Modelowanie Języka Naturalnego

Piotr Wierzgała

2019-07-22

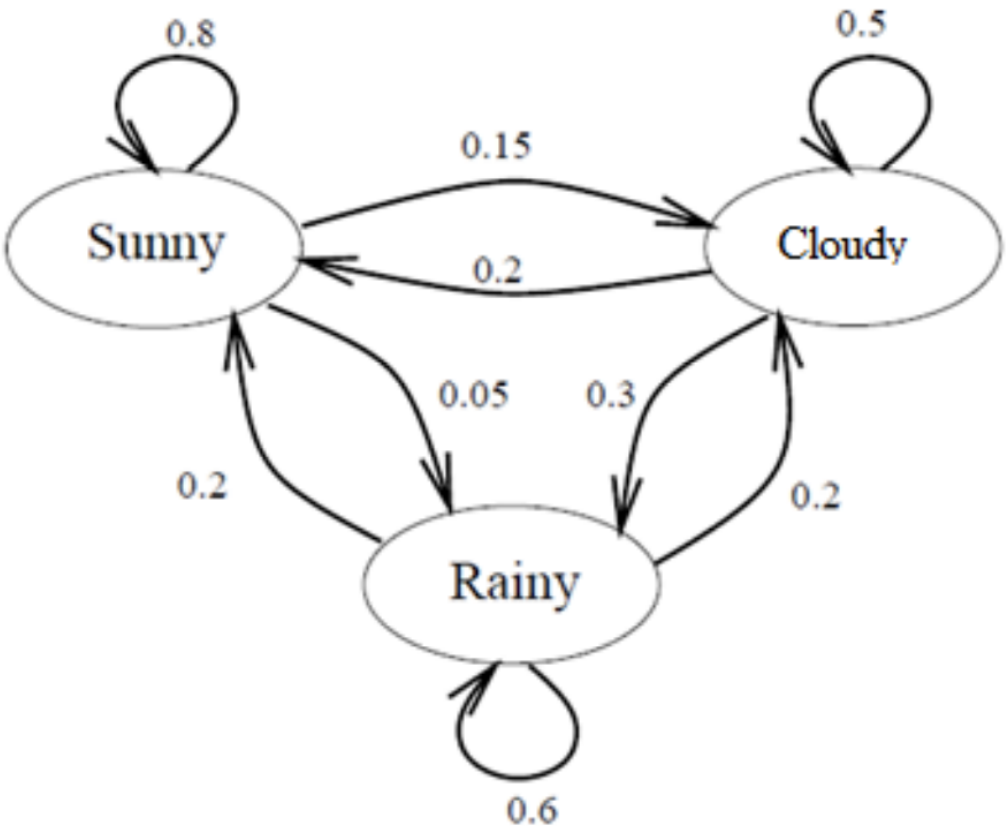
Przykładowe zastosowania

- Ocena jakości tłumaczenia
- Ocena jakości transkrypcji mowy
- Generowanie tekstu
- Korygowanie błędów językowych

Modele probabilistyczne

- Gdzie znajduje się salon optyczny?
- Gdzie znajduje się salon apteczny?

Model Markowa, własność Markowa



$$P(Sunny|Sunny) = 0.8$$
$$P(Rainy|Sunny) = 0.05$$
$$P(Cloudy|Sunny) = 0.15$$
$$\Big\} \; 1$$

$$P(Sunny|Rainy) = 0.2$$
$$P(Rainy|Rainy) = 0.6$$
$$P(Cloudy|Rainy) = 0.2$$
$$\Big\} \; 1$$

$$P(Sunny|Cloudy) = 0.2$$
$$P(Rainy|Cloudy) = 0.3$$
$$P(Cloudy|Cloudy) = 0.5$$
$$\Big\} \; 1$$

Rysunek: Model Markova przedstawiony w postaci grafu.

