Przetwarzanie języka naturalnego/04

2024-03-21

Krzysztof Misztal
misztal.edu.pl

Spis treści

Spis treści

Prawdopodobieństwo zdań

- tłumaczenie "mam dobre oceny":
  P("i have **good** grades") > P("I have **righteous** grades")
- korekcja
  P("I'll be late fifteen **minutes**") > P("I'll be late fifteen **mints**")
- rozpoznawanie mowy
  P("I saw rice bowl") > P("eye sour ice bow")

Prawdopodobieństwo zdań

- Cel: obliczyć prawdopodobieństwo zdań

$$P(W) = P(w_1, ..., w_n)$$

- Podobne: prawdopodobieństwa kolejnego słowa

$$P(w_k | w_1, ..., w_{k-1})$$

- Dowolny z takich rozkładów $P(W)$ lub $P(w_k | w_1, ..., w_{k-1})$ nazywamy **statystycznym modelem języka**.

Jak to policzyć?

- Cel: policzyć prawdopodobieństwo (joint probability)

$$P(w_1 = \text{My}, w_2 = \text{dog}, w_3 = \text{is}, w_4 = \text{the}, w_5 = \text{best})$$

Jak to policzyć?

- Cel: policzyć prawdopodobieństwo (joint probability)

$$P(w_1 = \text{My}, w_2 = \text{dog}, w_3 = \text{is}, w_4 = \text{the}, w_5 = \text{best})$$

- Intuicja: Polegamy na zasadzie łańcuchowej prawdopodobieństwa (prawdopodobieństwo całkowite)

Jak to policzyć?

- Cel: policzyć prawdopodobieństwo (joint probability)

$$P(w_1 = \text{My}, w_2 = \text{dog}, w_3 = \text{is}, w_4 = \text{the}, w_5 = \text{best})$$

- Intuicja: Polegamy na zasadzie łańcuchowej prawdopodobieństwa (prawdopodobieństwo całkowite)

$$P(A, B, C, D) = P(A)$$

Jak to policzyć?

- Cel: policzyć prawdopodobieństwo (joint probability)

$$P(w_1 = \text{My}, w_2 = \text{dog}, w_3 = \text{is}, w_4 = \text{the}, w_5 = \text{best})$$

- Intuicja: Polegamy na zasadzie łańcuchowej prawdopodobieństwa (prawdopodobieństwo całkowite)

$$P(A, B, C, D) = P(A)P(B|A)$$

Jak to policzyć?

- Cel: policzyć prawdopodobieństwo (joint probability)

$$P(w_1 = \text{My}, w_2 = \text{dog}, w_3 = \text{is}, w_4 = \text{the}, w_5 = \text{best})$$

- Intuicja: Polegamy na zasadzie łańcuchowej prawdopodobieństwa (prawdopodobieństwo całkowite)

$$P(A, B, C, D) = P(A)P(B|A)P(C|A, B)$$

Jak to policzyć?

- Cel: policzyć prawdopodobieństwo (joint probability)

$$P(w_1 = \mathrm{My}, w_2 = \mathrm{dog}, w_3 = \mathrm{is}, w_4 = \mathrm{the}, w_5 = \mathrm{best})$$

- Intuicja: Polegamy na zasadzie łańcuchowej prawdopodobieństwa (prawdopodobieństwo całkowite)

$$P(A, B, C, D) = P(A)P(B|A)P(C|A, B)P(D|A, B, C)$$

Zasada łańcuchowa [NLP]

$$P(w_1, \ldots, w_n) = \prod_i P(w_i | w_1, \ldots, w_{i-1})$$

Zasada łańcuchowa [NLP]

$$P(w_1, \dots, w_n) = \prod_i P(w_i | w_1, \dots, w_{i-1})$$

$$P(\text{"My dog is the best"}) =$$
$$= P(\text{"My"}) \cdot P(\text{"dog"}|\text{"My"}) \cdot P(\text{"is"}|\text{"My dog"}) \cdot$$
$$\cdot P(\text{"the"}|\text{"My dog is"}) \cdot P(\text{"best"}|\text{"My dog is the"})$$

Estymacja [NLP]

$$\hat{P}(w_i|w_1, \ldots, w_{i-1})$$

Estymacja [NLP]

Może tak jak w przypadku NB?

$$\frac{count((w_1, \ldots, w_i)) + \alpha}{count((w_1, \ldots, w_{i-1}))}$$

Estymacja [NLP]

Może tak jak w przypadku NB?

$$\frac{count((w_1, \ldots, w_i)) + \alpha}{count((w_1, \ldots, w_{i-1}))}$$

Złe rozwiązanie - w praktyce prawie zawsze licznik/mianownik będzie równy 0 (α).

Własność Markova

Markov property

A stochastic process has the Markov property if the conditional probability distribution of **future** states of the process (conditional on both past and present values) **depends only upon the present** state; that is, given the present, the future does not depend on the past. A process with this property is said to be Markovian or a Markov process.

