

Sztuczna Inteligencja

Modelowanie silnych przekonań

Alicja Kowalewska 165726

Alicja Smaruj 165151

Maciej Retowski 165429

maj 2018

1 Cele projektu

Celami projektu były ocena, na ile podstawowe metody przetwarzania tekstów pozwalają modelować zawarte w nich silne przekonania, oraz sprawdzenie, czy zastosowane modelowanie będzie wystarczające do stworzenia programu umożliwiającego klasyfikację artykułów zgodnie z wyrażanym w nich poglądem.

Programy rozpoznające przekonania w tekście pisanym w języku naturalnym mogą mieć bardzo wiele zastosowań, na przykład rozróżnianie pozytywnych i negatywnych komentarzy czy łatwe wyszukiwanie artykułów zawierających określone stanowisko.

2 Opis projektu

Przekonania, których modelowanie spróbowano przeprowadzić, dotyczyły katastrofy smoleńskiej. Część artykułów dostępnych w internecie odzwierciedla pogląd, iż rzeczywiście była to katastrofa, podczas gdy w innych zdarzenie to jest traktowane jako zamach.

Podjęta została próba przeanalizowania podobieństw i różnic pomiędzy artykułami wyrażającymi te dwa przekonania. Oczekiwany rezultat było pokazanie, że podstawowe metody przetwarzania tekstu umożliwiają dostrzeżenie (i określenie) różnicy pomiędzy artykułami wyrażającymi przeciwne poglądy.

Skrypty stworzone podczas pracy nad projektem, o których mowa w dalszej części sprawozdania, oraz pełne wyniki znajdują się pod adresem: <https://github.com/mreto/SI>.

3 Zastosowane metody

3.1 Zgromadzenie danych

Stworzyliśmy trzy zbiory artykułów:

- wyrażających przekonanie o katastrofie,
- wyrażających przekonanie o zamachu,
- neutralnych

oraz

- zbiór słów kluczowych.

Do stworzenia pierwszych dwóch katalogów zostały wykorzystane materiały wcześniej zgromadzone, zaklasyfikowane i przygotowane przez inną grupę projektową. Zawierały one artykuły o wydarzeniu z 10 kwietnia 2010 roku opublikowane w latach 2010 - 2016: 29 o katastrofie i 41 o zamachu.

Oprócz tego użyto zbioru 60 słów kluczowych – najważniejszych słów związanych z tematem artykułów (te również wybrane zostały przez grupę pracującą nad podobnym zagadnieniem w ubiegłych latach).

154	kabina	pancerny	Smolensk
analiza	Kaczynski	pilot	spis
Antoni	katastrofa	pogoda	stenogram
blad	klamstwo	prezydent	strzał
Blasik	komisja	prokuratura	swiadek
bomba	Komorowski	przyczyny	trotyl
Bronisław	Krauze	Putin	tupolew
brzoza	krzyk	Rosja	Tusk
Donald	krzyż	rosyjski	wieża
dym	lotnisko	samolot	wrak
ekspert	Macierewicz	sekcja	wybuch
film	mgła	skrzydło	zamach
general	nacisk	skrzynka	zaniedbanie
hanba	nalega	śląd	zbrodnia
identyfikacja	ofiary	sledztwo	zdrada

Rysunek 1: Lista słów kluczowych.

W celu zebrania neutralnych danych zastosowano skrypt `pars_wiki.py`¹ pobierający artykuły z Wikipedii zawierające wybrane słowa kluczowe i w ten sposób otrzymano zbiór 46 neutralnych artykułów. Poddano je wstępnemu przetworzeniu ułatwiającemu dalszą pracę – pozbyto się polskich znaków, wielkich liter oraz znaków nowej linii.

3.2 Przetwarzanie tekstów

Każdą z grup artykułów przetworzono do postaci macierzy za pomocą skryptu `script.py`² w następujący sposób:

- każdemu słowu kluczowemu odpowiada wiersz w macierzy – wektor słów kontekstowych
- dla każdego wystąpienia wybranego słowa kluczowego w tekście znaleźć należy słowa występujące w pobliżu – określany jest margines (przyjęto granicę równą 5), ile słów następujących po słowie kluczowym i ile słów poprzedzających je, będzie brane pod uwagę
- jeśli w granicy tego marginesu pojawiają się znaki sygnalizujące koniec zdania (. ! ?), słowa nienależące do tego samego zdania zostają odrzucone
- im bliżej słowo jest położone względem słowa kluczowego, tym większą otrzymuje wagę (..., 0, 1, 2, 3, 4, 5, SłowoKluczowe, 5, 4, 3, 2, 1, 0, 0 ...)
- przy pomocy odpowiedniego słownika³ przeprowadzona zostaje lematyzacja słowa (czyli wyznaczony odpowiadający danej formie gramatycznej leksem)

