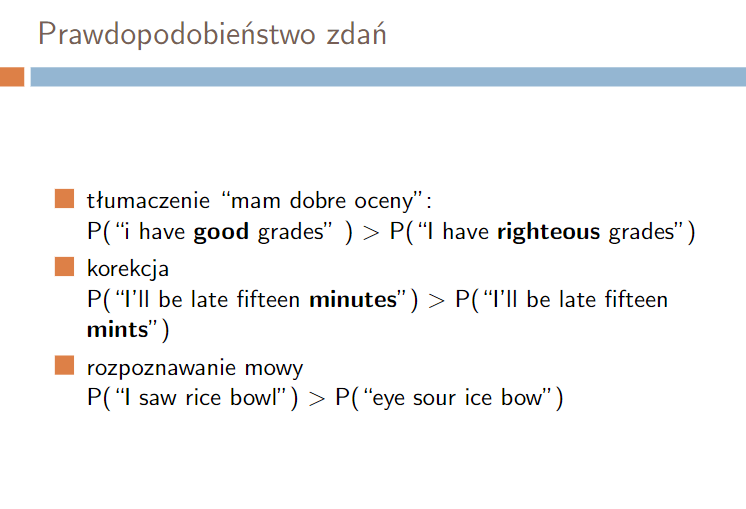
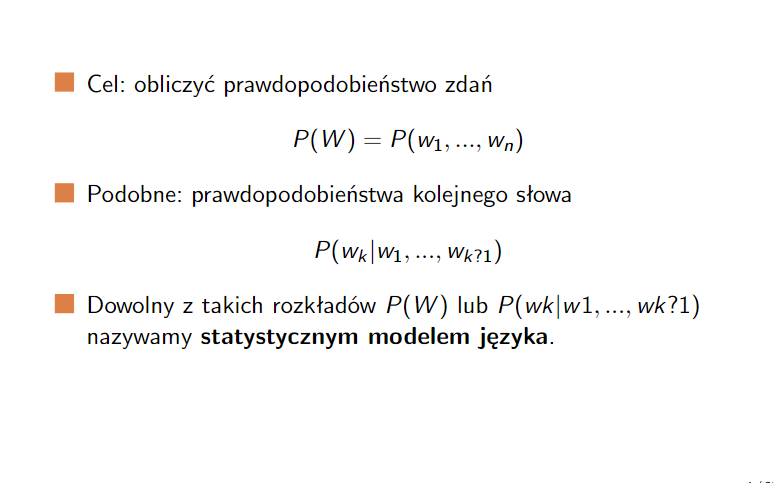
Przetwarzanie języka naturalnego 4.

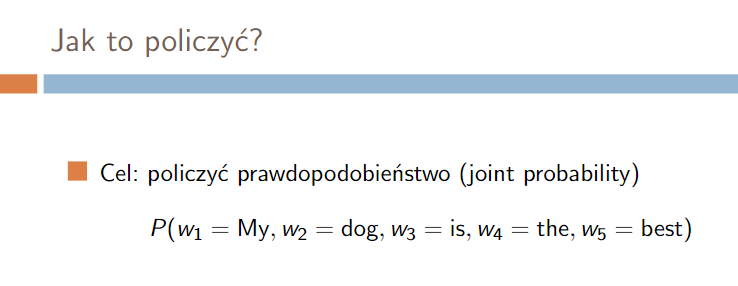
Statystyczne modele języka



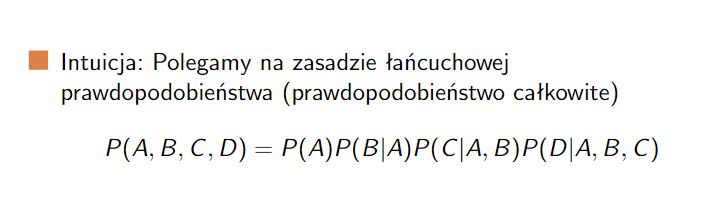
Mówi o ogarnięciu kontekstu słów, czyli o oszacowaniu prawdopodobieństwa zdań, i że są sposoby oszacowania tego.

