Марковски вериги и вероятностни автомати. Скрити Марковски модели за разпознаване на говорни сигнали. Понятие за кепстър на говорен сигнал. Приложение при разпознаване на ръкопис

# Марковски вериги

Марковска верига е математическа система, която изпълнява преходи между изброим брой състояния.

Формално, Марковска верига е редица от случайни величини *X1, X2, ..., Xn* имащи свойството на Марков, а именно, че при дадено текущо състояние, бъдещите и миналите състояния са независими:

Веригите на Марков често се описват с насочен граф, където ребрата са означени с вероятностите за преход от едно състояние към друго.

# Вероятностни автомати

Вероятностите автомати са обобщение на недетерминираните крайни автомати и веригите на Марков – те включват вероятността за даден преход във функцията на преходите, т.е. матрицата на преходите е стохастична матрица.

# Скрити Марковски модели за разпознаване на говорни сигнали

## Разпознаване на отделни думи

Нека са дадени:

* Входна редица:
където *o(1), …, o(T)* са параметрични вектори изчислени за всеки „кадър“ на входния сигнал
* Речник от *N* думи

Искаме да разберем на коя дума отговаря входната редица. В термините на вероятности:

където е условната вероятност на думата *wi* при фиксирано *O*.

При разпознаването на говор, обаче не може да бъде изчислено директно. Прилагайки формулата на Бейс, получаваме:

* *P(w)* е априорната вероятност на думите, определена от езиковия модел. В частния случай, в който всички думи са равновероятни тя няма ефект за максимизирането.
* *P(O)* е априорната вероятност да наблюдаваме поредицата *O*. Тя не може да се изчисли, но и не зависи от думата *w*, съответно няма ефект за максимизирането.
* *P(O|w)* е вероятността за поредицата *O* при фиксирана дума *w*. За изчисляването и използваме *модела*на думата *w* – *M*.

Моделът е вероятностен автомат с няколко състояния. Всички състояния без първото и последното наричаме *предавателни* – те могат да генерират различни вектори. За всяко от тези състояния има специфицирана *вероятностна плътност (emission probability density function (epdf)) – bj(o)*.

**Преходните вероятности***aij* трябва да отговарят на следното свойство -

В примера ще приемем, че съществуват само три типа преходи – *ai,i,ai,i+1,ai,i+2*.



**Фигура 1: Примерна конфигурация на СММ. Редицата от състояние е *X = [1, 2, 2, 3, 4, 4, 5, 6]*.**

Модела може да бъде описан и с матрицата:

Болшинството от моделите за автоматично разпознаване на говор са „отляво-надясно“, т.е. не е възможно връщане в предходно състояние – *ai,j = 0, i> j*.

Съществуват два начина за дефиниране на функциите *epdf* – като дискретни и като непрекъснати. В дискретния случай, наблюденията първо се преобразуват в дискретни символи, чрез класификатор базиран на минимално разстояние и се задават в табличен вид.

В непрекъснатия случай, функциите се задават чрез разпределения, като най-честия избор е Гаусовото разпределение.

Нека първо разгледаме тривиалния случай – вектора *o* съдържа един елемент (практически това не се случва никога). Плътността на *j­-*тото състояние се задава с едномерното разпределение:

В действителност обаче, имаме многомерни параметрични вектори, съответно разпределението също е многомерно. Ако размерността на вектора е *N*, то:

В повечето системи обаче, параметрите във вектора не корелират, съответно матрицата на корелацията е диагонална. Благодарение на това, вместо да се налага да обучим *NxN* коефициента на корелация, е достатъчно да намерим приближенията на *N* стандартни отклонения – вектора *σj*. Като следствие, *N*-мерното разпределение е произведение на *N* независими едномерни разпределения:

По този начина модел ще има *2N* параметъра за всяко предавателно състояние, вместо *N + NxN*.

В много случаи моделирането само с един Гаусиан не е достатъчно за обхващането на характеристиките на всички говорещи.

Входната редица *O* може да „премине“ през модела използвайки редицата от състояние *X* (с дължина *T+2*). *X* наричаме *път* през модела. Всеки път съдържа двете формални състояния – началното и крайното. Тъй като обикновено пътя не е „видим“ извън системата, моделите се наричат *скрити*.