¹https://github.com/mreto/SI/blob/master/pars_wiki.py

²<https://github.com/mreto/SI/blob/master/script.py>

³Słownik, z którego skorzystano, został stworzony na podstawie słownika fleksyjnego Polimorf dostępnego w ramach programu Morfeusz (<http://sgjp.pl/morfeusz/index.html>).

- jeśli słowo kontekstowe nie wystąpiło wcześniej, zostaje dodane jako nowa kolumna macierzy i w komórce odpowiadającej słowu kluczowemu ustawiona zostaje odpowiednia waga, a jeśli wystąpiło ono już wcześniej, waga odpowiedniej komórki zostaje zwiększona
- usunięto tzw. *stop-words*, czyli słowa, które nie wносиły podczas analizy żadnej cennej informacji (głównie zaimki, przyimki i spójniki)

Po przetworzeniu wszystkich artykułów i stworzeniu macierzy przeanalizowano otrzymane rezultaty. Zredukowano liczbę kolumn o ok. 70%, odcinając słowa powtarzające się zbyt rzadko (mające sumaryczną wagę mniejszą niż 10 i takie, które ponad połowę swoich wystąpień uzyskały przy tym samym słowie kluczowym). Następnie gotowe wektory poddano normalizacji, aby umożliwić ich późniejsze porównywanie. W tym celu dla każdego wektora szukano maksimum i dzielono przez jego wartość wszystkie pozostałe wartości.

Poniżej przedstawionych zostało kilka przykładowych słów kluczowych wraz z ich słowami kontekstowymi, które uzyskały największe wartości wag (podane wagi przed normalizacją):

TUPOLEW		SPISEK		PRZYCZYNY	
Słowa kontekstowe	Waga	Słowa kontekstowe	Waga	Słowa kontekstowe	Waga
skrzydło	20	teoria	20	katastrofa	176
miec	19	przekonywac	5	okolicznosc	36
wrak	19	nowy	4	badanie	28
prezydencki	15	zamach	4	wypadek	23
rozbic	15	kolejny	4	prawdopodobny	20

Tabela 1: Wektory kontekstowe na podstawie artykułów zawierających przekonanie o zamachu

TUPOLEW		SPISEK		PRZYCZYNY	
Słowa kontekstowe	Waga	Słowa kontekstowe	Waga	Słowa kontekstowe	Waga
prezydencki	45	zamach	12	katastrofa	100
wrak	10	wielki	5	smolenski	36
skrzydło	10	lubic	5	badajacy	20
katastrofa	9	portal	5	komisja	16

Tabela 2: Wektory kontekstowe na podstawie artykułów zawierających przekonanie o katastrofie

Obszerniejsza analiza wszystkich słów kluczowych znajduje się pod adresem: https://github.com/mreto/SI/tree/master/analize_vectors.

Dodatkowo w późniejszym etapie zaimplementowano także selekcję cech (*mutual information*), aby uwzględniać tylko te słowa, które są najważniejsze dla klasyfikacji artykułów. Zastosowana metoda określa, jak duże znaczenie dla dokonania poprawnej klasyfikacji ma obecność lub brak danego słowa⁴. Rezultatem jej użycia była redukcja wektorów do długości rzędu kilku słów. Mapy samoorganizujące się oraz wykresy przedstawione w dalszej części sprawozdania uwzględniają wektory sprzed zastosowania selekcji cech.

3.3 Mapy samoorganizujące się

W celu umożliwienia interpretacji otrzymanych wektorów zastosowano tzw. mapy samoorganizujące się (inaczej mapy Kohonena lub SOM – Self Organizing Maps).

Sieć Kohonena, na której bazują SOM-y, jest to sieć neuronowa uczona w trybie bez nauczyciela w celu sklasyfikowania wielowymiarowych danych wejściowych w taki sposób, aby umożliwić ich reprezentację w mniejszej liczbie wymiarów, a jednocześnie jak najwierniej odwzorować wewnętrzną strukturę wektora wejściowego.

