Wykład 6

sobota, 21 listopada 2020

13:07

Ukryty Model Markowa 
Rozważmy sygnalizację Świetlną na skrzyżowaniu 
Stop 
Prepare to 
prcxeed 
Prepare to stop 
Każdy stan jest zależny od poprzedniego, System jest 
deterministyczny. Możemy rozważać go jako proces Markowa. 

Slajd 4:

Model języka oparty o Model Markowa.

Rozważmy sygnalizację świetlna na skrzyżowaniu. Są 4 stany jakie sygnalizacja świetlna może generować.

Stop, żółte, zielone i znowu żółte.

Jeżeli my zauważymy światło, to wiemy z jakiego stanu przyszliśmy i do jakich świateł przejrzejmy w następnym kroku. Czyli w pełny, determistyczny sposób - obserwując aktualny stan, jesteśmy zauważyć gdzie jesteśmy, dokąd zmierzamy i skąd przyszliśmy. **Taki stan możemy rozważać jako proces Markowa.**

Ukryty Model Markowa 
Jkryty Vodel Markcwa 
Trojke (r, T, E) nazywamy Ukrytym Modelem Markowa 
(Hidden Markov Model. HMM). gdy jest æktorem rozkladu 
prawdopodobieöstwa rozmiaru n, T jest macierzq n x n, ktörej 
kaidy rzqd jest wektorem rozkladu prawdopodobiehstwa a E jest 
macierzq n x m o tej samej vvlasno'ci. 

Slajd 6:

Dla naszych potrzeb będziemy potrzebować zdefiniować sobie 3 elementy. Pierwszy element to jest **∏(pi**) i nazwiemy go sobie **wektorem rozkładu pozycji początkowych**. **T będzie macierzą transmisji i E to będzie macierz emisji.** W ukrytym modelu markowa będziemy się starali żebyśmy zawsze byli w stanie oszacować tę trójkę (macierze i wektor).

Podstawowe zadania HMM 
Dla zadanego HMM i zadanego ciągu obserwacji (Ol, . 
oceń jak jest to prawdopodobne zdarzenie 
Dla zadanego HMM i zadanego ciągu obserwacji (Ol, . 
znajdź najbardziej prawdopodobny ciąg stanÓN ukrytych 
(HI.....Hk) 
Dla zbioru obserwacji {(01. 
, —1 znajdź 
najbardziej prawdopodobny HMM o określonej strukturze 
O Dla zadanego zbioru HMM i Zadanego ciagu obserwacji 
( Ol, Ok) wskaz najbardziej prawdopodobny HMM 

Slajd 8:

**Ukryty model Markowa** tak naprawdę będzie nam definiował, ze względu na macierz pi, z jakich stanów będziemy zaczynać. To znaczy **jaka jest szansa występowania danego stanu.** Natomiast pozostałe dwie **macierze** będą opisywać w jaki sposób się **poruszać po taki grafie**. Ten graf jest zadany przez model Markowa i będziemy poruszać się pomiędzy tymi grafami zgodnie z regułami zapisanymi w macierzach - **graf skierowany**.

Ukryte modele Markowa zadania:

Mamy nasze macierze ∏ (pi), T, E i **stany od 1 do N** ( i pomiędzy tymi stanami chcemy się poruszać, jest to ciąg skończony). Jeżeli ustawimy sobie model markowa i będziemy na bieżąco notować w jakich stanach byliśmy, to może się okazać, że zostanie wygenerowana pewnego rodzaju **obserwacja**.

