

Głębokie sieci neuronowe w przetwarzaniu języka naturalnego

CaseWeek 2018

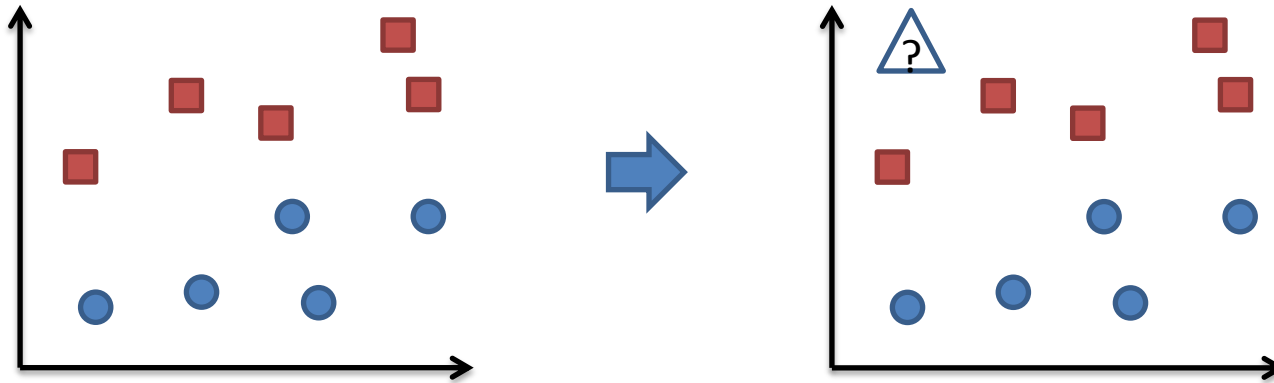
Tomasz Limisiewicz, Szymon Łęski

SAMSUNG

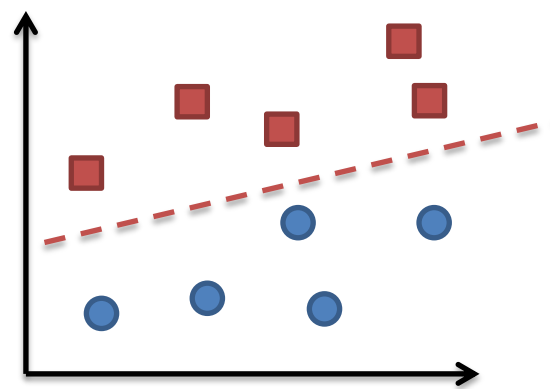
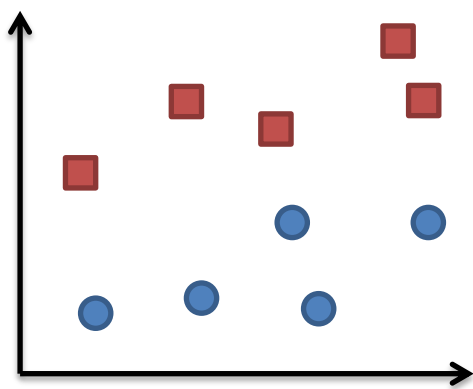
Agenda

- Sieci neuronowe – wstęp
- Część praktyczna: MNIST / Keras
- Sieci rekurencyjne
- Zastosowania sieci w NLP
- Modele języka
- Część praktyczna: modele języka

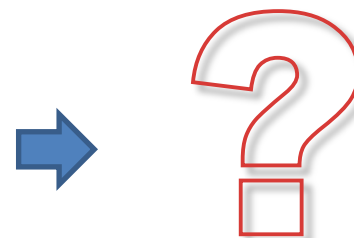
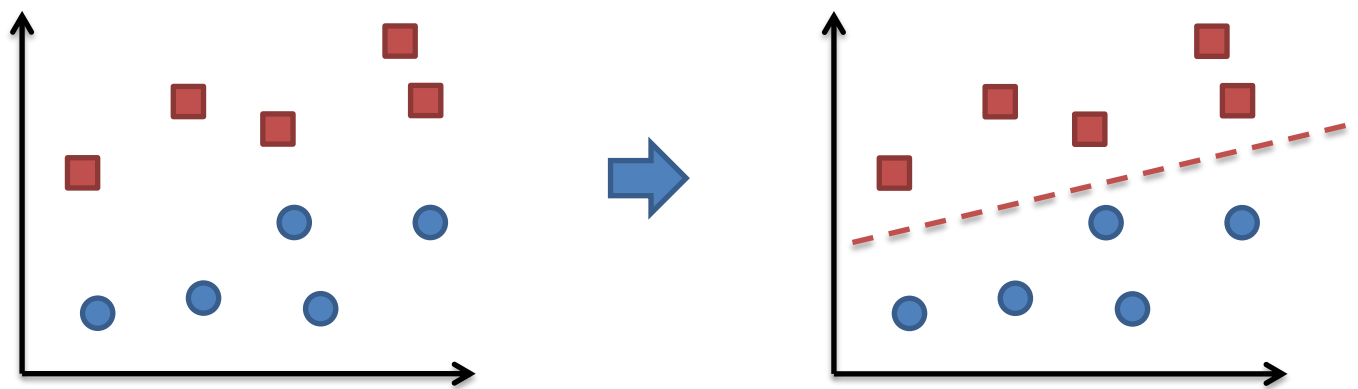
Klasyfikacja