Do generowania map użyto gotowych funkcji z pakietu SomToolBox dla GNU Octave⁵. Początkowo zastosowano domyślne parametry i rozmiar mapy 12 x 12. W ramach testów sprawdzono

⁴<https://nlp.stanford.edu/IR-book/html/htmledition/mutual-information-1.html>

⁵<https://github.com/limosek/somtoolbox>

skuteczność także dla innych parametrów: metody początkowej inicjalizacji 'randint', dla map o większym rozmiarze (30 x 30) oraz dla algorytmu treningu 'seq'.

3.4 Metody analizy wyników

Aby umożliwić ocenę zgodności działania map samoorganizujących się z oczekiwaniami, wygenerowano mapy:

- dla zbioru artykułów zgodnych z przekonaniem o katastrofie,
- dla zbioru artykułów zgodnych z przekonaniem o zamachu,
- dla zbioru artykułów neutralnych,
- dla 29 stadiów pośrednich pomiędzy artykułami stronnictwymi i neutralnymi uzyskanych poprzez dodawanie stopniowo do zbioru artykułów neutralnych coraz większych porcji artykułów stronnictwowych,
- dla 29 stadiów pośrednich pomiędzy artykułami stronnictwowymi i neutralnymi uzyskanych poprzez dodawanie do zbioru artykułów neutralnych kolejnych wielokrotności zbioru artykułów stronnictwowych, tak aby zmienić proporcje liczności zbioru artykułów stronnictwowych względem neutralnych.

Następnie stworzono macierze zawierające odległości pomiędzy wszystkimi parami słów kluczowych znajdującymi się na SOM-ach (czyli liczbową reprezentację wyników uzyskanych dzięki SOM-om) oraz obliczono podobieństwa między macierzami generowanymi w poszczególnych krokach a macierzami danych neutralnych i stronnictwowych według wzoru:

$$\text{podobienstwo} = \sum_{i=1}^n \sum_{j=1}^n |a_{i,j} - b_{i,j}|$$

gdzie:

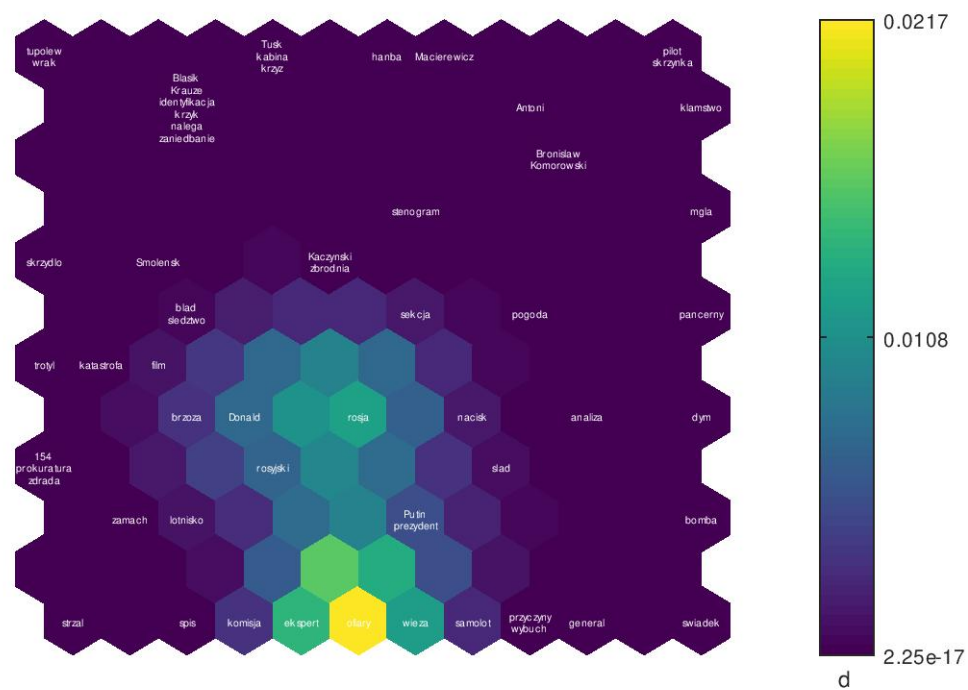
- n to liczba słów kluczowych,
- a to element macierzy odległości pomiędzy parami słów kluczowych na mapie przedstawiającej badany pogląd,
- b to element macierzy odległości pomiędzy parami słów kluczowych na mapie przedstawiającej pogląd neutralny lub stronnictwowy.