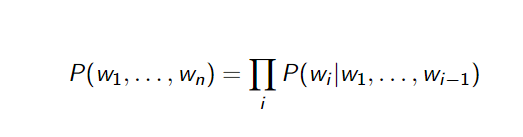


Celem jest oszacowania prawdopodobieństwa zdań, czyli prawdopodobieństwo zdania złożonego z kolejnych wyrazów, mamy P(W), gdzie W jest konkatenacją wyrazów będących w zdaniu (znaki zapytania na slajdzie wyżej oznaczają minusy, bo zrobił błąd na slajdzie) Można prościej oszacować jakie jest prawdopodobieństwo wystąpienia słowa Wk jeśli w podanym ciągu pojawiły się już wyrazy od W1 do Wk-1, czyli my jesteśmy w stanie, znając historię wyrazów poprzednich w łatwy sposób oszacować następny wyraz(przykładowo przewidywanie zapytania w Google, jak coś tam wpisujemy, to Google proponuje nam zapytanie) (Tutaj mówi coś o tym) jeżeli możemy oszacować jakie będzie następne zdanie w ciągu, to proces nazywamy statystycznym modelem języka

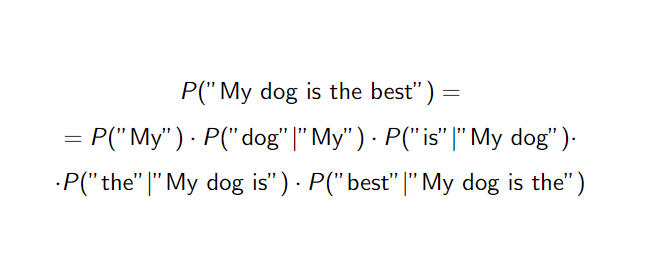


Mamy zdanie „My dog is the best” I mamy składowe od W1 do W5 słówka i jak tutaj widać, będziemy prawdopodobnie mogli skorzystać z jakiejś reguły łańcuchowej czyli tego prawdopodobieństwa całkowitego mówiącego ,

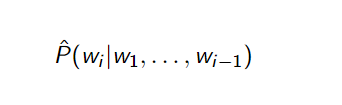
że jeżeli mamy kolejne zdarzenia A, B, C, D no ii wszystkie muszą zajść, no i jest to równoznaczne z P(A) pomnożone na P(A) że zaszło B \* P(C), że zaszło A i B \* P(D) , że zaszło A i B i C. Czyli takie prawdopodobieństwo (P(A,B, C, D) ) da się przełożyć na taki iloczyn prawdopodobieństw.



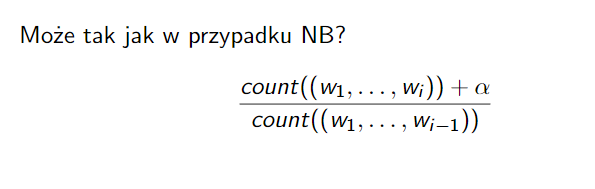
W związku z tym prawdopodobieństwo naszego zdania jest taki iloraz prawdopodobieństw po kolei (П - zaznaczenie tego ilorazu)



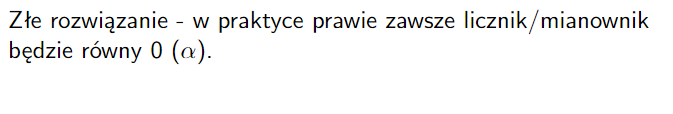
Tutaj podany przykład został literalnie przepisany , ale przy wiec skomplikowanych i długich zdaniach ten „ogon” iloczynu będzie dłuższy. (Ten przykład będzie wyjaśniony niżej)

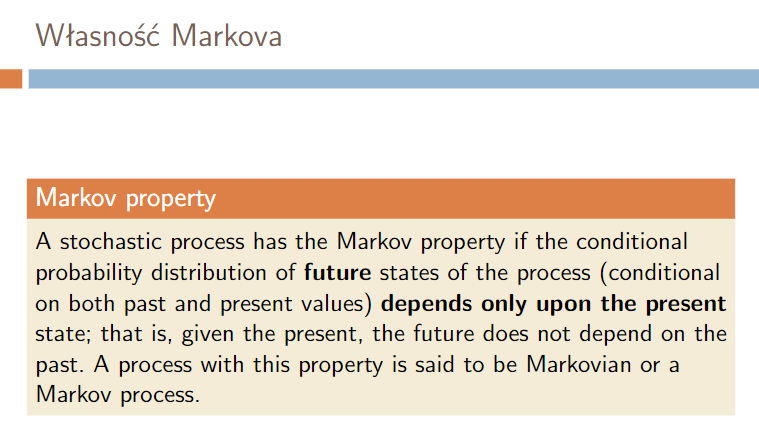


To jak estymujemy to prawdopodobieństwo wyżej? (Czyli jak to w ogóle liczymy)