**Изчисление на вероятността модела *M* да генерира редицата *O***

Вероятността за път *X* при входна редица *O* и модел *M* е

Практически, това е произведението на всички вероятности срещнати по пътя. Можем да обобщим това до всички пътища, който разпознават входната редица по два начина:

* Вероятност на Баум-Уелч –
* Вероятност на Витерби –

**Обучаване на параметрите**

За да обучим параметрите на модела *Mi* използваме обучаващите редици, които отговарят на думата *wi*. Обикновено на всяка дума съответстват голям брой обучаващи редици. В примера ще разгледаме, когато имаме една.

Първоначално построяваме груба оценка на математическото очакване и на диагонала на матрицата на ковариацията използвайки глобалното средно и глобалния вектор на отклонението. След това итеративно подобряваме това приближение. Макар че, е възможно да се фиксират теглата на векторите, по-добрия вариант е те да се преизчисляват на всяка итерация, базирайки се на това колко допринасят за преизчисляването на новото състояние. Нека бележим с *Lj(t)*приноса на вектор към състоянието. Тогава можем да преизчислим по следния начин:

За изчисляването на функциите на приноса е практично да изчисляваме вероятността на всички пътища приключващи в състояние *j* в момента *t* и вероятността на всички пътища започващи от състояние *j* в момента *t*, които ще бележим съответно с *αj(t)*и *βj(t). αj(t)* можем да изчислим по следния начин

Аналогично за *βj(t)*:

Така пресметнатите вероятности можем да комбинираме по следния начин:

Нормализирайки с пълната вероятност за предаване:

# Понятие за кепстър на говорен сигнал

Кепстър (cepstrum – анаграма на spectrum) на сигнал се нарича прилагането на трансформацията на Фурие върху логаритъма от спектъра на сигнала. Т.е. кепстър на сигнала *c(t)* можем да дефинираме по следния начин:

Където параметъра *τ* наричаме quefrency (анаграма на frequency). Съществено удобство на работата в кепстралното пространство е свойството, че конволюцията на два сигнала се изразява чрез сума на техните кепструми, т.е. ако:

то:

Кепстъра се използва в обработката на говорен сигнал, тъй като нискочестотните периодични скокове от гласните струни и високочестотните сигнали от устната кухина конволюират във времевия домен, умножават се в честотния, но са адитивни във quefrency домена.

# Приложение на СММ при разпознаване на ръкопис

Първите приложения използващи СММ за разпознаване на текст се появяват през 90те и има разработени редица различни подходи. В частност ще разгледаме подхода на K. Aasи L. Eikvil за разпознаване на ръкопис чрез наситеност на сивото. Основната идея на този подход е да се избегнат трудностите, които се пораждат при разделно сегментиране и разпознаване, тъй като на практика много приложения довеждат до лоши резултати именно поради лошо качество на сегментирането. Използвайки СММ, авторите предлагат процедура, която да извършва едновременно сегментиране и разпознаване. В допълнение СММ позволяват лесното интегриране на контекстуална информация – както лингвистична, така и свързана с конкретния текст.

В предложения алгоритъм авторите разделят текста на отделни думи и разделят всяка дума на колони съответстващи на векторите от атрибути и фиксират дължината на векторите на 18. По този начин те получават времеви серии с различна дължина за всяка дума.

**Фигура 2. Разделяне на буква на 6 състояния.**

Символите биват моделирани с 6 състояния и за всяка дума последователността от вектори от атрибути се съпоставя на всички възможни състояния, за да се открие най-добре пасващото. Макар изчислително сложен, този подход се стреми към намирането на оптимално решение.

# Източници

Jan Cernocky; Hidden Markov Models – an Introduction

KjerstiAas, Line Eikvil, Ragnar Bang Huseby; Application of Hidden Markov Chains in Image Analysis

Lawrence Rabiner, Biing-Hwang Juang; Fundamentals of Speech Recognition

Lawrence Rabiner, Ronald Schafer; Introduction to Digital Speech Processing

Daniel Ramage; Hidden Markov Models Fundamentals

Wikimedia Foundation; http://en.wikipedia.org