Wyliczenie powyższych wartości umożliwia dosyć łatwe porównywanie podobieństw między mapami. Na tej podstawie spróbowano stworzyć prosty klasyfikator rozpoznający, do którego poglądu badany artykuł jest bardziej podobny.

4 Wyniki

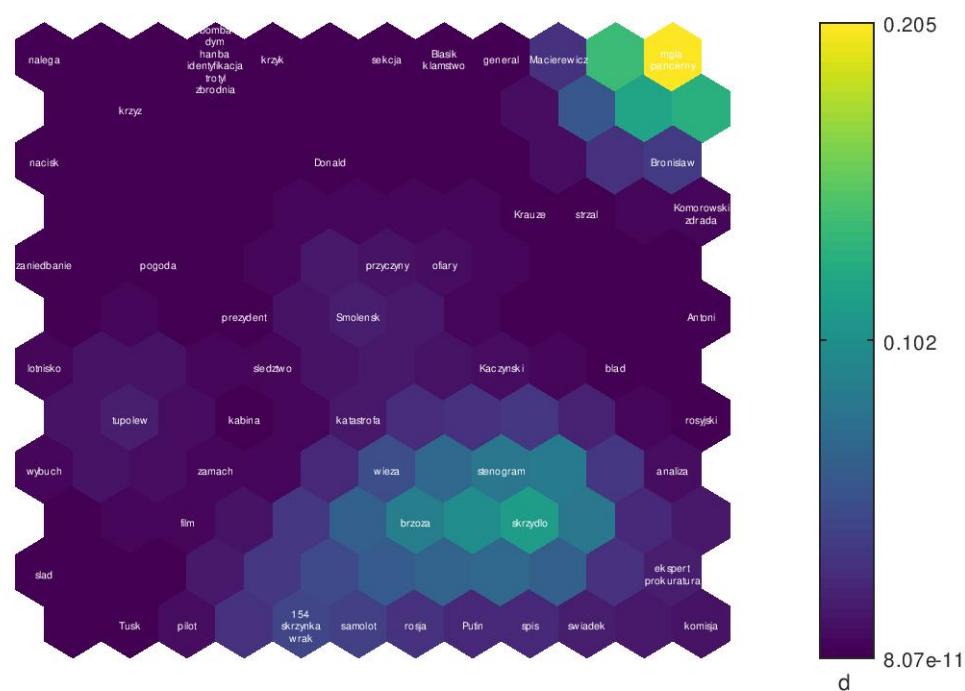
4.1 Mapy samoorganizujące się

Poniżej znajdują się mapy wygenerowane przy domyślnych ustawieniach kolejno dla danych: neutralnych, wyrażających przekonanie o katastrofie oraz wyrażających przekonanie o zamachu.



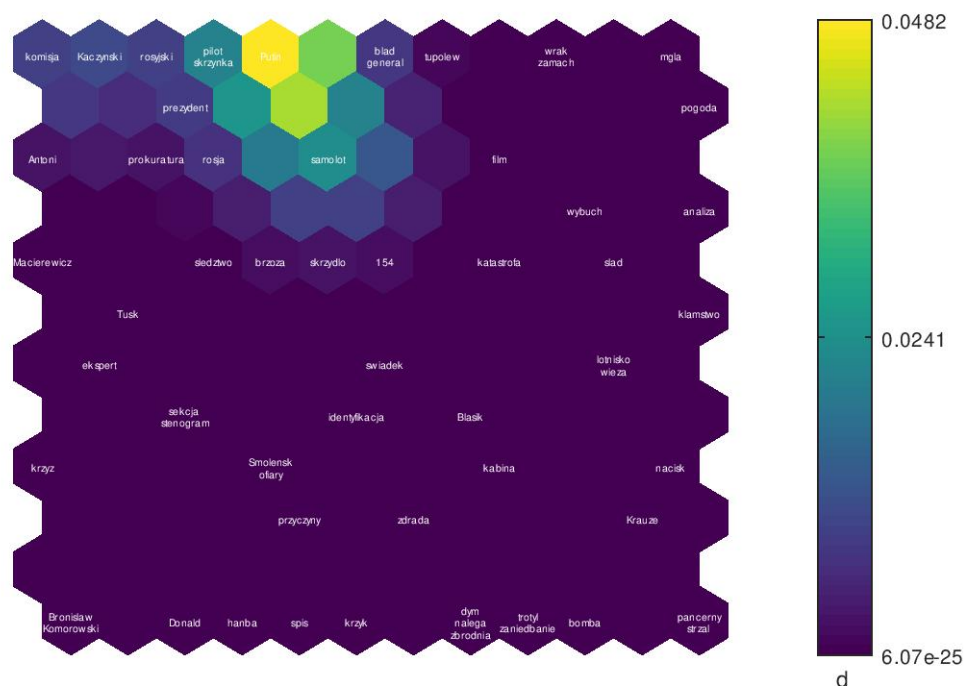
SOM 16-May-2018

Rysunek 2: SOM dla danych neutralnych.