Klasyfikacja – uczenie maszynowe



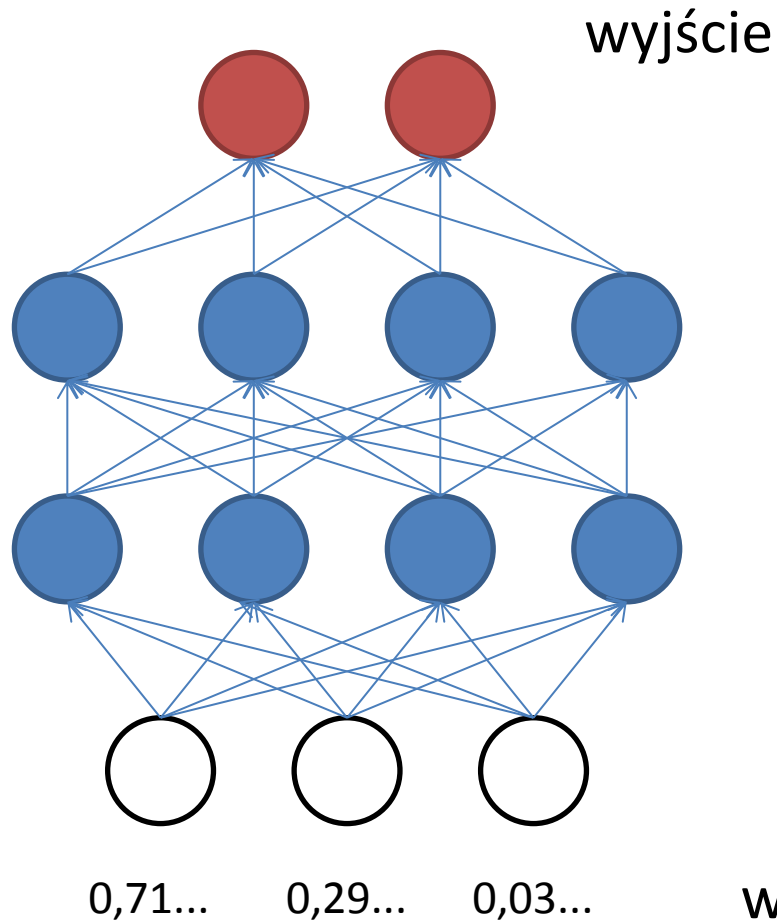
Klasyfikacja – sieci neuronowe



Sieci neuronowe

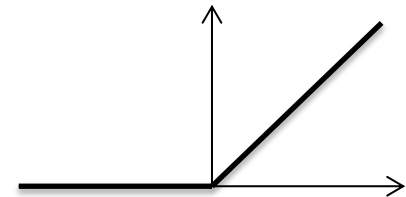
- Algorytm uczenia maszynowego
- Inspiracja: próba naśladowania mózgu
- Są uniwersalne
- Mogą wykonywać bardzo złożone operacje
- Mogą uczyć się, które cechy danych wejściowych są istotne

Sieci neuronowe

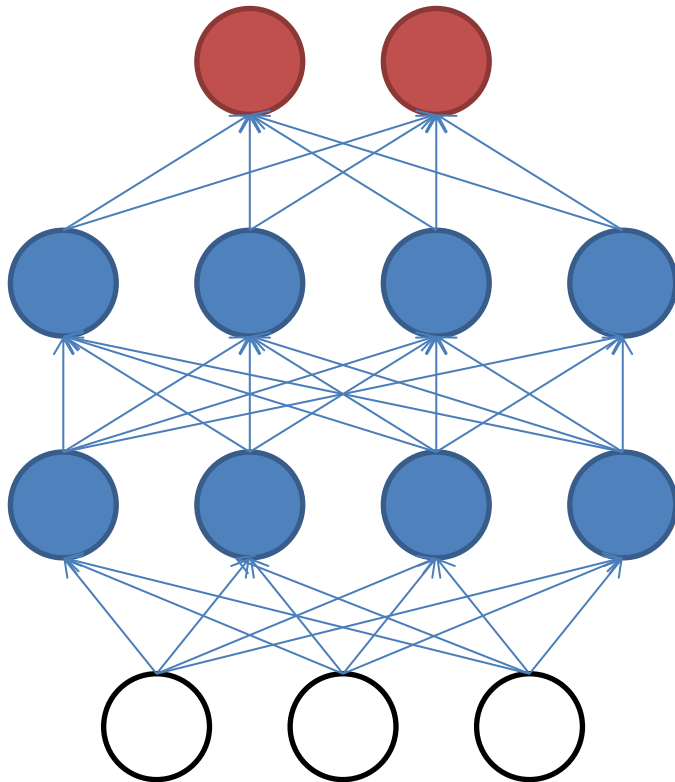


warstwy ukryte

● : suma ważona wejść,
dodatkowo funkcja nieliniowa
na wyjściu

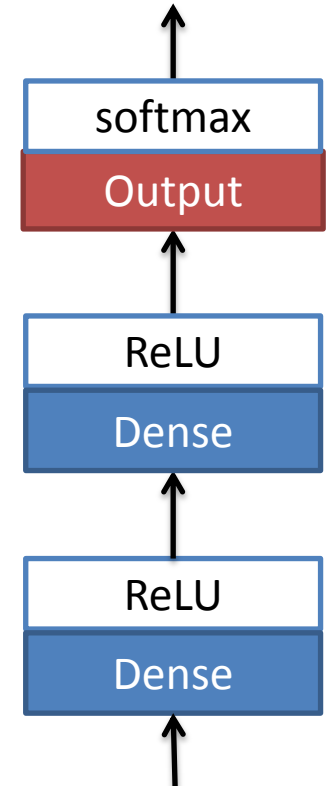
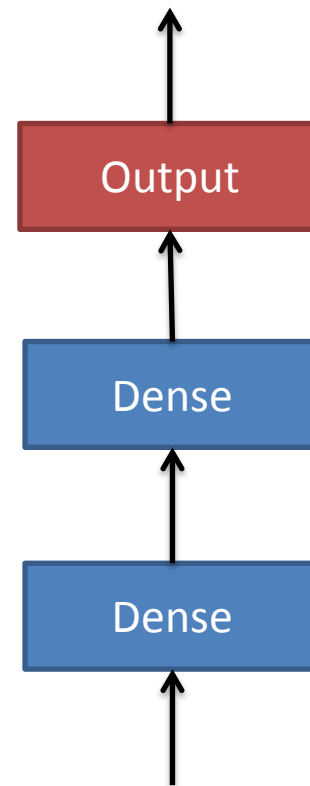
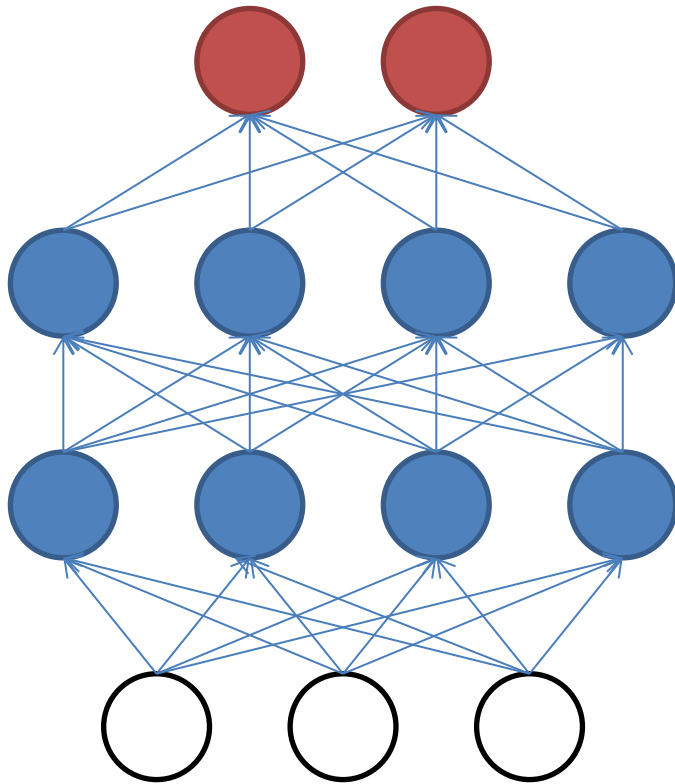