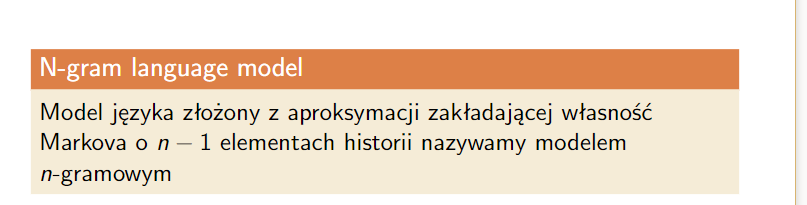


Proponuje wykorzystać Naiwnego Baesa (o nim było na innym wykładzie),

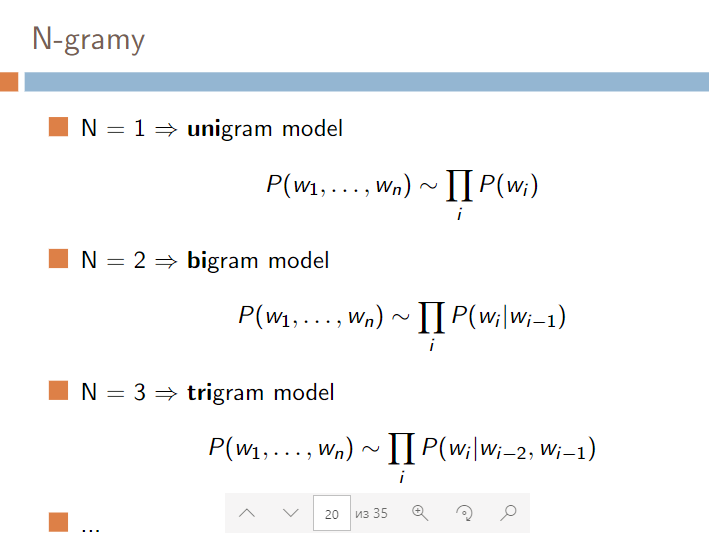
ale może się zdarzyć że licznik lub mianownik = 0 i nie możemy NB już używać.



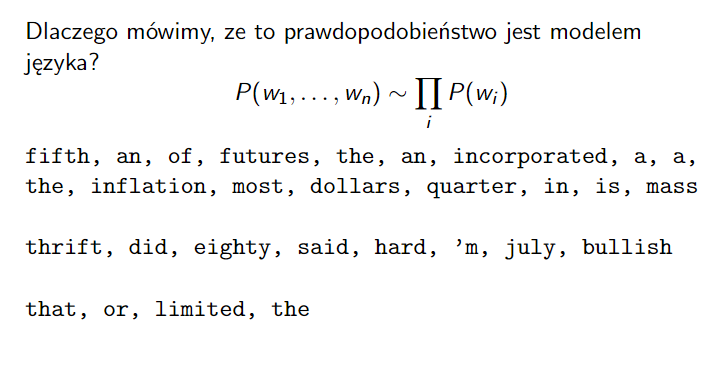
Łańcuchy Markowa –zasada taka: patrzeć nie na całą historie wyrazu(czyli nie na wszystkie poprzednie słówka) , a tylko na n elementów wstecz (czyli na n słów wcześniej) , czyli nie analizować całego zdania od początku do końca, tylko będziemy analizować ilość wyrazów wstecz.