**Obserwacje notujemy pod literki od O1 do Ok.**

Na pierwszej obserwacji byliśmy w stanie 1, potem znowu byliśmy w stanie 1 i potem przeskoczyliśmy do stanu 5 i potem znowu wróciliśmy do stanu 1

Jeśli mamy to ciąg obserwacji, czyli mamy ciąg obserwacji 1,1,5,1, będących numerami kolejnych stanów w których byliśmy. Pierwsze zadanie będzie polegało na tym, jak bardzo prawdopodobne było uzyskanie takiej obserwacji. Czyli mamy cały układ, stany, mamy ukryty model markowa, który mówi nam jak mamy się poruszać po naszym grafie. I my zaczynamy w pewnym momencie obserwacje. Obserwujemy ileś tam chwil czasowych. Pierwsze pytanie: **Jak bardzo prawdopodobne było uzyskanie danej obserwacji?** Czy tę obserwację uzyskujemy w 90%, czy ta obserwacja jest bardzo rzadka? Należy sobie zdawać sprawę, że **im więcej stanów, tym prawdopodobieństwo każdej obserwacji może być niższe. Im dłuższa obserwacja, tym prawdopodobieństwo uzyskania danej obserwacji może być niższe.**

**Zadanie 1:** Jeśli mielibyśmy dwie obserwacje, to chcielibyśmy móc je ze sobą porównać i oszacować, która jest bardziej prawdopodobna. Czyli móc obliczyć prawdopodobieństwo ciągu obserwacji dla zdefiniowanego układu.

Nasze stany tak naprawdę, ze względu na HMM, będą one powiązane ze stanami ukrytymi.

**Zadanie 2:** Zastanawiamy się, mając nasze obserwacje, jak na te obserwacje wpływały stany, których nie widzimy. Będziemy się starali oszacować najbardziej prawdopodobny ciąg stanów ukrytych. Czyli tego, czego nie widzimy, który doprowadził do uzyskania danych obserwacji.

**Zadanie 3:** Mamy zbiór ciągu obserwacji i naszym zdaniem jest znalezienie najbardziej prawdopodobnego ukrytego modelu Markowa o zadanej strukturze. Czyli mamy ciąg obserwacji i musimy sami oszacować prawdopodobieństwa, jak to wszystko przebiega, zbudować model Markowa, by móc wnioskować o rzeczywistości, która nam to generuje. Mamy czynniki, które wpływają na nasze obserwacje. Nie jesteśmy w stanie bezpośrednio badać tych czynników, ale znamy zależności pomiędzy nimi. Chcemy zbudować układ, w którym będziemy w łatwy sposób monitorować i przewidywać nasze obserwacje.

**Zadanie 4:** Mamy zbiór wielu HMM, zbudowanych dla danego ciągu w taki sposób, że jesteśmy w stanie w nich zaobserwować dany ciąg obserwacji. Mamy ciąg obserwacji, na nim budujemy dowolną liczbę HMM. Pytanie: Jak wybrać ten najbardziej prawdopodobny dla naszego stanu?

Podstawowe zadania HMM 
Dla Zadanego HMM i Zadanego ciągu obserwacji (Ol, 
oceń jak jest to prawdopodobne zdarzenie 
. Ok) 

Slajd 10:

Jaka jest szansa, że wygenerujemy ciąg obserwacji?

Im więcej jest stanów i obserwacja jest dłuższa , tym prawdopodobieństwa będą niższe. **Ale my będziemy w stanie na tej podstawie porównać ze sobą dwa ciągi obserwacji i powiedzieć, który jest bardziej prawdopodobny.**

Podstawowe zadania НММ 
О 
= (01, ф, 03) (dry.damp,soggy) 

Slajd 11:

Przykład: Czerwoną linią oddzielone jest to co widzimy, od tego czego nie widzimy.

My stoimy na czerwonej linii zwróceni plecami do słoneczek, nasza twarz jest zwrócona do góry. Obserwujemy niebieskie kółka. Obserwujemy wodorosty i wilgotność wodorostów. Mamy stan rozmoczony, wilgotny, wysychający i suchy. Czego nie widzimy? Nie widzimy stanów ukrytych. Za nasze stany ukryte rozumiemy stany odpowiadające pogodzie: słońce, zachmurzenie i deszcz.

Będziemy się starali dokonać oszacowania prawdopodobieństwa, jak bardzo prawdopodobne jest uzyskanie obserwacji (stanu w naszym układzie).