Model Markowa dla tekstu

Lata osa koło nosa
Lata mucha koło ucha
Lata bąk koło rąk

	bąk	koło	lata	mucha	nosa	osa	ucha	rąk
bąk	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
koło	0.0	0.0	0.0	0.0	1.0	0.0	1.0	1.0
lata	1.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0
mucha	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
nosa	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
osa	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
ucha	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0

(a) Tablica liczebności

	bąk	koło	lata	mucha	nosa	osa	ucha	rąk
bąk	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
koło	0.00	0.00	0.00	0.00	0.33	0.00	0.33	0.33
lata	0.33	0.00	0.00	0.33	0.00	0.33	0.00	0.00
mucha	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
nosa	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
osa	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
ucha	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00

(b) Tablica prawdopodobieństw

Model Markowa, rząd modelu

$P(s_i|s_{i-1})$ - Model Markowa pierwszego rzędu.
$P(s_i|s_{i-1}, \ldots, s_{i-n})$ - Model Markowa n rzędu.

	koło	nosa	lata	mucha	ucha	bąk	rąk
bąk koło	0.0	0.0	0.0	0.0	0.0	0.0	1.0
koło nosa	0.0	0.0	1.0	0.0	0.0	0.0	0.0
koło ucha	0.0	0.0	1.0	0.0	0.0	0.0	0.0
lata bąk	1.0	0.0	0.0	0.0	0.0	0.0	0.0
lata mucha	1.0	0.0	0.0	0.0	0.0	0.0	0.0
lata osa	1.0	0.0	0.0	0.0	0.0	0.0	0.0
mucha koło	0.0	0.0	0.0	0.0	1.0	0.0	0.0
nosa lata	0.0	0.0	0.0	1.0	0.0	0.0	0.0
osa koło	0.0	1.0	0.0	0.0	0.0	0.0	0.0
ucha lata	0.0	0.0	0.0	0.0	0.0	1.0	0.0

Rysunek: Tablica liczebności dla Modelu Markowa drugiego rzędu.

Model Markowa

Przykład: generowanie tekstu

Modele wektorowe

Worek wyrazów

The car is driven on the road

- [100000] - The
- [010000] - car
- [001000] - is
- [000100] - driven
- [000010] - on
- [000001] - road

0: The car is driven on the road
1: The truck is driven on the highway

	car	driven	highway	is	on	road	the	truck
0	1	1	0	1	1	1	2	0
1	0	1	1	1	1	0	2	1

Rysunek: Wektorowa reprezentacja zdań.