Sieci neuronowe – uczenie



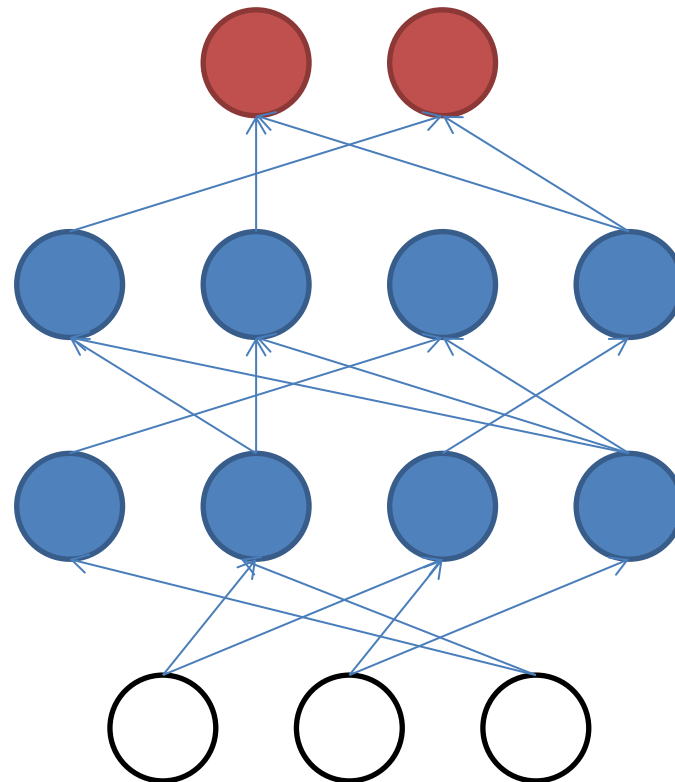
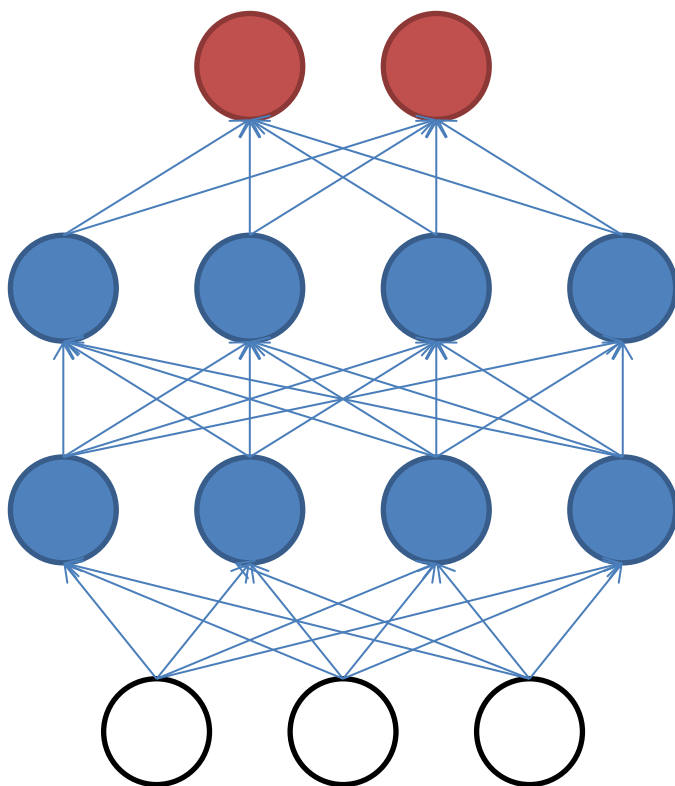
- Algorytm *propagacji wstecznej* (backpropagation)
- Inicjalizacja wag (losowa)
- Dla każdego przykładu treningowego:
 - Podajemy dane wejściowe
 - Obliczamy wynik działania sieci
 - Porównujemy z oczekiwanym wynikiem, obliczamy różnicę (błąd) na wyjściu
 - Błąd propagujemy wstecz, obliczamy o ile zmienić każdą wagę aby go zmniejszyć

Sieci neuronowe

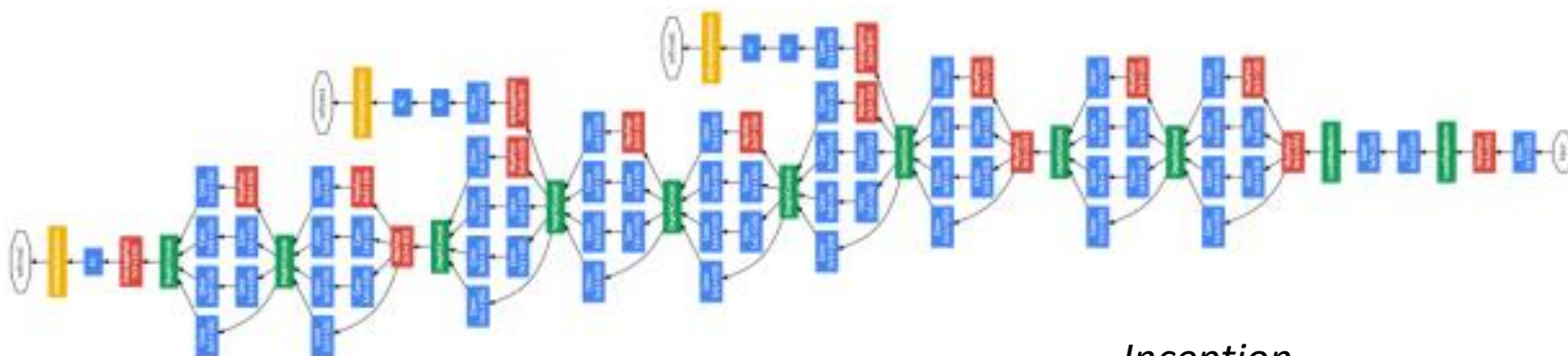
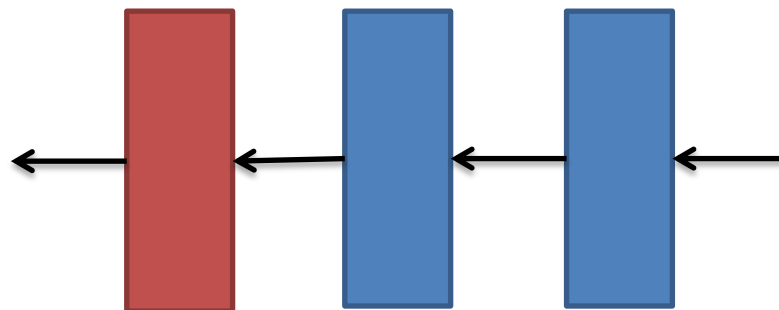