SOM 17-May-2018

Rysunek 3: SOM dla danych wyrażających przekonanie o katastrofie.



SOM 17-May-2018

Rysunek 4: SOM dla danych wyrażających przekonanie o zamachu.

Na przedstawionych powyżej mapach można zaobserwować wykryte powiązania pomiędzy słowami kluczowymi, np. blisko siebie leżą takie pary jak: "Bronisław" + "Komorowski", "mgła" + "pogoda", "brzoza" + "skrzydło", "trotyl" + "bomba", "identyfikacja" + "sekcja", "Rosja" + "Putin".

4.2 Przejścia pomiędzy mapą neutralną a stroniczą

4.2.1 Komentarz do wykresów

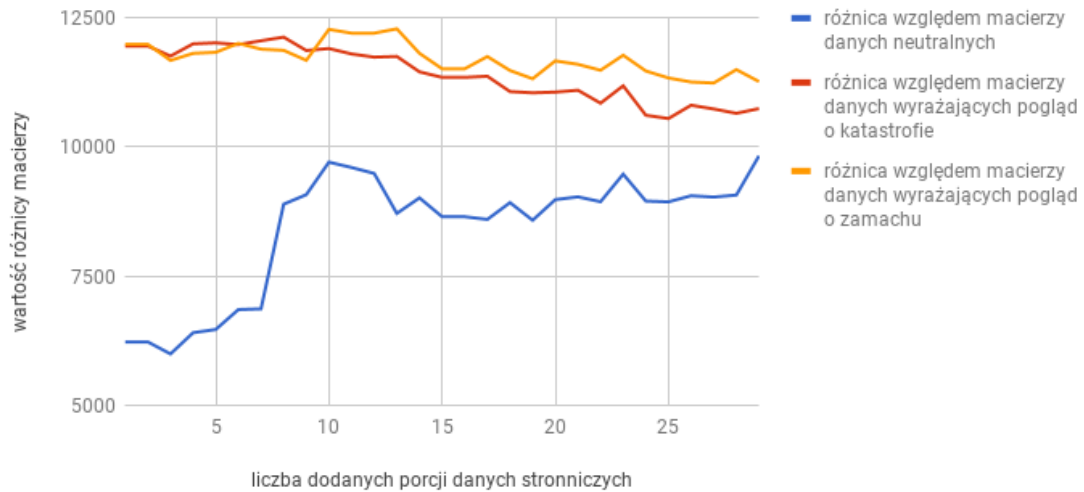
Widoczne poniżej rys. 5, 6, 7 i 8 przedstawiają wykresy pozwalające ocenić podobieństwa między mapami neutralnymi i związanymi z jakimś poglądem a wersjami pośrednimi.

Na rys. 5 widać, że gdy dodawano do mapy neutralnej kolejne porcje artykułów wyrażających pogląd o katastrofie, różnica względem mapy neutralnej wzrastała (zwłaszcza w początkowych dziesięciu iteracjach), a różnica względem mapy wyrażającej pogląd o katastrofie malała i pozostawała w większości przypadków niższa niż względem mapy wyrażającej pogląd o zamachu. Sytuacja jest analogiczna w przypadku rys. 6 (prezentującym zmiany podczas dodawania kolejnych porcji artykułów wyrażających pogląd o zamachu) – z tą różnicą, że krzywe katastrofy i zamachu są bliższe względem siebie i wielokrotnie się przecinają.

Na rys. 7 i 8 można zaobserwować, że przy wielokrotnym dodaniu zestawu artykułów związanych z danym poglądem do mapy neutralnej wynik staje się coraz bliższy temu pogładowi – czerwona krzywa na rys. 7 i żółta krzywa na rys. 8 mają tendencje malejące. Dodatkowo widać, że proces ten oddala mapę od neutralnej, ale jej różnica względem poglądu przeciwnego nie ulega znacznym zmianom.

