klasyfikatorów SVC, regresji liniowej oraz sztucznej sieci neuronowej w rozpoznawaniu par zdań typu przesłanka-konkluzja oraz par zdań losowych na podstawie podobieństwa semantycznego

Bartosz Swędrowski

Wstęp

Dzięki metodom uczenia maszynowego jesteśmy w stanie poddawać dyskursy analizie za pomocą algorytmów tej dziedziny i na skutek tego uzyskiwać cenne informacje o charakterze wypowiedzi w tych dyskursach. Jednym ze sposobów pozyskiwania informacji w temacie dyskursów politycznych jest analiza postów umieszczanych przez użytkowników na forach internetowych. Tematem niniejszego tekstu jest analiza porównawcza wyników pracy trzech rodzajów klasyfikatorów na fałszywych parach przesłanka-konkluzja oraz prawdziwych parach przesłanka-konkluzja wyabstrahowanych z danych zebranych z forów internetowych.

Wprowadzenie do użytych narzędzi

IAT

Teoria Zakotwiczania Inferencji (IAT) jest metodą anotacji transkrypcji dialogów, służącą do ekstrakcji struktur zdaniowych. Dzięki temu możliwa jest automatyzacja analiz procesów argumentacji, co pozwala na, między innymi, prowadzenie badań na języku naturalnym. (Visser et al., 2019)

Rodzaje argumentów w IAT

Mapy utworzone za pomocą IAT rozróżniają argumenty na szeregowy oraz równoległy:

- argument szeregowy to taki argument, którego przesłanki układają się w parę przesłanka-konkluzja w jednej linii. Jeśli jedna z nich z jakiegoś powodu zostanie wyeliminowana, argument przestaje mieć znaczenie,

- argument równoległy to taki argument, który posiada wiele wspierających go przesłanek, im więcej, tym lepiej jest poparty i tym samym mocniej wspiera teorię, której dotyczy. Dzięki swojej strukturze nawet po utracie jednej z przesłanek zachowuje swoją siłę.

W mapach utworzonych za pomocą Teorii Zakotwiczenia Inferencji wyróżniamy (Visser et al., 2019):

- lokucje: jednostki, w których znajduje się podzielony tekst korpusu. Lokucje formułuje się w formacie "MÓWCA: jednostka argumentatywna dyskursu",
- tranzycje: struktury uchwytyujące relację między lokucjami,
- illokucje: struktury uchwytyujące funkcje lokucji lub tranzycji pomiędzy nimi. Przykładowymi typami illokucji są zgodzenie się, nie zgodzenie się, poddawanie w wątpliwość,
- propozycje: tekst lokucji, który jest zrekonstruowany w sposób zależny od illokucji,
- inferencje: ukierunkowane indykatory relacji między propozycjami, odzwierciedlające poparcie propozycją akceptacji innej propozycji,
- konflikty: ukierunkowane indykatory relacji między propozycjami, odzwierciedlające wchodzenie jednej propozycji w konflikt z drugą propozycją lub relacją propozycyjną,
- parafrazy: ukierunkowane indykatory relacji między propozycjami, odzwierciedlające fakt, że propozycja jest parafrazą innej propozycji.

OVA

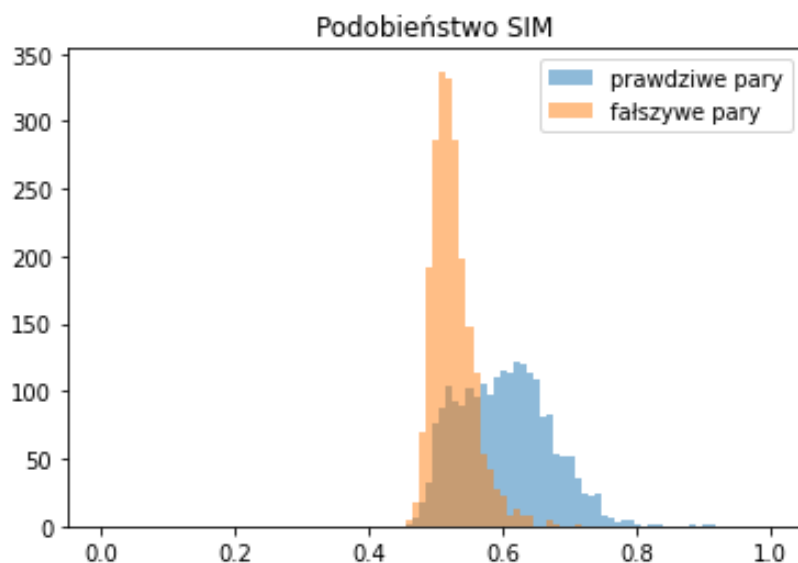
OvaArgTech (Janier et al., 2014) jest oprogramowaniem ułatwiającym pracę z korpusami. Umożliwia ono analizę transkryptów argumentacji w postaci tworzenia schematów argumentacyjnych konkretnych części korpusu za pomocą interfejsu, na przykład według struktury anotacji IAT. Jest dostępne online, bez wymagania opłaty dostępowej (Visser et al., 2019).