Sieć o n warstwach
 \approx program o n krokach

Dropout



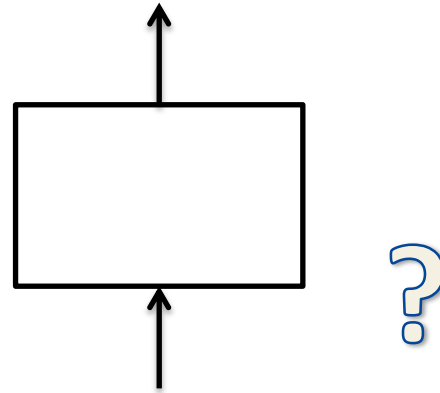
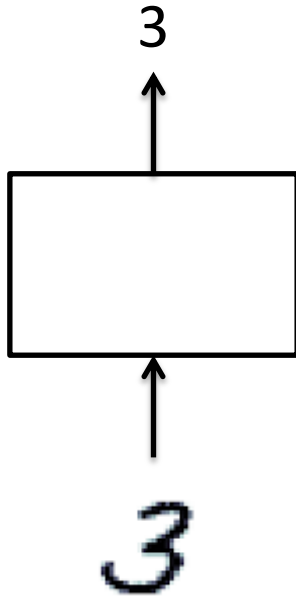
Głębokie sieci neuronowe



Inception

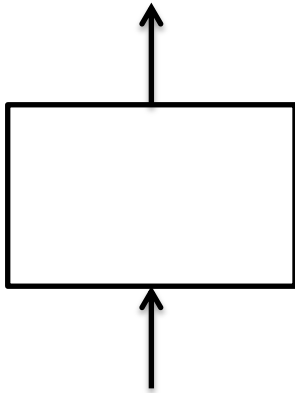
Część praktyczna: MNIST / Keras

Sieci rekurencyjne

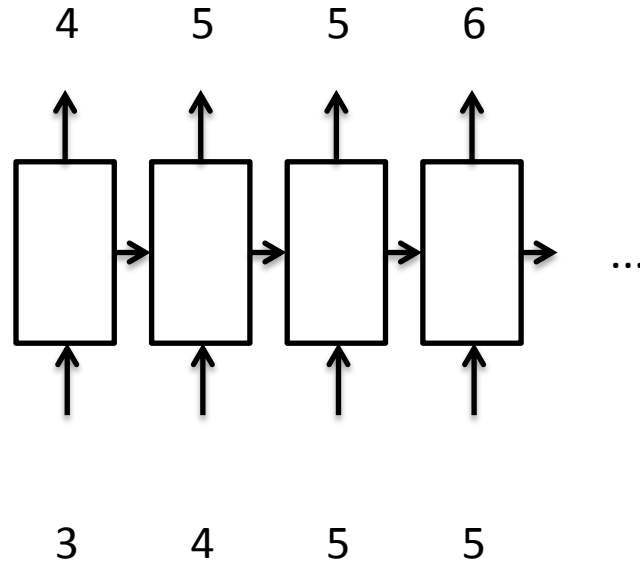


3, 4, 5, 5, 6, 4, 3, 0, 1, 2, 3, 3, 4, ...

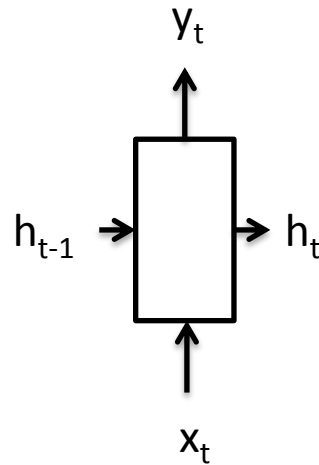
Sieci rekurencyjne



3, 4, 5, 5, 6, 4, 3, 0, 1, 2, 3, 3, 4, ...

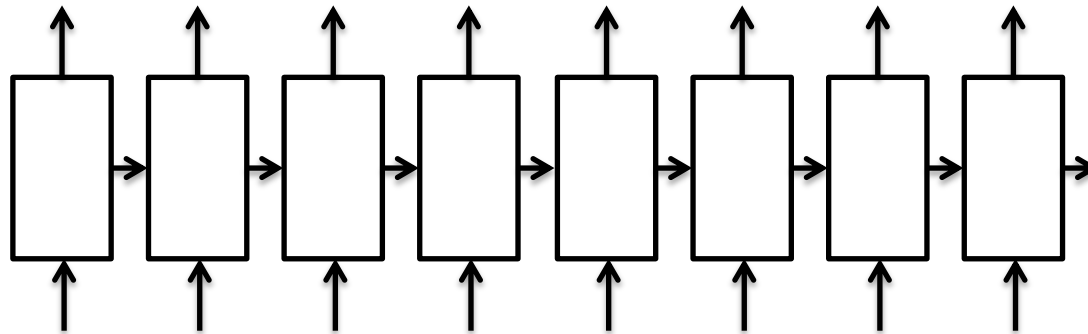


Sieci rekurencyjne

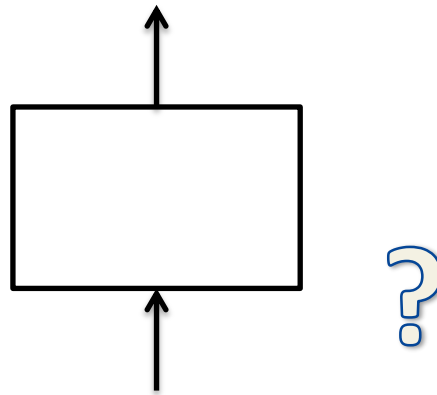


$$h_t = f(\mathbf{W} \cdot [h_{t-1}, x_t] + b_W)$$
$$y_t = g(\mathbf{V} \cdot h_t + b_V)$$

\mathbf{W}, \mathbf{V} – macierze wag
 b_W, b_V – wektory
 f, g – funkcje nieliniowe,
np. $f = \tanh, g = \text{softmax}$