Możemy się zastanowić, jak to wszystko przebiegało. Strzałki mogą pokazywać, jak to mniej więcej wygląda. Widać, że dane stany obserwujące są w pewien sposób emitowane przez stany ukryte. Czyli dane stany, które widzimy (wilgotności wodorostów) są w bezpośredni sposób zależne od aktualnych stanów ukrytych. Nie interesują nas prawdopodobieństwa pomiędzy stanami obserwowanymi. Ale chcemy przechodzić przez stany ukryty. Chcemy wiedzieć w jaki sposób w naszym układzie przemieszczamy się pomiędzy stanami ukrytymi. Szare zakręcone strzałki u dołu pokazują, w jaki sposób przechodzimy pomiędzy stanami. Możemy przejść między każdym stanem i jest też duże prawdopodobieństwo, że pozostaniemy w danym stanie (utrzymanie stanu). Są podane kierunki strzałek, więc tutaj mamy do czynienia z grafem skierowanym. Więc będziemy mieli oszacowane prawdopodobieństwa wystąpienia tego wszystkiego razem.

Podstawowe zadania HMM 
O (Or. 02, 03) (dry, damp, soggy) 
P(OIHMM) 
12/67 

Slajd 12:

Gdybyśmy rozpisali nasze zaobserwowane stany i powyżej nich wpisali stany ukryte, to tak naprawdę przechodząc przez wszystkie stany ukryte, jesteśmy w stanie podać obserwację danego stanu. Jakie jest prawdopodobieństwo, że przechodząc ze stanu słonecznego, przejdziemy do stanu pochmurnego i przejdziemy do stanu deszczowego i w tym momencie zaobserwujemy te stany, które chcemy. Czyli, że było wyschnięte, podeschnięte, wilgotne i rozmoczone.

Poruszając się po strzałkach między boksami, chcielibyśmy móc oszacować prawdopodobieństwo danych. Jeżeli jesteśmy w dowolnym stanie ukrytym, jesteśmy w stanie zaobserwować dowolny stan ze stanu, który obserwujemy (czyli np. w podczas słońca nasze wodorosty nie musza być suche). **Dowolny stan, który zaobserwowaliśmy, mógł zostać wyprodukowany przez dowolny stan ukryty.**

Podstawowe zadania HMM 
O 02, 03) (dry. damp, soggy) 
P(OIHMM) = 
Sunny, sunny)) 
sunny. cloudy)) + 
rain, rain)) 
H2, H3)) 

Slajd 13:

**Liczymy prawdopodobieństwa:**

* obserwacji pod warunkiem, że mieliśmy stany ukryte tylko słoneczne (pierwszy wiersz naszego rysunku ze slajdu 12). To oznacza, że zaczęliśmy od stanu słonecznego, zgodnie z macierzą przejść dla stanów ukrytych, przeszliśmy znowu do słonecznego i znowu do słonecznego.
* Wybieramy daną konfigurację stanów ukrytych. Rozważamy wszystkie takie konfiguracje i otrzymujemy prawdopodobieństwo naszego modelu
* **Ilość składników naszej sumy jest proporcjonalna do ilości stanów ukrytych do potęgi długości obserwacji.** Mamy w 3 stanach ukrytych i 3 obserwacjach 33 = 27. To nie jest duży wynik, jednak kiedy zmieni się którykolwiek ze składników, zwiększa się złożoność obliczeniowa.

Podejscie naiwne 
Zwraca povawny wynik 
Wymaga policzenia nlOl prawdopodobieństw, gdzie n to liczba 
stanów ukrytych a O to ciąg obserwacji, dla naszego 
przypadku 33 27 
Nie korzysta z własności Markowa 

Slajd 15:

**Podejście naiwne będzie nam zwracało poprawny wynik, ale wymaga ono obliczeń n|O|.** Wynik będzie poprawny, natomiast jest on okupiony dużymi obliczeniami.