Założenie o Markowości

$$P(w_1, \ldots, w_n) \sim \prod_i P(w_i | w_{i-k}, \ldots, w_{i-1})$$

N-gramy

N-gram language model

Model języka złożony z aproksymacji zakładającej własność Markova o $n-1$ elementach historii nazywamy modelem n-gramowym

N-gramy

- N $= 1 \Rightarrow$ **uni**gram model

$$P(w_1, \ldots, w_n) \sim \prod_i P(w_i)$$

- N $= 2 \Rightarrow$ **bi**gram model

$$P(w_1, \ldots, w_n) \sim \prod_i P(w_i | w_{i-1})$$

- N $= 3 \Rightarrow$ **tri**gram model

$$P(w_1, \ldots, w_n) \sim \prod_i P(w_i | w_{i-2}, w_{i-1})$$

- ...

Model języka?

Dlaczego mówimy, ze to prawdopodobieństwo jest modelem języka?

$$P(w_1, \ldots, w_n) \sim \prod_i P(w_i)$$

`fifth, an, of, futures, the, an, incorporated, a, a, the, inflation, most, dollars, quarter, in, is, mass`

`thrift, did, eighty, said, hard, 'm, july, bullish`

`that, or, limited, the`

Model języka?

Dlaczego mówimy, ze to prawdopodobieństwo jest modelem języka?

$$P(w_1, \ldots, w_n) \sim \prod_i P(w_i | w_{i-1})$$

`texaco, rose, one, in, this, issue, is, pursuing, growth, in, a, boiler, house, said, mr., gurria, mexico, 's, motion, control, proposal, without, permission, from, five, hundred, fifty, five, yen`

`outside, new, car, parking, lot, of, the, agreement, reached`

`this, would, be, a, record, november`

Jak to się ma do języka?

- Jest to dalekie od wymodelowania pełnej złożoności języka
  "My dog which has gone missing after the party is now safe"
- Sprawdza się całkiem dobrze w praktyce

Naive Bayes a modele języka

Naive Bayes jako model języka

Zakładając, ze w Naive Bayesie używamy wyłącznie słów jako cech, to buduje on **unigamowy** model języka używanego w każdej z analizowanych klas.

Ciekawostka

https://pdos.csail.mit.edu/archive/scigen/

Spis treści

Estymacja prawdopodobieństw

Ponownie, zacznijmy od MLE

$$\hat{P}(w_i|w_{i-1}) = \frac{count(w_{i-1}, w_i)}{count(w_{i-1})}$$

Uwaga: W dalszej części skracamy *count* do c.

Estymacja prawdopodobieństw – przykład

$$\hat{P}(w_i|w_{i-1}) = \frac{c(w_{i-1}, w_i)}{c(w_{i-1})}$$

I like dogs
dogs like dogs
dogs are like cats

Estymacja prawdopodobieństw – przykład

$$\hat{P}(w_i|w_{i-1}) = \frac{c(w_{i-1}, w_i)}{c(w_{i-1})}$$

@b I like dogs @e
@b dogs like dogs @e
@b dogs are like cats @e

Estymacja prawdopodobieństw – przykład

$$\hat{P}(w_i|w_{i-1}) = \frac{c(w_{i-1}, w_i)}{c(w_{i-1})}$$

@b I like dogs @e
@b dogs like dogs @e
@b dogs are like cats @e

$$\hat{P}(I|@b) = \frac{1}{3}$$

Estymacja prawdopodobieństw – przykład

$$\hat{P}(w_i|w_{i-1}) = \frac{c(w_{i-1}, w_i)}{c(w_{i-1})}$$

@b **I like** dogs @e
@b dogs like dogs @e
@b dogs are like cats @e

$$\hat{P}(I|@b) = \frac{1}{3},\ \hat{P}(like|I) = \frac{1}{1} = 1$$

Estymacja prawdopodobieństw – przykład

$$\hat{P}(w_i|w_{i-1}) = \frac{c(w_{i-1}, w_i)}{c(w_{i-1})}$$

@b **I like dogs** @e
@b dogs like dogs @e
@b dogs are like cats @e

$$\hat{P}(I|@b) = \frac{1}{3},\ \hat{P}(like|I) = 1,\ \hat{P}(dogs|like) = \frac{2}{3}$$

Estymacja prawdopodobieństw – przykład

@b I like dogs @e
@b dogs like dogs @e
@b dogs are like cats @e
$\hat{P}(I|@b) = \frac{1}{3},\ \hat{P}(like|I) = 1,\ \hat{P}(dogs|like) = \frac{2}{3}$
$\hat{P}(cats|like) = \frac{1}{3},\ \hat{P}(@e|cats) = \frac{1}{3}$

$$\hat{P}(@b\ I\ like\ cats\ @e) = \hat{P}(I|@b)\hat{P}(like|I)\hat{P}(cats|like)\hat{P}(@e|cats) =$$

$$= \frac{1}{27} \approx 0.037$$

Do własnego przejrzenia

`https://www.kaggle.com/alvations/n-gram-language-model-with-nltk`

Dziękuję za uwagę.