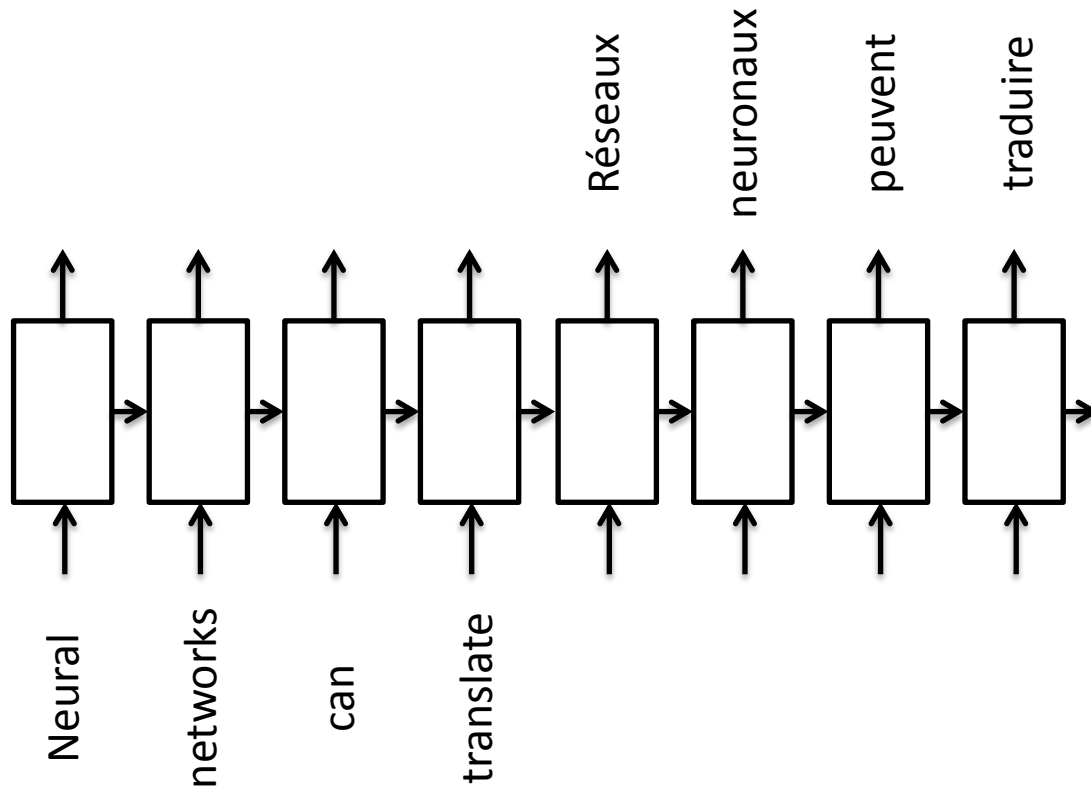


Sieci rekurencyjne

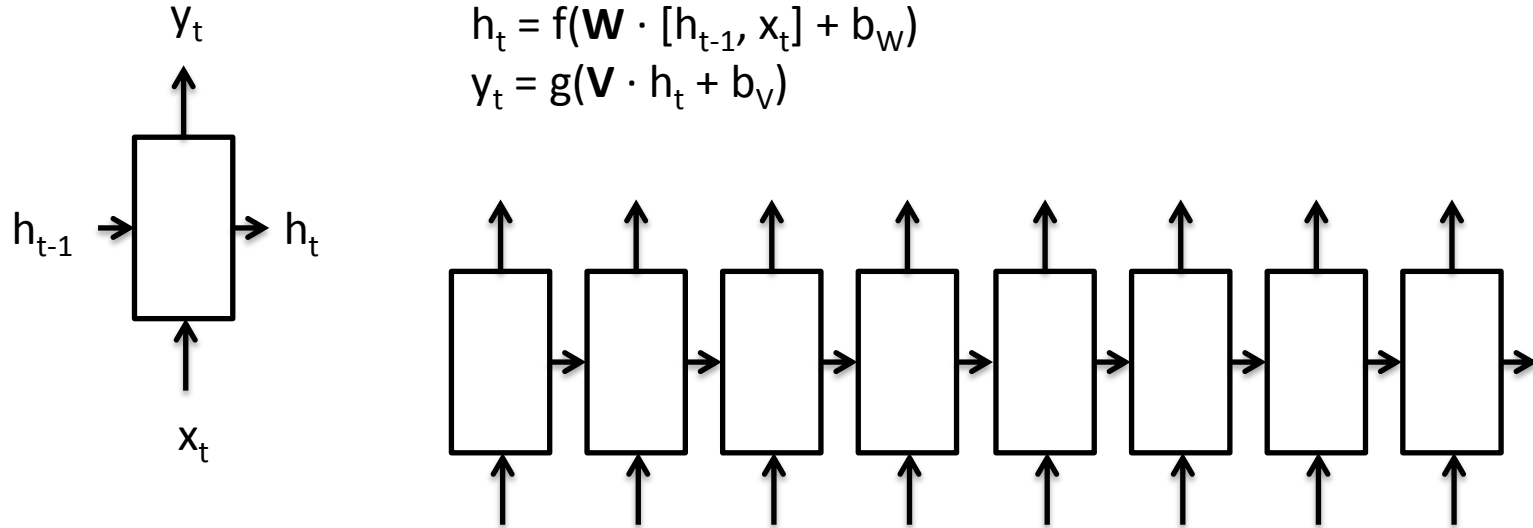


Neural networks can translate text from English to French.

Sieci rekurencyjne



Sieci rekurencyjne



Problem 1: nawet gdy $x = 0$, stan h zmienia się w czasie.

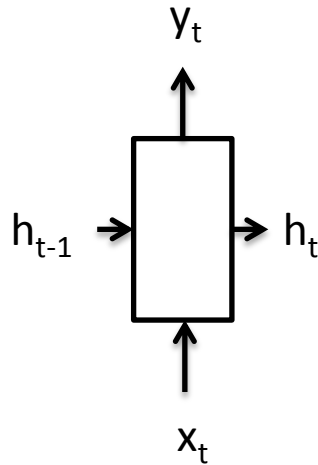
Problem 2: informację o błędzie trudno przenieść o wiele kroków wstecz.