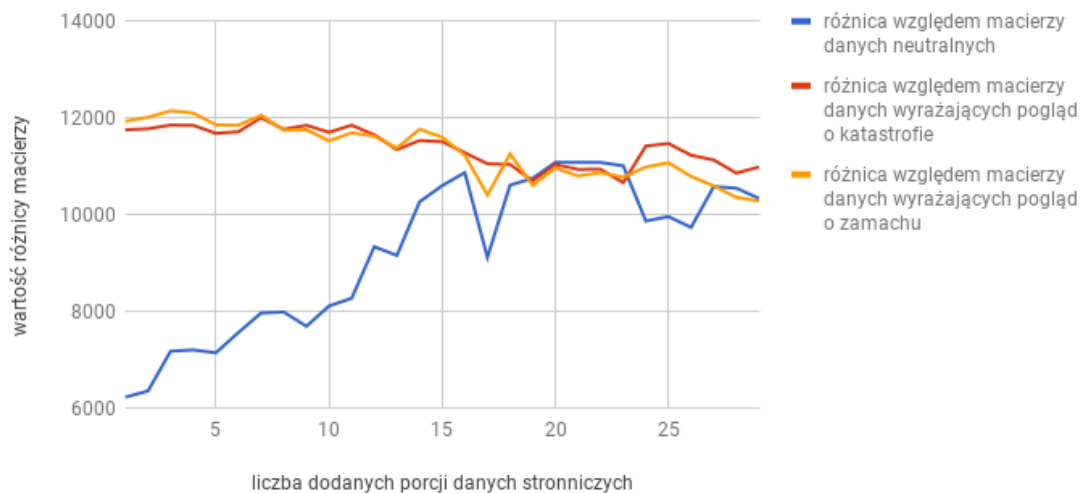
4.2.2 Wykresy

Wyniki przy dodawaniu porcji artykułów wyrażających pogląd o katastrofie



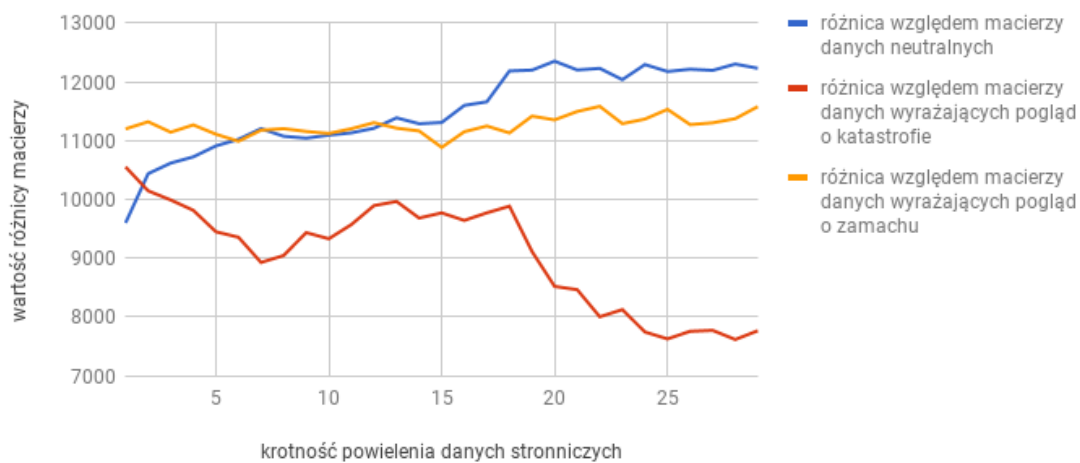
Rysunek 5: Wykres obrazujący zmiany map podczas dodawania kolejnych artykułów wyrażających przekonanie o katastrofie.