Funkcje rankingowe

Term-Frequency Inversed Document-Frequency

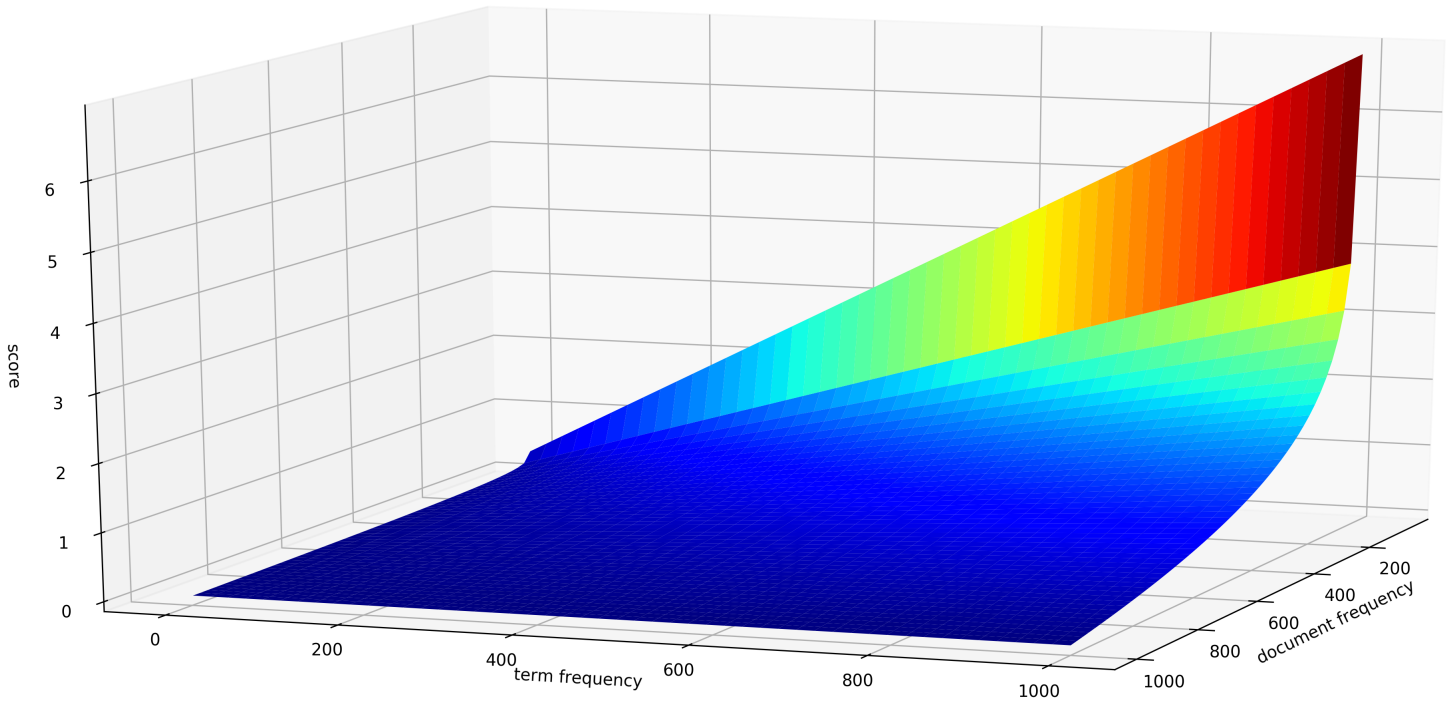
$$tfidf(t, d, D) = tf(t, d) \cdot idf(t, D)$$

gdzie:

- $tf(t, d)$ - Stosunek wystąpień wyrazu t w dokumencie d do wszystkich wyrazów w dokumencie d.
- $idf(t, D)$ - Stosunek liczby dokumentów, w których występuje wyraz t do liczby wszystkich dokumentów w korpusie (D).

Funkcje rankingowe

Term-Frequency Inversed Document-Frequency



Rysunek: Powierzchnia funkcji TF-IDF.

0: The car is driven on the road
1: The truck is driven on the highway

	car	driven	highway	is	on	road	the	truck
0	1	1	0	1	1	1	2	0
1	0	1	1	1	1	0	2	1

Rysunek: Wektorowa reprezentacja zdań.

	car	driven	highway	is	on	road	the	truck
0	0.043	0	0.000	0	0	0.043	0	0.000
1	0.000	0	0.043	0	0	0.000	0	0.043

Rysunek: Wektorowa reprezentacja zdań przetworzona przez TFIDF.

Funkcje rankingowe

Best Matching 25 (BM25)

$$bm25(t, d, D, k, b) = idf(t, D)\frac{tf(t, D) \cdot (k + 1)}{tf(t, D) + k \cdot (1 - b + b \cdot l(d, D))}$$

Funkcje rankingowe

Best Matching 25 (BM25)