Rozwiązanie: zmodyfikowane komórki RNN, takie jak:

LSTM – Long Short-Term Memory

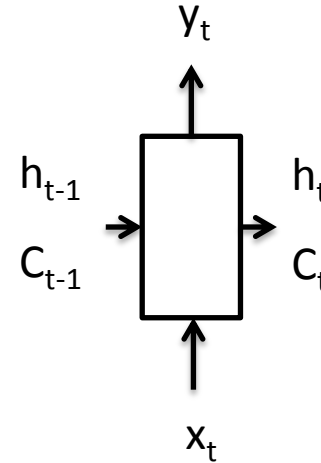
GRU – Gated Recurrent Unit

Sieci rekurencyjne LSTM



Stan:

$$h_t = \tanh(\mathbf{W} \cdot [h_{t-1}, x_t] + b_W)$$



Stan:

$$h_t = o_t \cdot \tanh(C_t)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tanh(\mathbf{W}_C \cdot [h_{t-1}, x_t] + b_C)$$

Bramki:

$$o_t = \sigma(\mathbf{W}_o \cdot [h_{t-1}, x_t] + b_o)$$

$$i_t = \sigma(\mathbf{W}_i \cdot [h_{t-1}, x_t] + b_i)$$

$$f_t = \sigma(\mathbf{W}_f \cdot [h_{t-1}, x_t] + b_f)$$

Modele języka

- Model języka przypisuje prawdopodobieństwo ciągowi słów

Anna poszła do kiosku i kupiła gazetę. $p = ?$

$$p = p(\text{Anna}) \cdot p(\text{poszła} | \text{Anna}) \cdot p(\text{do} | \text{Anna poszła}) \cdot \dots \\ \cdot p(. | \text{Anna poszła do kiosku i kupiła gazetę})$$

- Pozwala na porównywanie alternatywnych wersji zdania:

$$p(\text{Zaczyna się wiosna}) > p(\text{Zaczyna się wiosła})$$

Modele języka – zastosowania

- Rozpoznawanie mowy
- Rozpoznawanie tekstu, pisma (OCR)
- Tłumaczenie maszynowe
- Korekcja błędów
- Wprowadzanie tekstu
- ...

Modele języka – jak zbudować

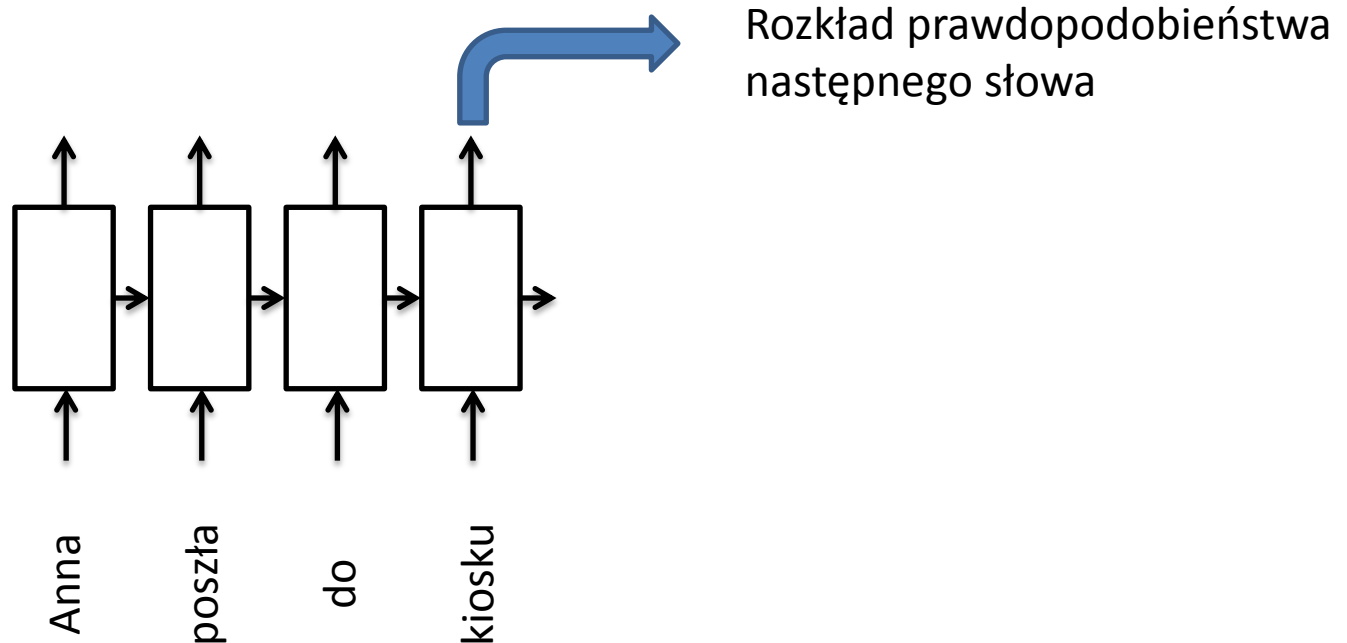
- $p(w_1, w_2, \dots w_N) = p(w_1) \cdot p(w_2 | w_1) \cdot$
 $\cdot p(w_3 | w_2, w_1) \cdot \dots \cdot p(w_N | w_1, w_2, \dots w_{N-1})$
- Model unigramowy:
 $p \approx p(w_1) \cdot p(w_2 | w_1) \cdot p(w_3 | w_2) \cdot \dots$
 $\cdot p(w_N | w_{N-1})$
- Model bigramowy:
 $p(x | y_1, y_2, \dots y_k) \approx p(x | y_1, y_2)$
- Prawdopodobieństwa szacowane z *korpusu*:
 $p(w_3 | w_1, w_2) = n(w_1, w_2, w_3) / n(w_1, w_2)$

n -gramowe modele języka

- Zalety: dobre skalowanie, szybkość
- Problemy:
 - dla większych n wiele n -gramów w ogóle nie wystąpi w korpusie (krótki kontekst)
 - nie uwzględniają pokrewieństwa słów:
np. poszedł \leftrightarrow pojechał; kot \leftrightarrow pies

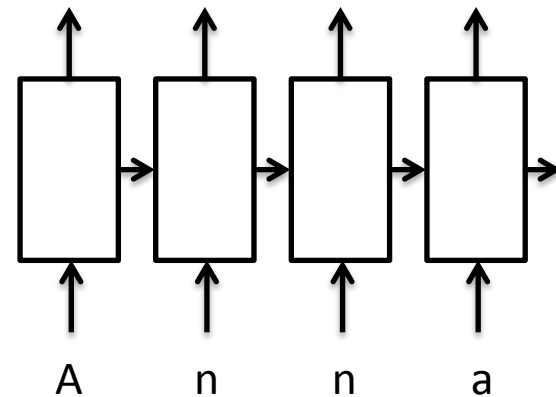
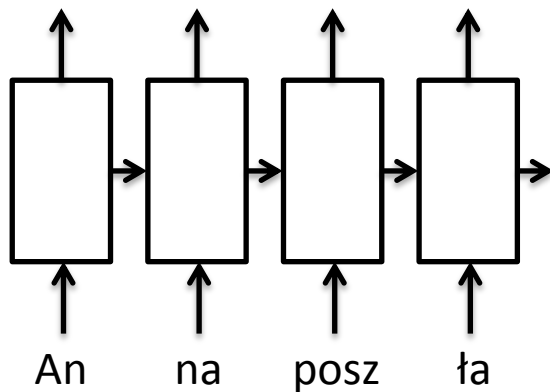
Modele języka oparte na RNN