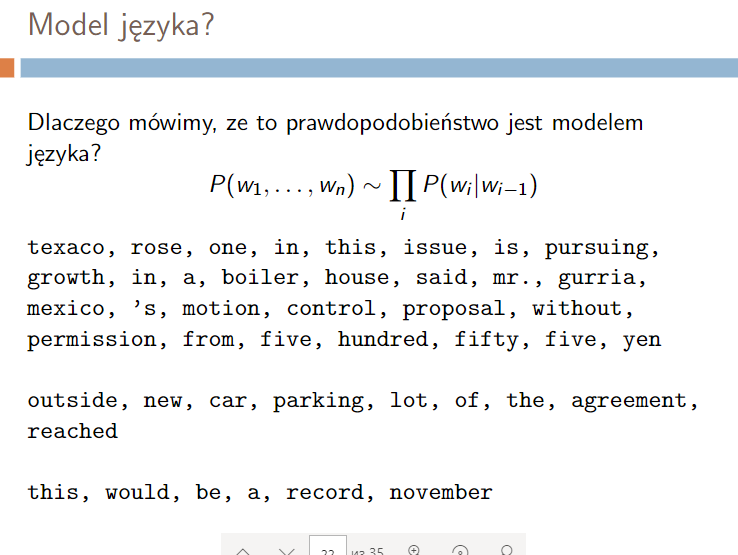
I ten N-gramowy model będziemy właśnie traktować w ten sposób , że będziemy patrzeć na te N elementów



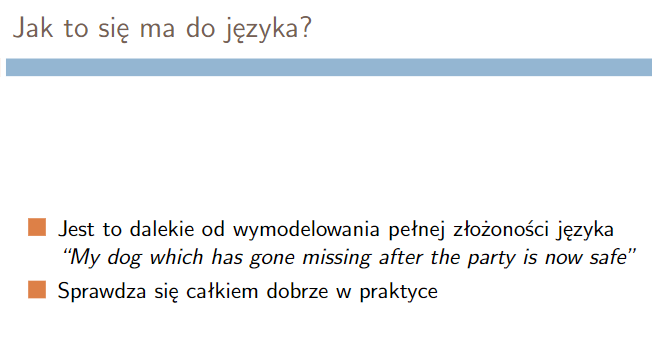
Podstawowy model – to N=1, czyli model unigramowy , który oszacowuje prawdopodobieństwo zdania jako P w którym bierze się pod uwagę pojedyncze wyrazy i nie patrząc na wyrazy występujące do tego. N = 2 (bigramowy model) dla każdego wyrazu patrzymy na wyraz , który był do tego i t.d. Im bardziej specyficzne i wieloelementowe N-gramy - tym lepiej (Można brać pod uwagę 5 wyrazów, ale oczywiście można próbować wykorzystać bardziej potężny model, czyli brać pod uwagę więcej wyrazów)



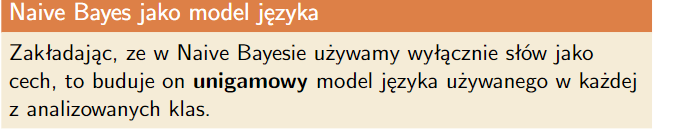
Weźmy przykład ciągu słów wyżej i jak można zrozumieć, patrzymy na prawdopodobieństwo tylko poszczególnych słów , więc mając dowolne zdanie , policzyli sobie 15 występowań i dalibyśmy radę coś oszacować. Pytanie jest na ile te przewidywanie jest tak naprawdę dobre.



Inny przykład, kiedy możemy zmienić kontekst , czyli nasz język , podawane z korpusu przykłady również będą modelować i zmieniać nasze prawdopodobieństwo , dodatkowo , jeżeli będziemy brali pod uwagę jakąś historię (stany poprzedzające nasze wyrazy) , to dokładniej będziemy odwzorowywać, czyli będziemy oszacowywać więcej realne wyniki niż w poprzednim przykładzie.

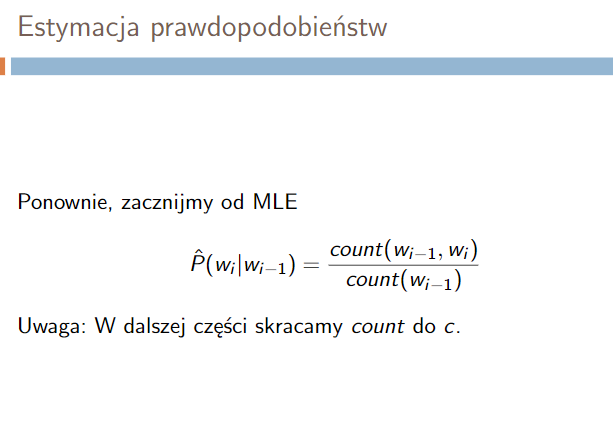


No i te modele N- gramy się nadają lepiej niż model NB. NB w szczególnych przypadkach jest odwzorowaniem tych modelów N-gram , czyli:

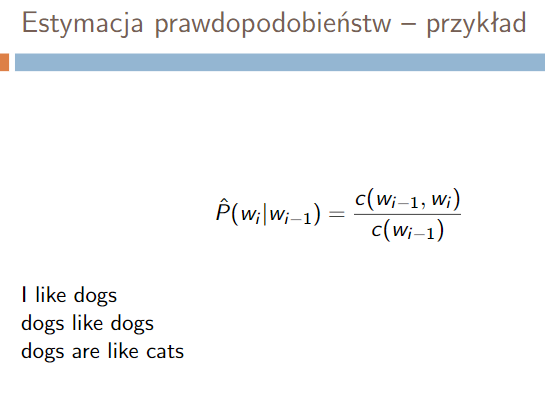


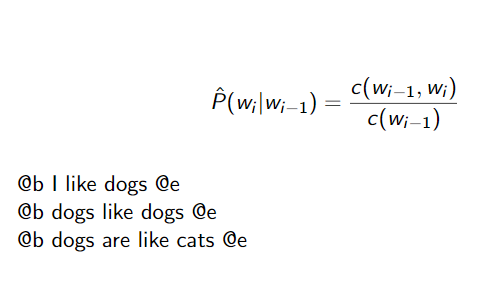
<https://pdos.csail.mit.edu/archive/scigen/> - podał link jako ciekawostkę(wygenerowania artykułów)

Markov – szacowanie prawdopodobieństwa

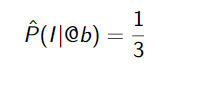


Bierzemy estymator największej wiarygodności , który oszacuje prawdopodobieństwo w modeli bigramowym. Czyli dzielimy ilość par analizowanych na ilość pierwszego el pary.

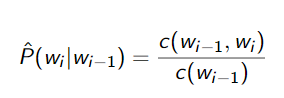


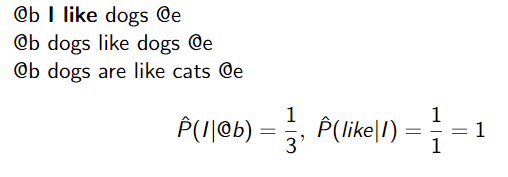
Wyżej przykład na którym pracujemy   


Dodajemy tagi, żeby oszacować początek i koniec zdania ich również bierzemy pod uwagę

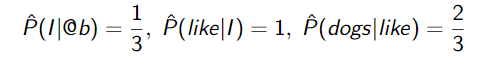


Prawdopodobieństwo że po tagu Begin będzie „I” = 1/3 ( bo mamy trzy zdania i tylko w 1 z tych zdań występuje „I” na początku (początki są w każdym zdaniu :p)



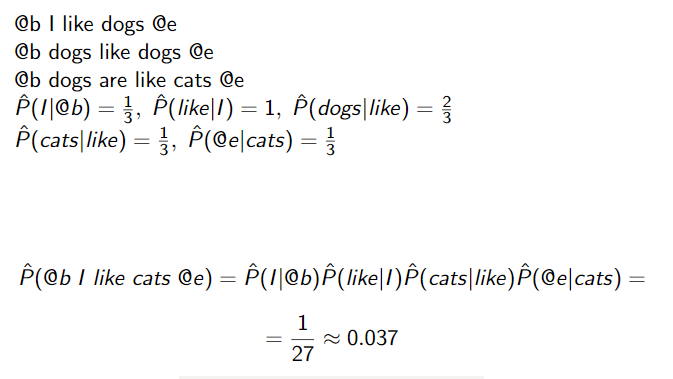


Z formuły wyżej patrzymy , „I” występuje 1 raz i „I like” również 1 raz ze wszystkich zdań , czyli 1/1 = 1



Tutaj „like dogs” występuje 2 razy w zdaniach, a „like” występuje 3 razy, czyli mamy P = 2/3

No i P(@e|dogs) = 2/4 , tak jak „dogs @e” mamy 2 razy, a ‘dogs’ = 4 razy



Tu podał przykład wyliczenia prawdopodobieństwa wyrazu „I like cats”

Dodał również stronę z info <https://www.kaggle.com/alvations/n-gram-language-model-with-nltk>

Tam jest realizacja tego wszystkiego w Pythonie