Przykład: BM25

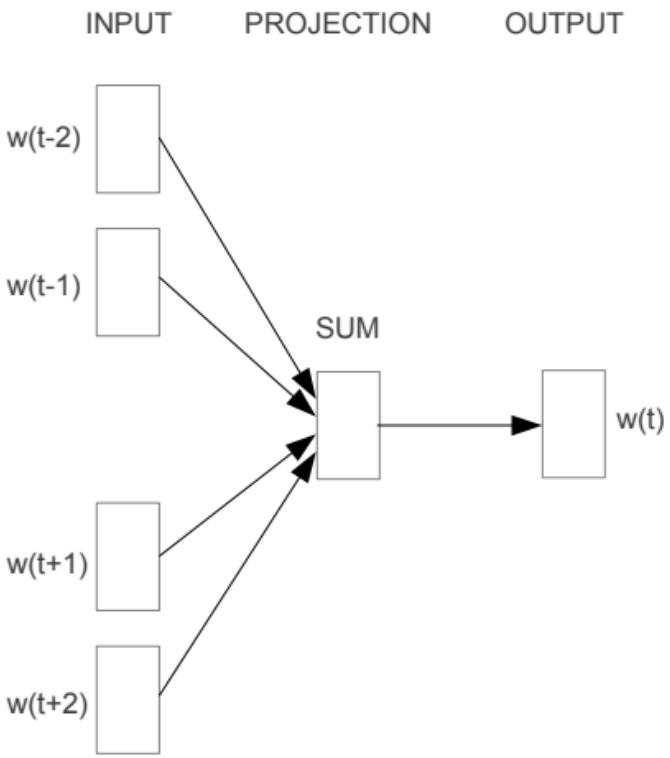
- "Oculist and eye-doctor occur in almost the same environments."
  Zellig Harris, 1954 r.
- "You shall know a word by the company it keeps!"
  John Firth, 1957 r.

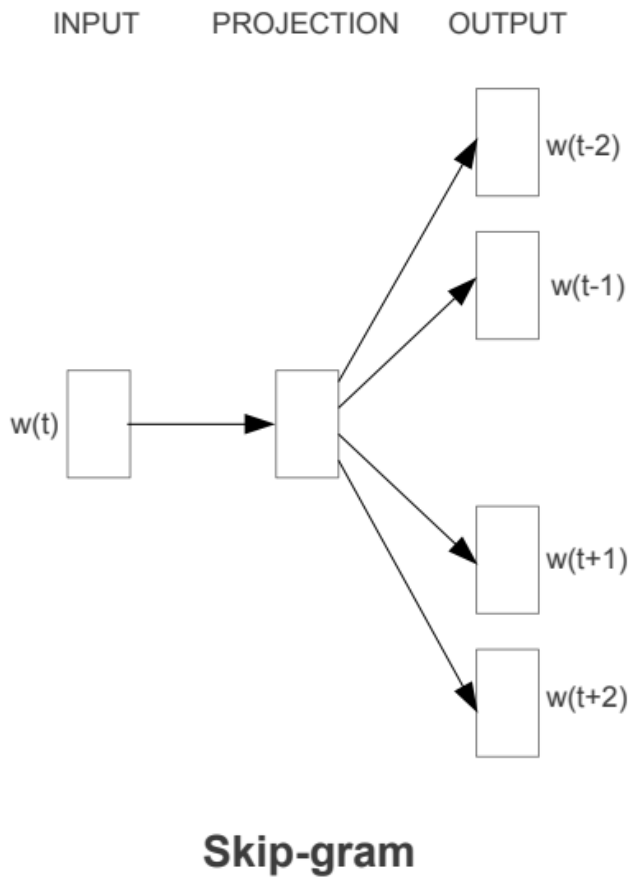
Przykład:

- Butelka tesgüino znajduje się na stole.
- Tesgüino uchodzi za smaczne.
- Spożywając tesgüino można się upić.
- Tesgüino wytwarzane jest z nasion kukurydzy.

Modele wektorowe

Wektory własnościowe





Rysunek: Architektury modeli zaproponowane w pracy Mikolov et al. 2013 (word2vec.