Wyniki przy dodawaniu porcji artykułów wyrażających pogląd o zamachu



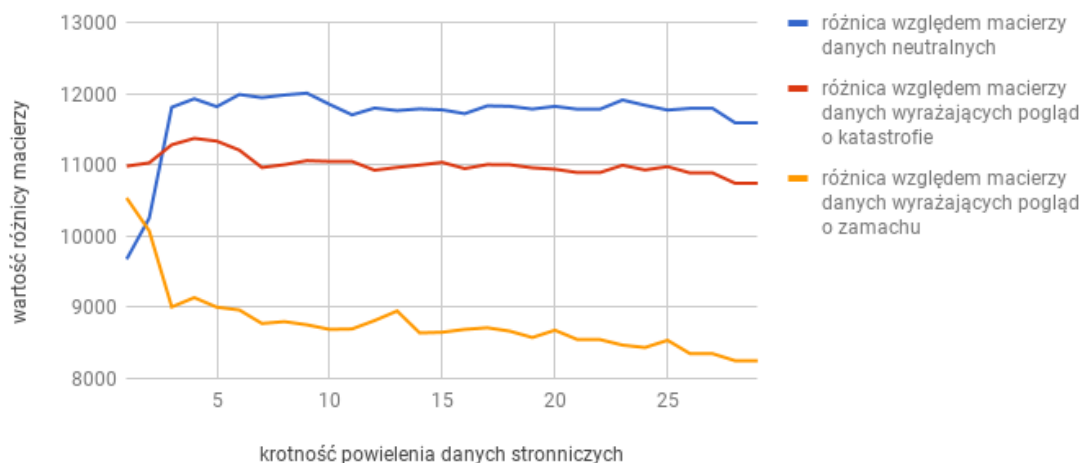
Rysunek 6: Wykres obrazujący zmiany map podczas dodawania kolejnych artykułów wyrażających przekonanie o zamachu.

Wyniki przy dodawaniu powielonych artykułów wyrażających pogląd o katastrofie



Rysunek 7: Wykres obrazujący zmiany map podczas dodawania kolejnych wielokrotności zbioru artykułów wyrażających przekonanie o katastrofie.

Wyniki przy dodawaniu powielonych artykułów wyrażających pogląd o zamachu



Rysunek 8: Wykres obrazujący zmiany map podczas dodawania kolejnych wielokrotności zbioru artykułów wyrażających przekonanie o zamachu.

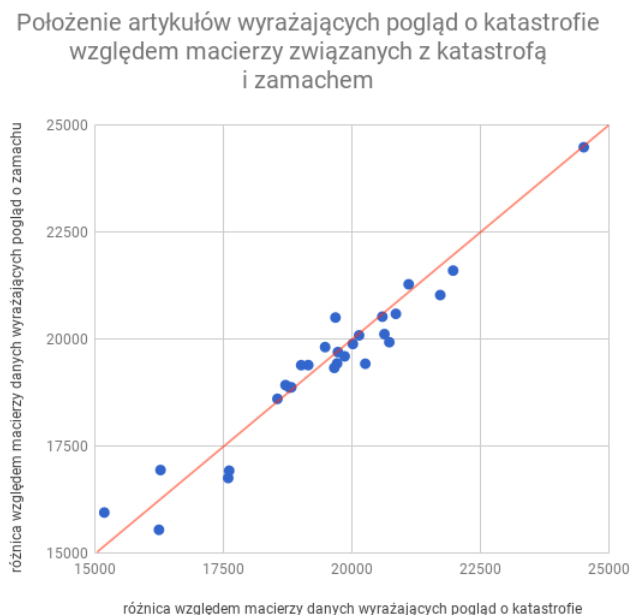
4.3 Wykresy obrazujące klasyfikację artykułów

4.3.1 Komentarz do wykresów

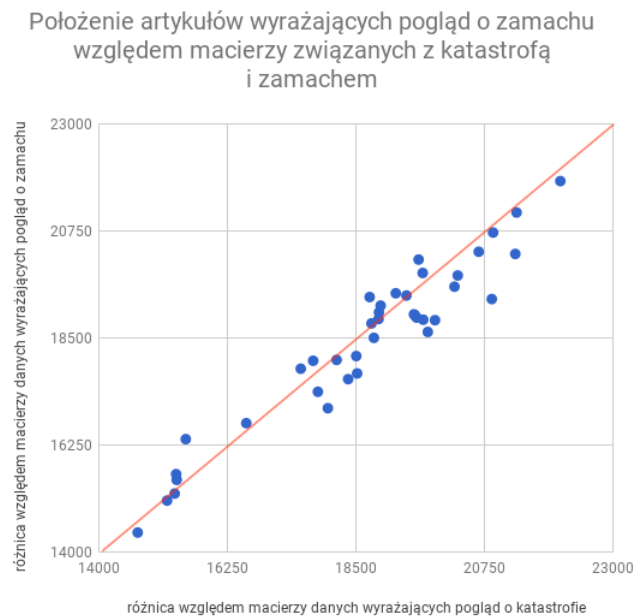
Dzięki rys. 9 i 10 można ocenić, czy mapy dla artykułów z poszczególnych grup rzeczywiście okazują się bliższe mapie dla wyrażanego w nich poglądu.