Forward algorithm 
Rozważmy obliczanie prawdopodobieństw wystąpienia 
sekwencji stanów zadanych przez HMM 

Forward algorithm 
= paths to x before t) 

Slajd 16 i 18 :

Alternatywą jest **algorytm Forward** - jego zadaniem jest akumulowanie prawdopodobieństw z poprzednich stanów i sekwencja obserwacji będzie zależna od markowowskości (?) naszego podejścia.

(2 aaopq x sq•ed (x) 'u 

Będziemy sumować wszystkie przeszłe zdarzenia w taki sposób, w jaki one wpływają lub doprowadzają do danego stanu w chwili t. Alfa jest kumulantą i zawiera prawdopodobieństwa obserwacji danego stanu pod warunkiem, że stan ukryty miał taką a nie inną wartość. Mnożymy to przez wszystkie ścieżki z przeszłości (przed aktualnym stanem t)

**Alfa jest kumulantą, zawiera prawdopodobieństwo obserwacji danego stanu, pod warunkiem, że stan ukryty miał taką a nie inną wartość**

Wyrażamy to przez wszystkie ścieżki przed t, które prowadzą do naszego aktualnego stanu.

Z filmiku:

Markov proces

**To zdarzenie, które jest bezpamięciowe** – oznacza to, że teraźniejsze zdarzenie jest zależne tylko od tego, co dzieje się teraz, a nie co zdarzyło się wcześniej.

Przykład: osoba podróżująca między lokacjami (stanami)

Prawdopodobieństwa tranzycji w tym przypadku określają prawdopodobieństwo tego, gdzie pojedzie ta osoba. Za każdym razem, gdy osoba podróżuje mówimy o zdarzeniu tranzycji. Gdy jest w domu, to ma określone prawdopodobieństwa podróży w dane lokacje. Żeby określić prawdopodobieństwo podróży w daną lokację, musimy najpierw zobaczyć cały proces podróżowania w te wszystkie lokacje i zobaczyć, jak często bywa w danej lokacji. Mając prawdopodobieństwa, możemy skonstruować macierz tranzycji – każda kolumna macierzy reprezentuje stan (lokację), z której nastąpiło przejście do innej, a wiersz reprezentuje stan (lokację), do której zmierza w podróży.

Aby użyć tej macierzy, potrzebujemy najpierw początkowego wektora prawdopodobieństwa. Gdy taką macierz przemnożymy przez ten wektor, to dostaniemy nowy wektor – odpowiada on za predykcję tego, gdzie będzie dana osoba w następnym kroku.

**Ważne: stan zależy wyłącznie od stanu poprzedniego i jest niezależny od czasu**

**Prawdopodobieństwo uzyskania danego stanu zależy tylko od stanu poprzedniego**

Ukryty model Markowa

Przykład z pogodą: określanie prawdopodobieństwa danej pogody.

**Prawdopodobieństwa wpisujemy w macierz przejść** – pokazuje ona jak nasze stany ukryte mogą się zmieniać.

**Macierz emisji** – macierz mówiąca jaka jest szansa, w danym stanie, obserwacji stanów obserwacyjnych, jak stany ukryte wpływają na te, które faktycznie obserwujemy

Do opisu potrzebna jest macierz transmisji i emisji oraz początkowe prawdopodobieństwo rozkładu **(wektor)**. Całą tą trójkę nazywamy ukrytym modelem Markowa. Będzie ona determinować, jakie stany będą obserwowane.

Forward algorithm 
Ol(x) = P(OIIHI = = 
— p(ot411Ht41 = x) = 
OIHMM) — 

Slajd 30:

**Algorytm forward** – sumuje ścieżki z przeszłości do danego stanu

Prawdopodobieństwo dla danego stanu: suma dla wszystkich i, gdzie alfa ma długość tyle, jaka jest długość obserwacji. Alfy liczymy dla wszystkich stanów ukrytych.

Prawdopodobieństwo obserwacji danego stanu w danym stanie – element macierzy emisji. Mówi nam np. jakie jest praw zaobserwowania suchego, kiedy byliśmy w stanie słonecznym. To jest pierwsza konfiguracja i ona jest w pierwszej kolumnie.