Modele wektorowe

Wektory własnościowe, okno kontekstowe

a	large	grey	cat	was	asleep	on	a	rocking	chair

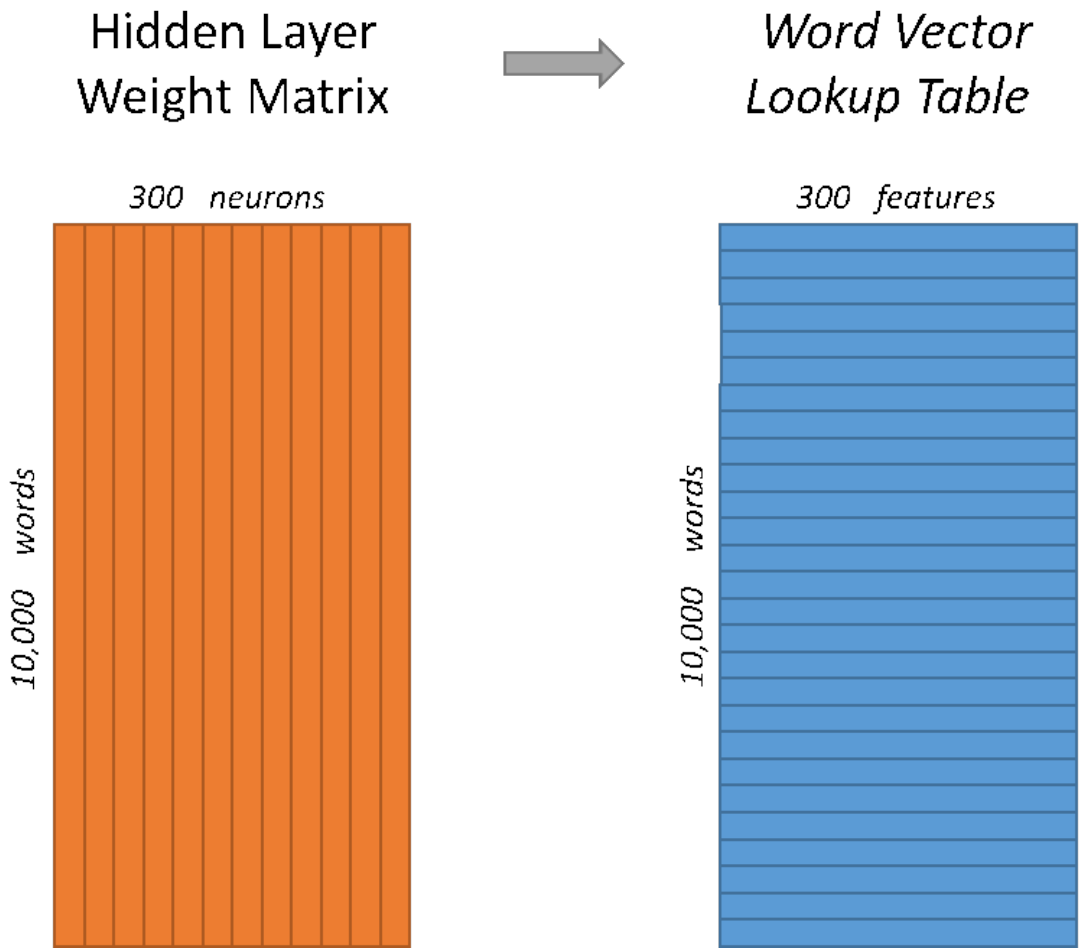
Rysunek: Rozmiar okna: 2, wektoryzowany wyraz “a”, pary wyrazów użyte do treningu sieci: (a, large), (a, grey).

a	large	grey	cat	was	asleep	on	a	rocking	chair

Rysunek: Rozmiar okna: 2, wektoryzowany wyraz “cat”, pary wyrazów użyte do treningu sieci: (cat, large), (cat, grey), (cat, was), (cat, asleep).

Modele wektorowe

Wektory własnościowe



Rysunek: Relacja pomiędzy modelem a wektorami własnościowymi.

Modele wektorowe

Wektory własnościowe

$$[0 \quad 0 \quad 0 \quad 1 \quad 0] \times \begin{bmatrix} 17 & 24 & 1 \\ 23 & 5 & 7 \\ 4 & 6 & 13 \\ 10 & 12 & 19 \\ 11 & 18 & 25 \end{bmatrix} = [10 \quad 12 \quad 19]$$

Rysunek: Przejście od wektora "1 z n" do wektora własnościowego.