- Kolejne słowa są wejściem do RNN w kolejnych krokach czasowych

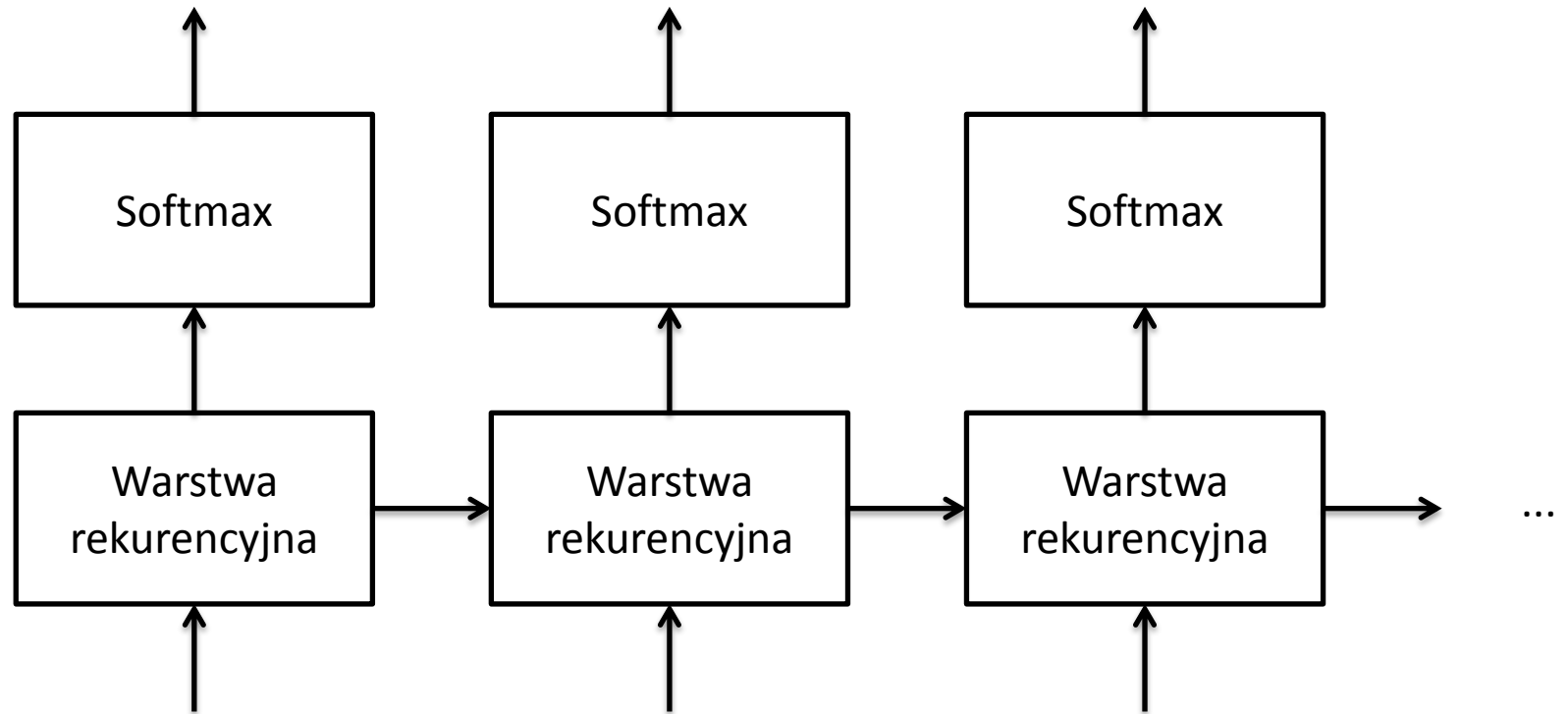


Modele języka oparte na RNN

- Mogą uchwycić nawet długie zależności (nie mają ograniczenia do n słów kontekstu)
- Mogą działać na poziomie słów, morfemów, znaków...



Modele języka oparte na RNN



PANDARUS:

Alas, I think he shall be come approached and the day
When little strain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:

They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:

Well, your wit is in the care of side and that.

Second Lord:

They would be ruled after this chamber, and
my fair nues begun out of the fact, to be conveyed,
Whose noble souls I'll have the heart of the wars.

Clown:

Come, sir, I will make did behold your worship.

VIOLA:

I'll drink it.

For $\bigoplus_{n=1, \dots, m} \mathcal{L}_{m,n} = 0$, hence we can find a closed subset \mathcal{H} in \mathcal{H} and any sets \mathcal{F} on X , U is a closed immersion of S , then $U \rightarrow T$ is a separated algebraic space.

Proof. Proof of (1). It also start we get

$$S = \mathrm{Spec}(R) = U \times_X U \times_X U$$

and the comparicoly in the fibre product covering we have to prove the lemma generated by $\coprod Z \times_U U \rightarrow V$. Consider the maps M along the set of points Sch_{fppf} and $U \rightarrow U$ is the fibre category of S in U in Section, ?? and the fact that any U affine, see Morphisms, Lemma ?? . Hence we obtain a scheme S and any open subset $W \subset U$ in $Sh(G)$ such that $\mathrm{Spec}(R') \rightarrow S$ is smooth or an

$$U = \bigcup U_i \times_{S_i} U_i$$

which has a nonzero morphism we may assume that f_i is of finite presentation over S . We claim that $\mathcal{O}_{X,x}$ is a scheme where $x, x', s'' \in S'$ such that $\mathcal{O}_{X,x'} \rightarrow \mathcal{O}'_{X',x'}$ is separated. By Algebra, Lemma ?? we can define a map of complexes $\mathrm{GL}_{S'}(x'/S'')$ and we win. \square

To prove study we see that $\mathcal{F}|_U$ is a covering of \mathcal{X}' , and \mathcal{T}_i is an object of $\mathcal{F}_{X/S}$ for $i > 0$ and \mathcal{F}_p exists and let \mathcal{F}_i be a presheaf of \mathcal{O}_X -modules on \mathcal{C} as a \mathcal{F} -module. In particular $\mathcal{F} = U/\mathcal{F}$ we have to show that

$$\widetilde{M}^\bullet = \mathcal{I}^\bullet \otimes_{\mathrm{Spec}(k)} \mathcal{O}_{S,s} - i_X^{-1} \mathcal{F}$$

is a unique morphism of algebraic stacks. Note that

$$\mathrm{Arrows} = (Sch/S)_{fppf}^{opp}, (Sch/S)_{fppf}$$

and

$$V = \Gamma(S, \mathcal{O}) \mapsto (U, \mathrm{Spec}(A))$$

is an open subset of X . Thus U is affine. This is a continuous map of X is the inverse, the groupoid scheme S .