Pierwszy element iloczynu dla alfat+1(x): prawdopodobieństwo np. soggy pod warunkiem cloudy.

Drugi element: prawdopodobieństwo, że ze wszystkich ścieżek poprzednich dotarliśmy do stanu x.

Forward algorithm 
Policz Ol(x) dla każdego stanu ukrytego x, dla pierwszej 
obserwacji Ol 
Dla każdej obserwacji Ot (t > l) i każdego Stanu ukrytego x 
policz nt(x) używając 
Zwróć sumę 0101(x) po wszystkich stanach ukrytych x 

Forward algorithm 
Zwraca poprawny wynik 
Ma złożoność IOln (zamiast "IQ) (przy ewaluacji Od lewej do 
prawej) 
Korzysta z własności Markowa 
32/78 

Slajd 32

Algorytm forward ma mniejszą złożoność, która nie jest potęgowa, ale jest iloczynem długości obserwacji i ilości stanów ukrytych. Korzystamy z własności Markowa, gdyż wpływ na to w jakim stanie jesteśmy, mają ścieżki, w których byliśmy (??????).

Tagging 
Tagowanie tekstu 
Mając dany tekst w formie ciągu słów d = wordl... , oraz 
ciągu tagÓw Cd = tagł. _ tagk celem łagowania jest zbudowanie 
funkcji f words rags. która dobrze odwzorcn•vuje to łagowanie. 
tj. minimalizuje jakąś funkcje błędu td). 
42,' 78 

Slajd 42

**Tagowanie tekstu** – określanie części mowy

Naiwne podejście

**Stany ukryte** – części mowy, **stany obserwowalne** – słowa z danego języka.

OkreSlanie cz$ci mowy - Part of speech 
Noun: a part of speech inflected for case. signifying a 
concrete or abstract entity 
Verb' a part of speech without case inflection. but inflected 
for tense, person and number, signifying an activity or 
performed or undergone 
Participle: a part of speech sharing the features of the verb 
and the noun 
Interjection: a part Of speech expressing emotion alone 
Pronoun: a part Of speech substitutable for a noun and 
marked for a person 
Preposition. a part of speech placed before other words in 
composition and in syntax 
Adverb. a part Of speech Without inflection, in modification Of 
Or in addition to a verb, adjective, clause, sentence, Or Other 
adverb 

Naiwne podejście 
Dla danego słowa w 
Jeśli w nie ma w zbiorze uczącym to zwróć tag, który 
występuje najczęściej w tym zbiorze 
W przeciwnym razie zwróć rag, który występuje w zbiorze 
uczącym najczęściej dla słowa w 
46/78 

**Viterbi**

Tekst alternatywny wygenerowany przez komputer:
Viterbi - najbardziej prawdopodobne ścieżki 
59/78 

Tekst alternatywny wygenerowany przez komputer:
Viterbi - 
najbardziej prawdopodobne ścieżki 
T<t-l 
Pbesr ) 
x 
60/78 

Algorytm wybiera najbardziej prawdopodobne ścieżki, żeby dojść do celu. Algorytm stara się wybierać te stany i przechodzić między tymi stanami, które są najbardziej prawdopodobne, które razem dają najbardziej prawdopodobną konfigurację stanów ukrytych.

Slajd 61

Tekst alternatywny wygenerowany przez komputer:
Viterbi - 
najbardziej prawdopodobne ścieżki 
Najlepsza ścieżka do x musi mieć postać: 
Pbest.t(X) = Pbest.t-l 
61/78 

Jak przechodzimy z chwili t-1 do chwili t – najlepsze ścieżki

**Musimy wybierać stany, w których jest największe prawdopodobieństwo**. Najbardziej praw stany ukryte to te stany, w których jest największe prawdopodobieństwo wygenerowania danej obserwacji.

Jak zrobimy tak dla każdego stanu, to niekoniecznie otrzymamy poprawne rozwiązanie, bo nie znamy historii (????).