AIFdb

AIFdb (Lawrence et al., 2012) to baza danych przechowująca korpusy zaanotowanych instancji tekstu, oferująca również ich wizualizację. AIFdb umożliwia wielu narzędziom na różne rodzaje interakcji z przechowywanymi danymi (Lawrence et al., 2012).

Opis korpusu IAC

The Internet Argument Corpus (IAC) (Walker et al., 2012) to korpus służący do badania debat politycznych na forach internetowych. Składa się z 1808 map zawierających w sobie około 11 tysięcy dyskusji, około 390 tysięcy postów oraz około 73 milionów słów. Podzbiory korpusu zostały zaanotowane między innymi w kategoriach tematu, postawy, zgody, sarkazmu oraz "przykrego charakteru". Dane korpusu dostępne są w formie plików JSON, z większością anotacji w formie plików CSV (Walker et al., 2012).



Rysunek 1: Wykres obrazujący stopień podobieństwa prawdziwych par przesłanka-konkluzja oraz fałszywych par przesłanka-konkluzja w korpusie IAC (opracowanie własne).

Opis kodu i klasyfikatorów

Całość procesu przebiegała w następującej kolejności:

- (i) tokenizacja i lematyzacja tekstu korpusu
- (ii) wydobywanie z korpusu par przesłanka-konkluzja oraz stworzenie fałszywych par,
- (iii) obliczenie wskaźnika podobieństwa semantycznego obu typów par,
- (iv) stworzenie zestawienia podobieństwo semantyczne-para/nie para,

- (v) podział zestawienia na zbiór treningowy i testowy,
- (vi) trening klasyfikatora SVC i ewaluacja stworzonego modelu,
- (vii) trening klasyfikatora regresji liniowej i ewaluacja stworzonego modelu,
- (viii) trening klasyfikatora skonstruowanego na podstawie sztucznej sieci neuronowej i ewaluacja stworzonego modelu.

Do realizacji zadania użyto bibliotek NumPy (Harris et al., 2020), pandas (pandas development team, 2020), scikit (Pedregosa et al., 2011), Tensorflow (Abadi et al., 2015) oraz Keras (Chollet et al., 2015).

Porównanie wyników klasyfikatorów

Porównanie wyników klasyfikatorów SVC, regresji liniowej oraz sztucznej sieci neuronowej w rozpoznawaniu par zdań typu przesłanka-konkluzja oraz par zdań losowych:

SVC

		Klasa predykowana	
		p	n
Klasa rzeczywista	p'	278 TP	145 FN
	n'	38 FP	406 TN

Rysunek 2: Macierz pomyłek klasyfikatora SVC (opracowanie własne).

Dokładność: 78,89%

Regresja liniowa

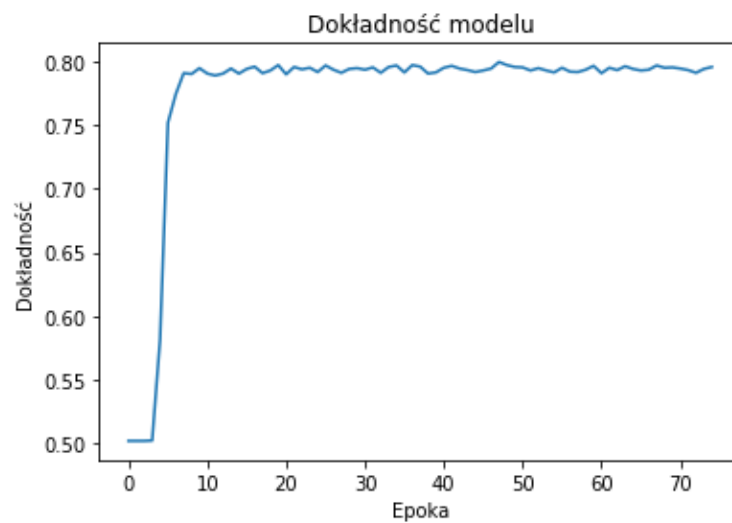
		Klasa predykowana	
		p	n
Klasa rzeczywista	p'	290 TP	133 FN
	n'	47 FP	397 TN

Rysunek 3: Macierz pomyłek klasyfikatora regresji liniowej (opracowanie własne).