Proof. See discussion of sheaves of sets. \square

The result for prove any open covering follows from the less of Example ?? . It may replace S by $X_{spaces, \acute{e}tale}$ which gives an open subspace of X and T equal to S_{Zar} , see Descent, Lemma ?? . Namely, by Lemma ?? we see that R is geometrically regular over S .

< *S* > With even more new technologies coming onto the market quickly during the past three years , an increasing number of companies now must tackle the ever-changing and ever-changing environmental challenges online . < *S* > Check back for updates on this breaking news story . < *S* > About 800 people gathered at Hever Castle on Long Beach from noon to 2pm , three to four times that of the funeral cortège . < *S* > We are aware of written instructions from the copyright holder not to , in any way , mention Rosenberg 's negative comments if they are relevant as indicated in the documents , " eBay said in a statement . < *S* > It is now known that coffee and cacao products can do no harm on the body . < *S* > Yuri Zhirkov was in attendance at the Stamford Bridge at the start of the second half but neither Drogba nor Malouda was able to push on through the Barcelona defence .

Kodowanie wejścia do sieci

- znaki – one-hot:

$\{a, b, c, d, \dots, z, \text{ `` } \}$ n różnych znaków

$a \rightarrow [1, 0, 0, \dots, 0]$ długość = n

$b \rightarrow [0, 1, 0, \dots, 0]$

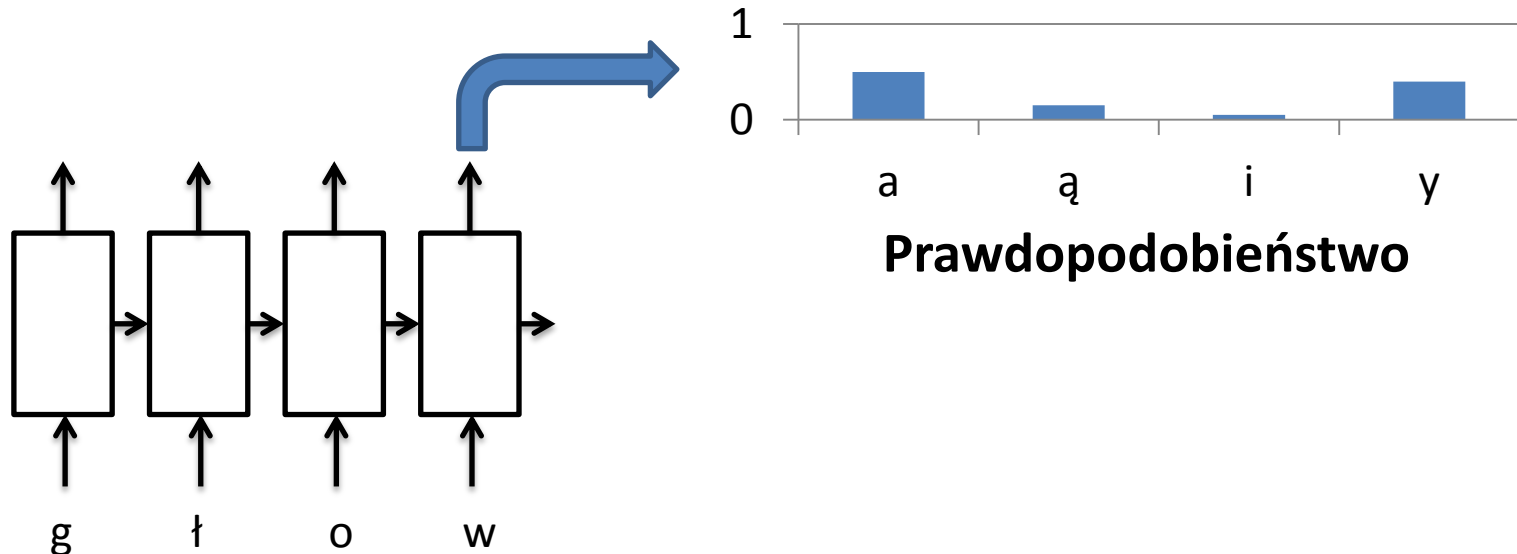
$z \rightarrow [0, 0, \dots, 1, 0]$

$\text{ `` } \rightarrow [0, 0, \dots, 0, 1]$

- fragment tekstu: macierz o wymiarach
długość tekstu $\times n$

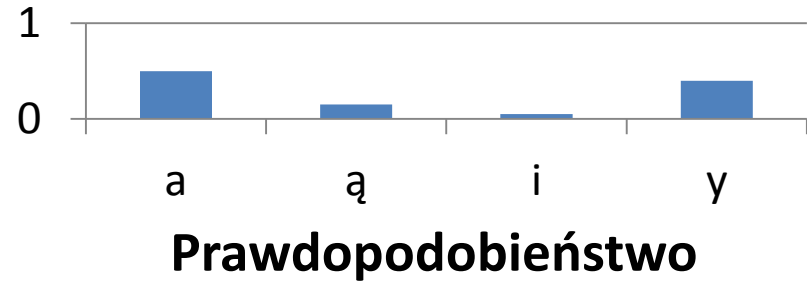
Dekodowanie

- = generowanie tekstu z modelu
- „zachłanne” (*greedy*) – zawsze wybieramy znak o największym prawdopodobieństwie, deterministyczne

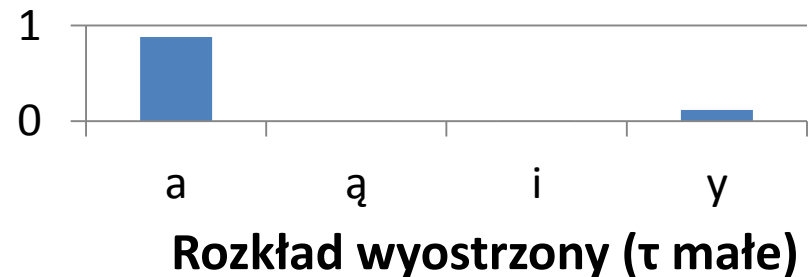
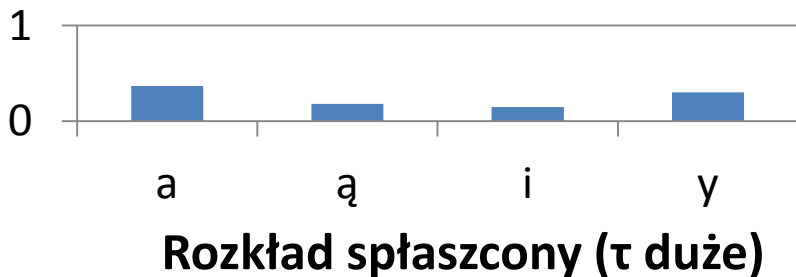


Dekodowanie

- losowanie z rozkładu



- losowanie z przekształconego rozkładu



Część praktyczna: modele języka