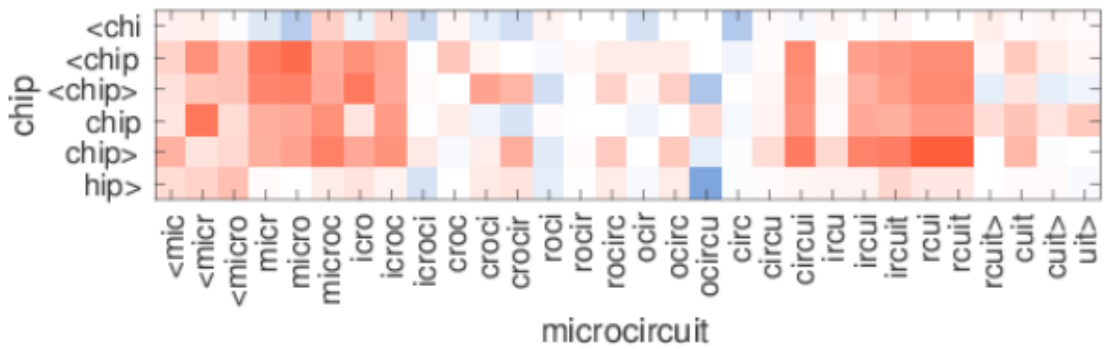
- Brak funkcji aktywacji dla warstwy ukrytej.
- Negative sampling.
- Subsampling.
  "a *large grey* **cat** *was asleep* on a chair"
  "large gray was asleep" $\longrightarrow$ "large gray asleep chair"

Modele wektorowe

Wektory własnościowe, fasttext, n-gramy znaków

N-gramy wyrazu apple dla n=3:

- ap
- app
- ppl
- ple
- le



Rysunek: Podobieństwo pomiędzy n-gramami znaków dla wyrazu spoza słownika.

Modele wektorowe

Wektory własnościowe, porównanie fasttext

Model	AG	Sogou	DBP	Yelp P.	Yelp F.	Yah. A.	Amz. F.	Amz. P.
BoW (Zhang et al., 2015)	88.8	92.9	96.6	92.2	58.0	68.9	54.6	90.4
ngrams (Zhang et al., 2015)	92.0	97.1	98.6	95.6	56.3	68.5	54.3	92.0
ngrams TFIDF (Zhang et al., 2015)	92.4	97.2	98.7	95.4	54.8	68.5	52.4	91.5
char-CNN (Zhang and LeCun, 2015)	87.2	95.1	98.3	94.7	62.0	71.2	59.5	94.5
char-CRNN (Xiao and Cho, 2016)	91.4	95.2	98.6	94.5	61.8	71.7	59.2	94.1
VDCNN (Conneau et al., 2016)	91.3	96.8	98.7	95.7	64.7	73.4	63.0	95.7
`fastText`, $h = 10$	91.5	93.9	98.1	93.8	60.4	72.0	55.8	91.2
`fastText`, $h = 10$, bigram	92.5	96.8	98.6	95.7	63.9	72.3	60.2	94.6

Table 1: Test accuracy [%] on sentiment datasets. `FastText` has been run with the same parameters for all the datasets. It has 10 hidden units and we evaluate it with and without bigrams. For char-CNN, we show the best reported numbers without data augmentation.

	Zhang and LeCun (2015)		Conneau et al. (2016)			`fastText`
	small char-CNN	big char-CNN	depth=9	depth=17	depth=29	$h = 10$, bigram
AG	1h	3h	24m	37m	51m	1s
Sogou	-	-	25m	41m	56m	7s
DBpedia	2h	5h	27m	44m	1h	2s
Yelp P.	-	-	28m	43m	1h09	3s
Yelp F.	-	-	29m	45m	1h12	4s
Yah. A.	8h	1d	1h	1h33	2h	5s
Amz. F.	2d	5d	2h45	4h20	7h	9s
Amz. P.	2d	5d	2h45	4h25	7h	10s

Table 2: Training time for a single epoch on sentiment analysis datasets compared to char-CNN and VDCNN.

Modele wektorowe

Wektory własnościowe

Przykład: king - man + woman = queen?

Podsumowanie

- Model Markowa
- Worek wyrazów (ang. bag of words)
- Funkcje rankingowe (ang. scoring functions)
- Wektory własnościowe (ang. word embeddings)
- Word2vec, Fasttext