Dokładność: 79,24%

Sztuczna sieć neuronowa

Sztuczna sieć neuronowa składała się z czterech warstw ukrytych, każda o wielkości 128 neuronów, z funkcją aktywacji *Rectified Linear Unit* (ReLU). Warstwa wyjściowa składała się z dwóch neuronów z sigmoidalną funkcją aktywacji. Wykorzystano optymalizator typu *Adaptive Moment Estimation* (ADAM) o wartości współczynnika uczenia 0.0001, funkcji straty *Sparse Categorical Cross-entropy*. Użyto również metryki *accuracy*. Model wytrenowano przy wartości batch size równą 24, z liczbą epok równą 75.



Rysunek 4: Wykres dokładności modelu na przestrzeni epok (opracowanie własne).

		Klasa predykowana	
		p	n
Klasa rzeczywista	p'	296 TP	127 FN
	n'	53 FP	391 TN

Rysunek 5: Macierz pomyłek klasyfikatora stworzonego przy użyciu sztucznej sieci neuronowej (opracowanie własne).

Dokładność: 79,01%

Podsumowanie wyników

Wszystkie klasyfikatory uzyskały dokładność klasyfikacji na poziomie około 79% na zbiorze testowym o wielkości 433 par typu para przesłanka-konkluzja oraz 434 par losowych. Podobny wynik uzyskany przez algorytmy świadczy o odpowiednim dostosowaniu poziomu trudności zadania do wszystkich zastosowanych mechanizmów klasyfikacji. Z dużym prawdopodobieństwem wynik mógłby zostać podwyższony dzięki zwiększeniu ilości analizowanych map.

Zakończenie

Celem projektu było stworzenie oraz porównanie trzech dwuklasowych klasyfikatorów (predyktorów) danych dotyczących par zdań (bycia parą zdań losowych lub parą przesłanka-konkluzja) na podstawie wskaźnika podobieństwa semantycznego tych zdań. Zbiory treningowe i testowe składały się z obliczonych wskaźników podobieństwa semantycznego par fraz oraz wskaźnika bycia lub nie bycia prawdziwą parą przesłanka-konkluzja. Kategoria par typu para przesłanka-konkluzja zawierała 2166 instancje. Kategoria par typu losowego zawierała 2166 instancje. Do zrealizowania zadania wykorzystano sztuczne sieci neuronowe, algorytm regresji liniowej oraz algorytm Support Vector Machine. Średnia dokładność klasyfikacji (stosunek TP+TN do P+N) wyniosła 79,05%.

Literatura

- M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL <https://www.tensorflow.org/>. Software available from tensorflow.org.
- F. Chollet et al. Keras, 2015. URL <https://github.com/fchollet/keras>.
- C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, Sept. 2020. doi: 10.1038/s41586-020-2649-2. URL <https://doi.org/10.1038/s41586-020-2649-2>.
- M. Janier, J. Lawrence, and C. Reed. Ova+: an argument analysis interface. In S. Parsons, N. Oren, C. Reed, and F. Cerutti, editors, *Computational Models of Argument*, Frontiers in artificial intelligence and applications,

- pages 463–464, Netherlands, 2014. IOS Press. ISBN 9781614994350. doi: 10.3233/978-1-61499-436-7-463. URL <http://www.comma-conf.org/>. Fifth International Conference on Computational Models of Argument, COMMA 2014 ; Conference date: 09-09-2014 Through 12-09-2014.
- J. Lawrence, F. Bex, M. Snaith, and C. Reed. Aifdb: Infrastructure for the argument web. volume 245, 01 2012. doi: 10.3233/978-1-61499-111-3-515.
- T. pandas development team. pandas-dev/pandas: Pandas, Feb. 2020. URL <https://doi.org/10.5281/zenodo.3509134>.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85): 2825–2830, 2011. URL <http://jmlr.org/papers/v12/pedregosa11a.html>.
- J. Visser, B. Konat, R. Duthie, M. Koszowy, K. Budzynska, and C. Reed. Argumentation in the 2016 us presidential elections: annotated corpora of television debates and social media reaction. 2 2019.
- M. Walker, P. Anand, J. Fox Tree, R. Abbott, and J. King. A corpus for research on deliberation and debate. In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC). 2012.