Na rys. 9 i 10 widać, że zarówno artykuły wyrażające pogląd o katastrofie, jak i wyrażające pogląd o zamachu układają się bardzo blisko przekątnej wykresu, co oznacza, że zachowują podobne odległości od obu poglądów.

4.3.2 Wykresy



Rysunek 9: Wykres odległości pojedynczych artykułów związanych z poglądem o katastrofie względem map w pełni stronicznych.



Rysunek 10: Wykres odległości pojedynczych artykułów związanych z poglądem o zamachu względem map w pełni stronicznych.

4.4 Klasyfikator

Prosty klasyfikator porównujący mapę każdego artykułu z mapą wyrażającą przekonanie o katastrofie i mapą wyrażającą przekonanie o zamachu osiągnął następujące wyniki:

1. przy domyślnych ustawieniach - skuteczność 53,2% (rozpoznał 62,5% artykułów wyrażających przekonanie o zamachu i 40% artykułów wyrażających przekonanie o katastrofie),
2. dla map o rozmiarze 30x30 - skuteczność 53,5% (rozpoznał 70% artykułów wyrażających przekonanie o zamachu i 30% artykułów wyrażających przekonanie o katastrofie),
3. dla inicjalizacji 'randinit' - skuteczność 45,6% (rozpoznał 25% artykułów wyrażających przekonanie o zamachu i 75% artykułów wyrażających przekonanie o katastrofie),
4. dla algorytmu treningu 'seq' - skuteczność osiąga w każdym wykonaniu zupełnie inną wartość, nie sposób więc sugerować się otrzymywanymi w ten sposób rezultatami.

5 Wnioski

Z zaprezentowanych powyżej wykresów (rys. 5 - 8) można wywnioskować, iż przygotowane mapy Kohonena w pewnym stopniu odzwierciedlają przekonanie zawarte w zbiorze tekstów. Dodawanie stronicznych artykułów do początkowo neutralnego zbioru rzeczywiście powoduje odchyłanie wartości w odpowiednich kierunkach.

Mapy samoorganizujące się dla pojedynczych tekstów nie wykazują jednakże zbyt dużego podobieństwa do żadnej z grup. Różnice pojedynczych artykułów względem przekonania o katastrofie a względem przekonania o zamachu (rys. 9 i 10) nie są wystarczająco duże, aby można było na ich podstawie stworzyć klasyfikator. Prawdopodobnie jest to związane z faktem, iż testowane artykuły są zbyt krótkie, aby na bazie pojedynczego tekstu mapa była w stanie wiarygodnie zobrazować zawarte w nim zależności - jeden artykuł nie zawiera większości słów kluczowych, a średnia długość wektorów kontekstowych wynosi ok. 17 słów (podczas gdy dla pełnych zbiorów było to ok. 500). Wyjaśniałoby to niezbyt dobrą skuteczność klasyfikatora.

Pewną poprawę rezultatów mogłoby spowodować:

- zwiększenie liczby badanych artykułów,
- dobór innego zestawu słów kluczowych,
- wychwytywanie zaprzeczeń w tekstach.

Zasadnicza przyczyna niskiej skuteczności wydaje się jednak tkwić nie w zastosowanych metodach, ale w samych artykułach. Teksty zostały przyporządkowane do poszczególnych kategorii przez poprzednie grupy projektowe. Po weryfikacji okazało się, że przekonanie o katastrofie lub zamachu w wielu artykułach nie tylko nie jest wyrażone wprost, ale wręcz wcale nie daje się na podstawie tekstu określić.

Zaimplementowane w ramach projektu metody mogłyby więc okazać się skuteczne dla lepszego zbioru danych.

6 Bibliografia

- Duch W., *Memetics and Neural Models of Conspiracy Theories*, <https://arxiv.org/abs/1508.04561>, data dostępu: 15.06.2018.
- Analizator morfologiczny *Morfeusz*, <http://sgjp.pl/morfeusz/index.html>, data dostępu: 17.05.2018.
- SOM Toolbox dla GNU Octave, <https://github.com/limosek/somtoolbox>, data dostępu: 15.05.2018.
- Selekcja cech (metoda *mutual information*), <https://nlp.stanford.edu/IR-book/html/htmledition/mutual-information-1.html>, data dostępu: 25.06.2018.